

The use of time during lexical processing and segmentation: A review

SVEN L. MATTYS

State University of New York, Stony Brook, New York

Speech, by its very nature, is a time-based phenomenon. Speech sounds are temporally distributed, with the presentation of one sound roughly conditioned by the fading of the previous one. In this review, three classes of models are discussed with respect to the sequential nature of speech. It is argued that the three resulting conceptions of time are linked to the type of segmentation process proposed by these models to deal with speech continuity. In the first one, lexical activation is viewed as perfectly synchronized with the temporal deployment of speech. This type of model corresponds to the traditional left-to-right (proactive) account of lexical processing. Because serious segmentation problems exist for such an approach (e.g., *car* and *card* are embedded in *cardinal*), the second type of model treats word recognition as the result of a mechanism that sometimes delays commitment on word identity beyond word offset. Lexical activation, instead of shadowing the unfolding of time, lags behind it until an unambiguous decision can be made. The temporarily unprocessed information is stored in a memory buffer. In the third approach, a prosodic cue (lexical stress) contributes actively to speech segmentation and lexical processing. Every stressed syllable encountered in the signal is postulated as a word onset and thus constitutes the starting point of lexical activation. However, with non-initial-stressed words, retroactive procedures going “back in time” must be used. Finally, the use of time (including proactive, delayed, and retroactive procedures) is discussed in light of cross-linguistic phonological differences.

How spoken words are recognized is a major concern in psycholinguistics. Models posited to explain this complex cognitive skill have traditionally ascribed a central position to the notion of *lexical processing*, which corresponds to the multiple steps that lead to the recognition of a word (Frauenfelder & Tyler, 1987). Lexical processing includes three stages: (1) the initial contact between the input signal and the mental lexicon, (2) the activation of candidates in the lexicon and their progressive selection, and (3) the recognition of the actual word. The term *lexical access* is usually reserved to describe the moment at which lexical information becomes available to the speech processor.

Approaching lexical processing and segmentation from a time perspective is unconventional in the field of word recognition. Traditionally, research has mainly focused on the informational side of speech processing (amount of information, number of phonemes, type of perceptual units, quality of the input, etc.). However, it is possible to examine and classify the models built on this type of evidence along a time-based typology. In this review, I have

isolated three categories of time uses in which most current models of word recognition can fit. Interestingly, the description of these categories happens to be based largely on the evidence related to the speech segmentation mechanisms supported by the models under categorization. This overlap is due to the intimate relationship between segmentation mechanisms and the use of time. My goal in the present paper is to unravel this relationship.

The way in which lexical processing is carried out over time can be modeled on a continuum between two contrasting principles. According to the first principle, activation in the lexicon follows the deployment of speech in a time-shadowing way. In other words, speech information occurring first in time is always processed first. For example, with the word *technique*, the syllable *tech* would be processed before the syllable *nique* because *tech* occurs first in time. According to the second principle, even though the physical variable of time is defined as a before/after sequential order, the information occurring first in time is not always processed first. In this view, processing primacy is not given to temporally early information but to that which is perceptually salient (e.g., stressed syllables). As a consequence, late information may sometimes be processed before earlier information. For example, if stressed syllables are considered to be the leading salient events, the word *technique* would be recognized through the early processing of *nique*. The syllable *tech*, although coming earlier in time, would be decoded retroactively with the help of the lexical information provided through *nique*.

This work was supported by National Institute of Mental Health Grant R01 MH5166301. I am very grateful to Arty Samuel for his supervision of this work and to the following members of the speech laboratory: Donna Kat and Lee Wurm. Thanks are also due Susan Brennan, Richard Gerrig, and John Robinson for discussing the ideas presented in this paper with me, as well as Robert Crowder, Dominic Massaro, and two anonymous reviewers for their helpful comments on previous drafts. Correspondence concerning this article should be addressed to S. L. Mattys, Department of Psychology, SUNY Stony Brook, Stony Brook, NY 11794-2500 (e-mail: smattys@psych1.psy.sunysb.edu).

The respective implications of these two views are not trivial. On one end of the continuum, the temporality of speech processing can be thought of as sequential and universally identical. All languages would be processed in a proactive way, with early information contributing to the processing of later information. In contrast, the other end of the continuum opens a realm of new predictions regarding, among others, cross-linguistic differences. For example, insofar as the nature of perceptually salient information varies from one language to another, the use of time during lexical processing could, too, be different across languages. Because of their phonological specificities, languages may require different degrees of proactive and retroactive processing.

In the present paper, I review the evidence supporting various theoretical positions on the sequential/nonsequential continuum as it pertains to English. To do so, I first examine how the notion of time has traditionally been considered (or avoided) in studies on word recognition. The goal is to show that, despite the intrinsic correlation between time and the speech information that it brings to the listener, these two variables have an independent impact on lexical processing.

Next, I discuss three classic types of models which, although originally designed to deal with lexical processing and segmentation, embody specific implications concerning the use of time. The differences among the models originate principally from the distinct solutions that they offer to the problem of speech continuity and the absence of word-boundary cues in the signal. The first view suggests that lexical activation closely shadows the unfolding of time; words in the lexicon are activated sequentially as the information appears. The second view posits that recognition cannot always be successfully achieved in a time-shadowing fashion. Often, information beyond a word's offset has to be processed before recognition can take place. Thus, in this account, the system is required to delay the decision about the word's identity in order to avoid incorrect lexical hypotheses. The information temporarily left unprocessed during this operation is stored in a buffer. The nature and capacity of this buffer are discussed.

Contrary to the first two models, which posit segmentation as a by-product of word recognition, the third model considers segmentation an independent and explicit mechanism. In this model, the ambiguity at word boundaries (responsible for delayed commitment in the second view) is dealt with by calling upon a suprasegmental cue such as lexical stress in order to segment the speech stream. For example, initiating lexical access on every stressed syllable of the input would be a relatively successful heuristic to use in English, because a high proportion of English words begin with a stressed syllable. With initial-stressed words, lexical activation would progress sequentially (or proactively), from left to right. However, with non-initial-stressed words, the stress-based strategy would fail to segment the input properly. The system would repair faulty segmentation by proceeding retro-

actively, against the unfolding of time, and by accessing the lexicon with an earlier (weak) syllable.

Finally, I propose that realistic models of word recognition require a combination of proactive, delayed, and retroactive procedures. I suggest that the respective contribution of these three temporal procedures is tied to the structural particularities of the lexicon and to the phonological characteristics of a given language, two aspects that often determine the extent of segmentation problems. Illustrations of this language-dependent time principle are briefly presented for non-English languages.

THE NATURAL CONFOUNDING OF TIME AND INFORMATION

To examine how lexical processing is temporally carried out by the speech system, we first need to understand the temporality of the signal itself—that is, the way in which it is produced. For anatomical reasons, speech is produced fairly sequentially, with limited overlap between phonemes. For example, it is impossible to pronounce all the phonemes of the word *speech* at once. Phonemes are instead successively uttered in rapid and continuous strings that constitute words and sentences. Nevertheless, more meticulous spectral analysis reveals that each short segment of speech carries information on one phoneme together with its bordering phonemes (A. M. Liberman, 1970a, 1970b; A. M. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Massaro, 1975; Studdert-Kennedy, 1975). For example, what one could isolate as the spectral representation of the phoneme /s/ appears to differ in the syllables /su/ and /si/ because of the specific articulatory constraints generated by the adjacent vowel. This phenomenon, which can sometimes span over several phonemes, is usually referred to as *coarticulation*. Thus, in the same way that physical and muscular aspects of the vocal tract limit the number of phonemes that can be uttered at once, they also prevent pure sequentiality. However, despite this segmental overlap, speech production still has an overriding sequential component. The unfolding of information follows the unfolding of time rather closely.

Does lexical processing follow the sequential way in which speech is produced or does it unfold in a more time-independent way? To answer this question, it is important to understand what the notion of time refers to in speech processing and, specifically, how time and the sensory information that it conveys combine. The passage of time is usually filled with incoming sensory information. Conversely, speech information takes time to be produced. For example, the sequences /klær/ and /klæri/ differ not only in the amount of phonemic information that they bring to the processor but also in the amount of time that it takes to present them. Thus, when researchers study how different amounts of information affect subjects' responses, they also measure the effect of processing time as it is confounded with information input. Both the informational and the durational factors

could be responsible for the activation state of the system at a given time and, hence, for the subjects' responses (Zwitserslood & Schriefers, 1995).

Most current models of spoken word recognition are based on research done at the informational level only. Their main goal is to determine the activation state of the lexicon when the system is presented with a given input and, eventually, to specify the amount of information necessary for recognition of a word (Grosjean, 1980; Marslen-Wilson, 1985; McClelland & Elman, 1986; Salasoo & Pisoni, 1985; Tyler, 1984; Zwitserslood, 1989). For example, in the gating paradigm (Grosjean, 1980), subjects are presented with words that are interrupted before their offset (e.g., the presentation of a word such as *spaghetti* could be interrupted after /spæg/). The subject's task is to guess the identity of the full word. By varying the locus of the interruption, one can establish how much of the word must be heard for it to reach reliable recognition. However, processing time is usually not controlled. As a result, researchers overlook two levels at which time could intervene. First, because speech rate is not manipulated (i.e., the amount of information per unit of time is always the same), it is impossible to determine the contribution of time alone in the recognition performance. Second, reaction times are traditionally not measured in gating experiments. Subjects have as much time as they need (for an exception, however, see Zwitserslood, 1989). It is thus impossible to determine the amount of time that is necessary in order to produce a response.

On the other hand, the durational aspect of speech processing—that is, the role of time *per se*—has been addressed in various research areas as well. Although most research has been done at the acoustical level (e.g., Bastian, Eimas, & A. M. Liberman, 1961; Dorman, Raphael, & A. M. Liberman, 1979; Lisker, 1957; Marcus, 1978; Port, 1976; Summerfield & Bailey, 1977), a small number of studies have focused on the effect of time at the word level. In these studies, time, generally referred to as *processing time*, is used as an independent variable in itself, either as part of the experimental design or in an exploratory post hoc way. The rationale is that, given that the cognitive operations involved in speech recognition can presumably differ depending on the amount of processing time available, it is important to analyze the results at successive points in time during the task (Dupoux, 1993). Specifically, data analyses differentiating between fast and slow respondents can be of great interest if one believes that fast responses are, as a rule, generated at an early processing level whereas slower responses may originate from more attentional, strategy-guided stages. Experiments done with data partitioned on the basis of latency range, or with speeded/delayed response designs, have revealed patterns of results that varied considerably as a function of the amount of processing time (e.g., Dupoux, 1989; Fox, 1984; J. L. Miller & Dexter, 1988; Sebastián-Gallés, Dupoux, Segui, & Mehler, 1992).

The "temporal inertia" in the lexicon is clearly illustrated in studies of lexical ambiguity and multiple lexi-

cal activation (Tabossi, 1988; Zwitserslood, 1989; see Simpson, 1984, for a review of earlier work). In this type of study, the life span of the meanings of a homophonic word (e.g., *ring*) is gauged by varying the time interval between the offset of this word presented in a sentential context and the occurrence of a visual word related to one of its meanings (e.g., *finger* or *bell*). Latencies to perform lexical decision on the two visual words are compared at different time intervals. Research has shown that, although all the different meanings of an ambiguous word may be activated at first (Swinney, 1979), the context is soon used to select the appropriate meaning and suppress the inappropriate one(s) (Dooling, 1972; Oden & Spira, 1983; Seidenberg, Tanenhaus, Leiman, & Bienkowsky, 1982; Tanenhaus, Leiman, & Seidenberg, 1979). Seidenberg et al. (1982) and Tanenhaus et al. (1979) suggest that this selection process is completed within about 200 msec (but see Burgess, Tanenhaus, & Seidenberg, 1989, and Glucksberg, Kreuz, & Rho, 1986, for diverging results).

In a similar way, Cutler (1986) found that the stress pattern of a speech signal needed a certain amount of time to constrain lexical access. For example, a word like *DISCOUNT*, bearing a first stressed syllable and a second unstressed syllable, was shown to produce a priming effect on the meanings associated with both the noun *DISCOUNT* and the verb *DISCOUNT*. However, when the test word was presented 750 msec after the prime instead of immediately after it, a priming effect was found for the stress-associated test word only. Taken together, the results obtained with homophonic and prosodic cross-modal priming show that the lexical activation produced by a given input can change simply by allowing extra processing time.

The relationship between processing time and amount of sensory information has been analyzed explicitly in only a few studies. For example, Massaro (1972b) found that the pitch perception of single tones could be improved either by increasing the duration of the test tone presented to the subjects or by allowing more processing time after the test tone offset. In fact, processing time turned out to be a more critical factor than did the duration of the stimulus, with an optimal processing time of 250 msec. Similarly, using vowel recognition tasks, Massaro (1974) and Massaro and Idson (1978) demonstrated that a short fragment of a vowel plus additional processing time improved recognition performance to the same extent as did the same vowel presented for the cumulative duration. Thus, it appears not only that processing time is a critical factor during speech recognition but also that it can compensate for a shortage of sensory input.

Amount of processing time and amount of information seem to interact similarly at the level of word recognition. In a cross-modal experiment carried out in Dutch, Zwitserslood and Schriefers (1995) compared the effects of three auditory primes on the lexical decision of a subsequent visual target. The three primes were: (1) a short onset fragment of a word (e.g., /kae/ from *kapitein* [*captain*]), (2) a long onset fragment from the same word (/kaep/),

and (3) the short fragment (/kae/) followed by an inter-stimulus interval (ISI) amounting to the difference in duration between short and long primes in a given set. The authors found that the short fragment was not sufficient to produce a priming effect on the lexical decision. However, the long fragment and the short + ISI conditions both yielded reliable priming effects. On the basis of these results, Zwitserlood and Schriefers concluded that the information present in the short fragment was sufficient to activate the lexicon and produce priming, provided that it was given enough time to become operative. These findings concur with those reported on pitch perception and vowel recognition (Massaro, 1972b, 1974; Massaro & Idson, 1978) in that they illustrate the distinct contribution of processing time and sensory information to speech recognition, despite their confounded nature in normal listening situations.

Another way to disentangle the respective role of time and information is to vary the amount of sensory information per unit of time through speech compression or expansion. For example, numerous researchers have analyzed the effect of speech rate on the comprehension of a spoken message and on the intelligibility of its individual words (e.g., Aaronson, 1974; Fairbanks, Guttman, & Miron, 1957; Foulke, 1969; Garvey, 1953). One very consistent finding is that listeners are able to deal with various speech rates successfully,¹ probably by taking into account the rate at which the speech is produced, a process called *perceptual normalization of the input* (J. L. Miller, 1981, 1987; J. L. Miller & A. M. Liberman, 1979). For example, investigations of time-compressed speech reveal that the accuracy with which a word presented in isolation can be reported remains above 90% as long as the compression rate does not exceed 50% (Fairbanks et al., 1957; Garvey, 1953). Thus, listening comprehension, on the whole, appears to be relatively unaffected by variations in the amount of information per unit of time.

A few studies have explored the separate effects of speech rate and amount of information with isolated words. When Dupoux and Mehler (1990) measured the detection latencies for initial stop consonants in high-frequency and low-frequency French words, they found a frequency effect in monosyllabic words but not in bisyllabic words. Subjects were faster at detecting an initial /t/ in high-frequency monosyllables such as *texte* (*text*) than in low-frequency monosyllables such as *teinte* (*tint*). There was no significant difference between high-frequency bisyllables such as *terrible* (*terrible*) and low-frequency bisyllables such as *terrasse* (*terrace*). This interaction was explained in terms of different processing routes used with monosyllabic and bisyllabic words (Cutler & Norris, 1979; Foss & Blank, 1980; Newman & Dell, 1978). Monosyllables would be processed via an access to the lexical information—the lexical route—thus creating a lexical effect. The bisyllables would be too long to yield lexicon-based reaction times. A simple acoustic/phonetic decoding—the prelexical route—would be sufficient. Such a route would not be affected by a lexical parameter such as frequency. In order to de-

termine whether the length effect was structural (one vs. two syllables) or durational in nature, the authors compressed all items to half their durations. Despite this change, the same results emerged. Compressed bisyllables, which were even shorter in duration than uncompressed monosyllables, still showed no frequency effect. Thus, the interaction between word length and word frequency was apparently due to a structural difference and not to a durational difference. These results suggest that lexical processing, as indicated by the frequency effect, depends more on the amount of information being processed than on the amount of time to process it.

Recently, this conclusion was moderated by Pitt and Samuel (1995), who showed that both the amount of information and the processing time allowed to the listener can have an effect on the activation state of the lexicon. In their experiments, the authors measured the detection latencies for various consonants in words and pseudowords. The words differed in the position of their uniqueness point. The uniqueness point (or UP) of a word is that point in the word where only one item in the lexicon matches the input (Marslen-Wilson & Welsh, 1978). For example, the UP of *spaghetti* is at /g/ because it is the only word that starts with the sequence /spæg/. In Pitt and Samuel's study, half the words had a UP in their early portion (early-unique words) and half had a UP in their late portion (late-unique words). Matched pseudowords were created by changing one or more phonemes in the target words. The pseudowords' *deviation point* (or DP, point at which the string diverges from any word in the lexicon) occurred by the end of the second or third phoneme in most cases. Initial results showed that both pre- and post-UP target consonants in late-unique words were detected faster than identical consonants occurring at the corresponding location in the pseudowords. In contrast, with early-unique words, the word advantage was found only on the detection of post-UP consonants. The authors suggested that the absence of a lexical effect before the UP in early-unique items was due to either insufficient information at such an early point in the words or insufficient processing time.

To tease apart these two confounded possibilities, Pitt and Samuel (1995) compressed the late-unique items to align their UP with the UP of the early-unique items. With this manipulation, early- and late-unique items became unique at the same point in time, though with different amounts of sensory information having been presented. Despite the compression, the lexical effect before the UP of the late-unique words was still present. This result suggests that reduced processing time does not affect the lexical activation produced by a sufficient amount of information, a finding consistent with Dupoux and Mehler's (1990). However, when Pitt and Samuel expanded the early-unique items to align their UP with that of the late-unique items, a lexical effect emerged before the UP of the early-unique words. This result suggests that even a few phonemes of input can produce the lexical activation necessary for lexical effects to appear, provided that the system is given enough processing time.

This finding is compatible with the previously reported work on pitch perception (Massaro, 1972b), vowel identification (Massaro, 1974; Massaro & Idson, 1978), and, more importantly, lexical processing (Zwitserslood & Schriefers, 1995). Indeed, all these studies point out the separate influences of processing time and sensory information on speech processing. In particular, they suggest that lexical activation lags slightly behind the temporal deployment of information. This delay would explain why extra processing time sometimes appears to compensate for a shortage of information.

In conclusion, time emerges as a critical factor during speech perception and lexical processing and seems to be independent of the amount of information supplied. However, the joint operation of time and information can be disentangled only indirectly because of the intrinsically confounded nature of these two variables in running speech. In the next section, I describe how time, both as a critical factor in itself and as a conduit for information, is the central aspect of sequential models of lexical processing, the first type of model presented in this review.

MODEL I

The Sequential Nature of Speech Processing

Until recently, there was fairly strong agreement among speech recognition models that the most straightforward way in which lexical processing could proceed was to follow the unfolding of time in a left-to-right manner (see, e.g., Chodorow, 1979; Cole, 1973; Cole & Jakimik, 1978, 1980a, 1980b; Marslen-Wilson, 1973, 1984; Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Zwitserslood, 1989; Nooteboom, 1981). This idea has been computationally developed in the influential *Cohort* model (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978), which, for a long time, dominated the field with an impressive body of evidence in several languages and various tasks. The present section is a review of this evidence and the ideas that it fostered. However, we will see in later sections that data have been accumulating in the last decade that show a much more complex relationship between time and lexical processing.

The idea behind sequential models of lexical processing is that in all languages, speech units (e.g., phonemes and syllables) are processed as they temporally occur, one by one. In this view, lexical activation starts as soon as any signal is available and the decision on the identity of the string occurs when there is enough evidence to do so without ambiguity. In some cases, the decision even occurs before the end of the word. The synchronicity of speech production and speech processing leads to what we might refer to as *time-shadowing decision*. The modeling of such a process received its first decisive application in the *Cohort* model (Marslen-Wilson & Welsh, 1978). In this model, a phase of lexical activation driven by the unfolding speech signal is followed, on line, by a phase of deactivation of the candidates that no longer match the gradually growing input. The result of this activation/selection process is the isolation of a unique can-

didate that matches the input while the others no longer do. The remaining candidate theoretically corresponds to the word to recognize. As indicated in the previous section, the point at which a word becomes unique (hence recognized), following this "left-to-right" process, is called the uniqueness point (UP). Such a model appears to be computationally inexpensive, because, in theory, only the speech before the UP must be processed to make word recognition possible. This is particularly economical with long words that have an early UP. For example, only the first four phonemes of the word *claustrophobic* would need to be decoded to achieve recognition, since it is the only word starting with the sequence /klɔs/. In addition, early recognition would allow the offset of a word to be anticipated, thus solving the complex problem of word segmentation in continuous speech (L. A. Taft, 1984).

A way to test the sequentiality of speech processing is to compare recognition of words with an early UP (early-unique words) to that of words with a late UP (late-unique words). If word recognition is a time-shadowing process, early-unique words should be recognized faster than late-unique words and, more generally, there should be a high correlation between the time that it takes to recognize a word and the position of its UP. Sequential effects, including UP effects, have been reported repeatedly in English in various tasks, such as shadowing (Marslen-Wilson, 1973), phoneme monitoring (Frauenfelder, Segui, & Dijkstra, 1990; Marslen-Wilson, 1984), lexical decision (Goodman & Huttenlocher, 1988; M. Taft & Hambly, 1986), mispronunciation detection (Jakimik, 1979; Ottevanger, 1984), gating (Grosjean, 1980; Tyler & Wessels, 1983), and phonemic restoration (Samuel, 1987). Similar effects have also been found in French in a task known as *gender classification*, in which subjects must decide whether the word they hear is masculine or feminine (Radeau, Mousty, & Bertelson, 1989), and in a shadowing task (Radeau & Morais, 1990). The time-shadowing aspect of word recognition has been clearly illustrated by Radeau et al. (1989), who pointed out that in their experiment, subjects responded before the end of the word in 20% of the early-UP items and 10% of the late-UP items.

These studies suggest that UP effects are robust across tasks and languages. However, Radeau, Mousty, and Padeloup (1994) have pointed out that these effects can be sensitive to time-related factors such as speech rate. When the authors speeded up the presentation of French test words from 2.2 to 5.6 syllables per second on the average, through either faster human delivery or computerized compression, UP effects disappeared. Interestingly, the fast rate corresponded to the average normal rate of continuous speech (Malécot, Johnston, & Kiziar, 1972). Therefore, Radeau et al. (1994) suggested that UP phenomena traditionally reported in the literature could arise from the abnormally slow speech rate used in the artificial laboratory situation. Word presentation slower than normal would provide subjects with extended processing time and hence enable them to use powerful strategies such as time-shadowing decision.

This study raises two comments. First, it is difficult to establish whether the UP effects in experiments other than Radeau et al.'s (1994) were really obtained with abnormally slow speech. Except for one UP study by Goodman and Huttenlocher (1988) in which the speech rate was reported and was indeed considerably slower than a natural rate, speech rates are usually not available. Second, and more importantly, although it is important to point out that UP effects might not occur at natural speech rates, Radeau et al.'s (1994) results do not necessarily demonstrate that sequential lexical processing is an artificial construct that occurs only with an unnaturally slow input. Indeed, as I have reported above, Zwitserlood and Schriefers (1995) indicated that the information that uniquely specifies a word may require some time before it can be reflected in the activation state of the recognition system and the subject's responses. Thus, with short presentation times, the temporal inertia of the system could mask UP effects. Moreover, we ought to remember that the UP and, a fortiori, UP effects are only theoretical concepts aimed at capturing a very general phenomenon—namely, the on-line nature of lexical processing. Such a cognitive mechanism may be measurable only with a slower speech input, a magnifying glass incidentally provided in most experimental studies on speech. The absence of UP effects at natural speech rates can therefore not be taken as evidence against the sequentiality of speech processing.

MODEL II

Speech Processing Delays Commitment: The Need for Temporary Storage

Critiques of the Sequential Models

Despite (and also because of) their simplicity, strictly sequential speech recognition models like the Cohort model have not gone unchallenged. The most salient problems are presented below. First, although there is convincing evidence that the signal is processed as it appears, there is also evidence that processing continues beyond the UP (e.g., more than the first four phonemes would be processed in *claustrophobic*). For example, M. Taft and Hambly (1986) obtained shorter lexical decision times for pseudowords such as *mep* than they did for pseudowords such as *mepsig*, although in both items /p/ corresponded to the deviation point (DP)—that is, the point in the pseudoword where the sequence becomes incompatible with any word in the lexicon. This result suggests that speech processing is not a strictly time-shadowing operation, with lexical decisions occurring when enough information is theoretically available. The decision on the identity of a stimulus might require the processing of post-DP (or -UP) segments too (see Goodman & Huttenlocher, 1988, for similar results).

Parallel to these studies, there is a growing body of research moderating the popular assumption, held by the proponents of strictly sequential models (Cole, 1973; Cole & Jakimik, 1980a, 1980b; Marslen-Wilson & Welsh,

1978; Marslen-Wilson & Zwitserlood, 1989), that word onsets are crucial for successful recognition. First, the use of cross-modal priming suggests that distortions of word-initial information do not preclude recognition much more than distortions of noninitial information (Connine, Blasko, & Titone, 1993). Second, late segments of words have been shown to generate important lexical activity in the lexicon, a result not predicted by sequential models. For instance, Slowiaczek, Nusbaum, and Pisoni (1987) observed that the identification of monosyllabic words played in noise was facilitated by prior presentation of a prime sharing one or more phonemes with the target, regardless of the location of the shared phonemes. The role of position in phonological priming has been investigated in many other studies (e.g., Radeau, Morais, & Segui, 1995; Salasoo & Pisoni, 1985; Shillcock, 1990; see Tabossi, 1993, for a discussion).

The relevance of the UP has also been questioned by Marcus and Frauenfelder (1985), who showed that it is possible to obtain UP effects within a system that does not postulate left-to-right processing. These authors found that the position of the UP is confounded with the point in a word after which it starts to differ from the closest nonidentical word in the lexicon (in terms of number of phonemes), regardless of the position of the deviating phonemes. They demonstrated this point by using the concept of minimal deviation. If minimal deviation is defined as the number of different phonemes between the segment of a word (starting from its onset) and the closest word in the lexicon that has at least as many phonemes as the sequence, then all words have a minimal deviation of zero before the UP and of one at the UP. After the UP, the minimal deviation can remain very small for some words (e.g., one for *aptitude* because a word like *altitude* diverges from the base word by one phoneme until its offset), but it can become quite large for others. For example, as shown below, the minimal deviation is four for the word *incestuous*, because the closest competitors, such as *accidental*, are four phonemes away from *incestuous* (the four different phonemes are underlined in each word):

incestuous	<u>i</u> <u>n</u> 's e s t f u ə s
accidental	æ <u>k</u> 's e <u>n</u> t f u ə l

However, Marcus and Frauenfelder (1985) showed that, despite some variation, the average minimal deviation increases in a linear way after the UP. Therefore, as the authors pointed out, the UP is not only the point at which a segment becomes unique through sequential elimination of its competitors, it is also the point after which it increasingly departs phonetically from them. The UP effects found in the literature could thus reflect minimal deviation as much as uniqueness. In such a case, UP effects may unwittingly support the possibility that processing goes beyond the UP, an idea that is not implemented in the Cohort model. In addition, the concept of minimal deviation suggests that words that would the-

oretically not enter the cohort because they do not share the initial phonemes with the input could turn out to be serious competitors (consider *mobility* and *nobility*).

The preceding remarks show that the notion of the UP as a recognition point is problematic. According to the early version of the Cohort model (Marslen-Wilson & Welsh, 1978), the listener simply cannot recognize a word if a pre-UP phoneme has not been properly decoded (e.g., the word *cigarette* could never enter the cohort generated by a mispronounced word such as *shigarette*). Because such misperceptions frequently happen in everyday speech without necessarily precluding recognition, it must be possible for the system to take advantage of "spare" information contained in post-UP segments of words. In order to remedy this drastic dichotomous feature, Marslen-Wilson (1987) designed Cohort II. In this model, variability in the signal quality is no longer an insuperable obstacle to recognition, because the all-or-none cohort membership principle has been replaced with a relative activation principle akin to that postulated in connectionist models (McClelland & Elman, 1986). In Cohort II, a lexical candidate belongs to the cohort to a certain extent, according to its degree of phonetic similarity with the incoming input. As a consequence, recognition becomes an affair of relative activation rather than a cohort's steady reduction to one candidate. Although Cohort II is probably more realistic than its earlier version, it loses its temporal predictability as to the recognition point. Recognition can now occur after the UP, depending on the quality of the input. This loss of predictability is ultimately the consequence of the distance separating Cohort II from strictly time-shadowing decision principles.

Yet, even with a perfect input, the sequentiality of Cohort I and Cohort II theoretically leads to inaccurate recognition a noteworthy number of times. McQueen, Cutler, Briscoe, and Norris (1995) reported that 84% of polysyllabic English words have at least one shorter word embedded within them (e.g., *carbon* contains *car*) and 63% contain more than one embedded word (e.g., *cardinal* contains *card* and *car*). Thus, according to a sequential view, words such as *card*, *carbon*, and *cardinal* remain in the cohort of the isolated word *car* until a pause is heard at its offset. In connected speech, where no reliable cues indicate word boundaries, *car* may need a few phonemes past its offset to be disambiguated.

As a matter of fact, given that short words are by far the most frequent occurrences in English, recognition would seldom be possible before word offset. Luce (1986) showed that, when the words in the lexicon are weighted by their frequency of occurrence, the probability of a word's becoming unique before its offset is .39. In addition, although semantic or syntactic constraints can potentially reduce the ambiguity introduced by embedded words (Cutler, 1993; Cutler & Carter, 1987; Marslen-Wilson & Tyler, 1980; Thorne, Bratley, & Dewar, 1968), they do not eliminate the segmentation problem entirely (McQueen et al., 1995).

Empirical evidence clearly highlights the inability of sequential processing to deal with embeddedness. When subjects are presented with word segments (or "gates") of various lengths in a sentential context, many words (particularly short ones) are not recognized before the onset of the following word (Bard, Shillcock, & Altmann, 1988; Grosjean, 1985). This result indicates that segmentation and word recognition are two variables that are quite interdependent and difficult to tease apart.

Marslen-Wilson (1987) argues that postponed segmentation due to "late selection" does not weaken the essential feature of the Cohort model, which is the sequential nature of the word recognition process. Although it is true that lateness of selection does not undermine the sequential principle *per se*, it may well constitute a problem for the assumption that a decision about the identity of a word is made as soon as enough evidence is available to the system to do so (the time-shadowing decision assumption). In other words, the Cohort model may not be as optimal as was once thought. Because of embeddedness, there may be some interference between the late selection stage of a word and the activation of the cohort of the next one. For example, Tabossi, Burani, and Scott (1995) showed that a lexical hypothesis (e.g., the Italian word *visite* [*visits*]) straddling a target word and the following word (e.g., *visi tediati* [*faces bored*]) produced a priming effect that remained active after the offset of the target word (*visi*). This result raises problems for models that present speech processing as a word-by-word operation, with the processing of one word starting only when the processing of the previous word has been completed. Tabossi et al.'s finding suggests that because of embeddedness, left-to-right lexical processing must be accompanied with a less time-linked procedure in order to reach successful word segmentation and recognition.

The Principle of Delayed Commitment

The problem of embeddedness can be overcome if the decision about word identity is delayed until sufficient information is available. Such a procedure, known as delayed commitment, was recently evaluated by McQueen et al. (1995). These authors suggested that cohorts are generated at any moment in the signal so that the lexical hypotheses span different portions of the input with many of them straddling word boundaries. For example, in the sequence *can it*, word candidates such as *can*, *canid*, *canister*, *annual*, and *anecdote* are activated and considered in parallel (this process is known as "multiple alignment") until information present in the subsequent word disambiguates the signal. In this example, only the occurrence of the segment /it/ reduces the activation of all competitors. The recognition of *can* is thus delayed until after its offset. In this type of model, the time-shadowing decision feature that was present in the Cohort model is replaced by delayed commitment such that, as McQueen et al. (1995) put it, "words uttered later in time can influence the recognition of those uttered earlier" (p. 310).

Some current computational models of speech recognition such as the interactive TRACE model (McClelland & Elman, 1986), the bottom-up Shortlist model (Norris, 1994), and even Cohort II (Marslen-Wilson, 1987) postulate multiple alignment and delayed commitment. However, two points on which these models differ are (1) the number of lexical hypotheses activated at any one time, and (2) whether or not there is lateral inhibition (i.e., inhibitory connections) between the lexical hypotheses. The number of words allowed to enter the competition is kept relatively small in Shortlist compared with TRACE and Cohort II, thanks to severe matching restrictions between the input and the candidates. It nevertheless remains considerably larger than in the purely sequential Cohort I as a result of multiple alignment. Concerning the competition mechanism between the activated candidates, only TRACE and Shortlist propose the principle of lateral inhibition. All candidates activated at one point in time compete with each other so that the candidates that best match the input decrease the activation level of the others. In Cohort II, competition is passively accomplished through the analysis of the goodness-of-fit between the candidates and the acoustic-phonetic information.

Models that postulate delayed commitment implicitly adopt a very conservative position in the debate between speed and accuracy in word recognition. The delayed commitment in word identification gives importance to accuracy. Recognition is put off until the entire input is segmented without ambiguity. One of the implications of this processing delay is that temporarily unused information has to be kept active until full disambiguation is achieved. The price for such delayed commitment is increased memory storage space.

Possible Memory Stores

According to Nootboom (1979), two types of memory could be used to hold acoustic-phonetic information while the lexical candidates compete for word identification: a very short-lived physical-acoustical buffer and a more perceptual-auditory short-term memory.

The physical-acoustical buffer has been referred to in a number of ways in the literature. For example, Nootboom (1979) talks about a "gliding time window" whose aperture defines the capacity of the "acoustic buffer." This acoustic buffer is (1) faithful to the physical input—that is, its contents are acoustic and not phonetic, (2) very brief; and (3) largely determined by the neurophysiological properties of the auditory system. The agreed-upon temporal capacity of the acoustic buffer seems to be about 250 msec (Brokx, 1979; Butcher, 1973; Huggins, 1975; Massaro, 1974; Plomp, 1964; Slis & van Nierop, 1970). This estimate roughly corresponds to the average duration of a syllable.

A similar description of the physical-acoustical buffer can be found in Huggins's (1975) *echoic memory* (slightly different from the echoic store proposed by Crowder and Morton, 1969, which will be discussed later). Huggins argues that the need for echoic storage is obvious in au-

dition because of the evanescent nature of the input: "sounds travel past the observer at 1,100 ft/sec, so there is no possibility of taking a second look, as there often is in vision" (p. 212). Furthermore, one of the functions of echoic memory is to integrate ongoing sounds into a continuous percept by "bridging the gaps" (e.g., pauses, extraneous noises) between the successive segments of speech. However, this "gap-bridging" process has only a limited span. Huggins observed that, when silent pauses of 125 msec or more were inserted in the speech signal at regular intervals, intelligibility was dramatically reduced. The reduction in intelligibility was interpreted by the author as the echoic memory's failure to link the too dramatically disconnected speech segments. This breakdown could in fact be related to both speech rate and absolute duration. Indeed, data showed that the intelligibility curve plotted against various silence durations was different with speeded speech from what it was with slowed speech (Huggins, 1975). Specifically, it appeared that the critical parameter controlling intelligibility was not the duration of silence alone, but rather the time between the start of a speech segment and the start of the next. This result indicates that the capacity of echoic memory may not be exclusively time related. Instead, it might be a combination of time and amount of information (see also Huggins, 1964; Wingfield & Wheale, 1975).

The composite nature of the physical-acoustical buffer appears much the same way in studies of *preperceptual auditory images* (Efron, 1970a, 1970b, 1970c; Massaro, 1972a, 1972b, 1974). Massaro argues that an auditory signal leaves an image, held in an acoustical store, that can persist after the end of the stimulus. An important function of this image is to assist recognition of the signal. After readout of the image, which occurs within approximately 250 msec, an almost total decay takes place. The relationship between a sensory input and its subsequent image is complex. For example, Efron showed that the minimal perceived duration of a tone or a noise burst was about 130 msec. If the stimulus was shorter than 130 msec, its auditory image combined with the actual signal to create the impression of a 130-msec stimulus. This finding indicates that, with very short presentations, the auditory image does not differ qualitatively from the stimulus itself. Thus, research on preperceptual auditory images is compatible with Huggins's (1975) data on echoic memory in that it suggests that the temporal capacity of the physical-acoustical buffer may not be fixed but, instead, intrinsically tied to the stimulus duration.

Finally, it should be noted that the contents of the physical-acoustical buffer may have a phonetic component in addition to the purely acoustical nature usually ascribed to it. In their dual coding model, Fujisaki and Kawashima (1969, 1970) proposed that two types of memories were serially consulted in phoneme perception: a phonetic memory and an auditory memory. The phonetic memory would encode the discrete phonetic label of the input (e.g., /k/, /æ/, /t/) whereas the auditory memory would realize a more fine-grained analysis in order to discriminate phonemes within the same pho-

netic category. Insofar as the dual memory system is linked to the physical–acoustical buffer, the data obtained by Fujisaki and Kawashima could suggest that the information maintained in the physical–acoustical buffer during delayed commitment procedures may have both a phonetic and an acoustic format.

As Nooteboom (1979) pointed out, the physical–acoustical buffer must not be confused with the more perceptual “auditory short-term memory,” although they both qualify for temporary storage in models that postulate delayed commitment. Auditory short-term memory has a wider span (up to several words) than the physical–acoustical buffer has, its contents are principally phonetic, and its decay time is variable. The capacity of short-term memory seems to be information related rather than time related (G. A. Miller, 1956; Simon, 1974).

In their theory of short-term memory, Crowder and Morton (1969) describe a type of “echoic” memory, called precategorical acoustical storage (PAS), that has a capacity of a few seconds and is, in its original account, only auditory. There is now evidence that this store may not be restricted solely to audition. This possibility is supported by the finding, in different modalities, of the suffix effect, the standard empirical evidence for PAS. This effect can be observed by comparing the serial recall performance of a list of several items presented auditorily with that of the same list followed by a “suffix”—that is, an additional item that must not be remembered. The suffix effect is indicated by a substantial decrease of the traditional recency effect (i.e., the performance peaks on the last two or three items) due to the presence of the suffix (Crowder, 1967). Later studies of this phenomenon have revealed that a silent lip-read suffix can produce comparable interference (Spoehr & Corin, 1978). Similar results were also obtained with an auditory suffix preceded by a lip-read list (Campbell & Dodd, 1982; Gathercole, 1987; Greene & Crowder, 1984) or by a written list silently mouthed (Greene & Crowder, 1984; Nairne & Crowder, 1982; Nairne & Walters, 1983). These results show that PAS does not have a purely auditory/acoustic basis but, instead, may share some components with seen and mouthed speech (Crowder, 1983). Thus, the compound nature of PAS could indicate the presence of multilevel representations within the storage buffer used in lexical processing.

Finally, the idea that the memory store has a multifaceted organizational structure is also supported by recent findings on the role of voice information and its retention on spoken word recognition. For example, Palmeri, Goldinger, and Pisoni (1993) have shown that the non-phonetic details of a spoken word resulting from the specificities of the utterer’s voice are retained for several seconds—sometimes up to a week (Goldinger, 1996)—and are used to assist in word recognition during that period. Although it is still unclear whether voice information and phonetic information are stored and retrieved together, this finding suggests that even short-term auditory memory may contain surface features of the speech signal.

MODEL III Speech Processing Goes Back in Time: The Need for a Processing Window

In the preceding section, I described how delayed commitment models deal efficiently with embeddedness by postponing recognition after word offset and by including a memory store in which ambiguous information is held. In those models, segmentation emerges as a by-product of word recognition; word boundary ambiguities are solved by semiexhaustive (e.g., Shortlist) or fully exhaustive (e.g., Cohort II and TRACE) activation of lexical candidates at any point in the signal. This “serendipitous” segmentation principle (Cutler, 1996) is not the only way in which segmentation can theoretically be achieved. For example, if some indicators of word boundaries could be detected in the signal itself, segmentation could be done explicitly, *prior to* lexical activation. Segmentation would thus not be the result of recognition but, instead, its starting point. Lexical access could be reset at every identified boundary and recognition achieved at word offset.

Speech scientists have attempted to pinpoint several markers for word junctures: pauses, fundamental frequency contours, bursts, aspiration, glottal stops, laryngealization, and allophonic variations (e.g., Garding, 1967; Lehiste, 1960; Nakatani & Dukes, 1977). However, none of them has proved stable or reliable (Klatt & Stevens, 1973; Lehiste, 1972; Reddy, 1976).

Since then, the search for word boundary cues has shifted to the prosodic arena. Studies in this domain have revealed that the lexical stress of words is a fairly reliable index of word boundaries (Cutler & Norris, 1988; Nakatani & Schaffer, 1978; Nooteboom, Brokx, & de Rooij, 1978). Words starting with a stressed syllable (i.e., containing a full vowel) appear to be the rule in English. When their frequency of occurrence is taken into account, about 90% of lexical (content) English words are found to start with a stressed syllable (Cutler, 1989; Cutler & Carter, 1987).² Therefore, a segmentation strategy that postulates a word boundary before every stressed syllable seems to be, in theory, an effective solution to deal with speech continuity. With such a procedure, the input would be mostly processed in a time-shadowing fashion. With initial-stressed words, a cohort of candidates would be generated on the strong syllable—the putative onset of the word—and would be steadily reduced until the uniqueness point or the next stressed syllable was reached. The word identity decision would thus be made at word offset, at the latest.

The obvious drawback of such a stress-based strategy is that about 10% of the words encountered daily would be segmented, and hence accessed, erroneously. Yet, in real life, non-initial-stressed words such as *appear* or *fantastic* do not seem to cause tremendous difficulty for listeners. Thus, stress-based models need to include a routine able to deal with such an input. In non-initial-stressed words, speech processing may not be entirely sequential. It may go back in time through some repair operation,

such as backtracking, in order to remedy the incorrect segmentation hypothesis. For such a backtracking mechanism to work properly, the system would have to include a memory space in which the unstressed syllable(s) preceding the stressed syllable could be stored.

In this section, I first review the evidence that stressed syllables constitute valid candidates to segment speech and initiate lexical processing sequentially. I then present support for the existence of some backtracking mechanism that allows the processing of initial unstressed syllables to be influenced by a later occurring stressed syllable. In this description, the notion of a processing window is central, because it defines the space in which both proactive (sequential) and retroactive mechanisms are completed.

Stressed Syllables Initiate Sequential Processing

The intelligibility of stressed syllables is a well-established fact in phonetics as well as in cognitive psycholinguistics. A quick overview of the literature suggests that stressed syllables have the following features:

Physical salience. Stressed syllables show salience in terms of pitch, duration, and amplitude (Chomsky & Halle, 1968; Fry, 1955, 1958; Lehiste, 1970).

Phonemic stability. Stressed syllables are more stable than unstressed syllables, in the sense that they are less sensitive to phonological modification. For instance, while the realization of unstressed vowels varies greatly depending on the type of English (e.g., colloquial speech vs. citation form), stressed vowels usually remain unaffected (Brown, 1977; Gimson, 1989; Shockey, 1983). The higher stability of stressed syllables makes them more efficient in partitioning the lexicon (Altmann & Carter, 1989; Huttenlocher & Zue, 1983).

Perceptual distinctiveness. Stressed syllables are less likely to be misinterpreted than unstressed syllables (Browman, 1978); they are detected more consistently than unstressed syllables in noisy environments (Kozhevnikov & Chistovich, 1966) and in words excised from fluent speech (Lieberman, 1965); hearers can locate clicks more accurately in stressed syllables than in unstressed syllables (Bond, 1971); phoneme detection is faster in stressed syllables than in unstressed syllables (Cutler & Foss, 1977); mispronunciations are detected more rapidly in stressed syllables than in unstressed syllables (Cole & Jakimik, 1978, 1980a); stressed syllables are misperceived less often than unstressed syllables (Bond & Garnes, 1980); they are also less fluently restored than unstressed syllables in shadowing tasks (Small & Bond, 1982). Overall, these studies show that stressed syllables consist of perceptual "islands of reliability" (Pisoni, 1981) that could be used as anchors in the recognition process.

In addition, research has shown that the preference for stressed syllables could also originate from their capacity to draw the listener's attention to the rhythmic deployment of speech (Allen, 1967; Cutler, 1976; Gow & Gordon, 1993; Huggins, 1972a, 1972b; Martin, 1972; Pitt & Samuel, 1990; Shields, McHugh, & Martin, 1974). The temporal regularity of the alternation between stressed

and unstressed syllables could cause the listener's attention to lock into this rhythm and jump from one stressed syllable to the next (Meltzer, Martin, Mills, Imhoff, & Zohar, 1976). This hypothesis, known as the *attentional bounce hypothesis*, was tested by Shields et al. (1974). These authors found that the phoneme /b/ starting a bisyllabic pseudoword was detected faster when it was predicted to be in a stressed syllable by the preceding sentence than when it was not. This effect was attributed to the fact that, when the rhythmicity of speech is predictable, particular attention is given to stressed syllables which, as a consequence, yield better detection performances. The possibility that only acoustic factors associated with stress were responsible for this effect was later discarded in a similar experiment by Pitt and Samuel (1990), in which the acoustic cues to stress were digitally eliminated. Moreover, Pitt and Samuel (1990) found that the facilitation for an expected stressed syllable was also present (although weakly) when the target was in the stressed syllable of a second-syllable-stressed bisyllabic word, a result that extends Shields et al.'s findings beyond just the initial syllable.

Studies on attention clearly demonstrate that stressed syllables have a particular status that goes beyond their local physical salience. Through prosody, the speech processor is able to anticipate the occurrence of upcoming stressed syllables and to focus its attention on them. There have been several attempts to establish the origin of such prosodic attentional effects. One view suggests that the expectancy for stressed syllables could originate from a higher level of representation going beyond the time variable. The processing of rhythmicity could derive from a hierarchical organization of stressed and unstressed syllables defining a rhythmic grid based on the foot (a foot includes a stressed syllable plus the following unstressed syllables). This organization would stretch over the entire input, from the syllable to the sentence as a whole (Hayes, 1981; M. Y. Liberman, 1975; M. Y. Liberman & Prince, 1977; Selkirk, 1980, 1984). A problem with this type of conceptualization is that, owing to its wide span, the layered structure could not be disentangled fast enough during on-line processing (Buxton, 1983; M. Y. Liberman & Prince, 1977).

By contrast, a model such as Martin's (1972), which gives more importance to the temporal factor than to the hierarchical structure, could account for stressed syllable expectancies during on-line processing while keeping an emphasis on prosody. Martin argues that temporal and rhythmic factors are related in such a way that the perception of salient temporal events like stressed syllables generates expectancies concerning later events such as the position of the next stressed syllable or the quality of the signal between two stressed syllables. In this account, stressed syllables are central during speech processing because they are "articulatory targets of ballistic movements" (p. 500) that can be spotted by the listener as the momentum of the utterances. Speech processing would thus proceed in two stages. First, the rhythmic pattern of the message would be identified through stressed

syllables. This operation would reveal a full outline of the prosodic contour of the ongoing utterance. Second, this analysis would permit rapid, because anticipated, segmental analysis.

The evidence reviewed so far suggests that stressed syllables have a particular status in the speech signal, either because they acoustically leap out of the signal or because they form a perceptually reliable rhythmic cue for attentional processes. The next step is to examine the evidence that stressed syllables are used to segment speech and initiate lexical processing.

Given that stressed syllables often start English content words and that they are salient in many respects, they would theoretically make excellent triggers for accessing the lexicon. Along the lines of the attentional studies reported above, Gow and Gordon (1993) obtained results suggesting that, among stressed syllables, those starting a word are particularly attended to. The authors presented subjects with sentences that carried a target syllable (e.g., *sub* or *ject*) belonging to either a stressed–unstressed word (e.g., *SUBject*) or an unstressed–stressed word (e.g., *subJECT*). The sentence was manipulated in such a way that, in one case, either the stressed–unstressed or the unstressed–stressed version of the carrier word was expected. In the other case, the stress pattern of the carrier could not be inferred by the sentential context. The results revealed that, with both unambiguous and ambiguous sentential contexts, the subjects were faster at detecting stressed than unstressed syllables in the word-initial position. The reverse pattern was observed in the second syllable of the word.

At first glance, these results contradict Pitt and Samuel's (1990) finding that the detection of a syllable-initial consonant is facilitated when the syllable is predicted to bear stress, regardless of whether stress is predicted in the first or the second syllable. However, with careful scrutiny, the asymmetry between first- and second-syllable-stressed words can be shown in this study too. Pitt and Samuel (1990) found that detection was reliably faster (24 msec) in syllables predicted to receive stress than in syllables predicted not to receive stress, in both first- and second-syllable-stressed words.³ There was no significant interaction between the predicted position of the stress and the position of the target. However, an inspection of the effect of stress predictability reveals a 36-msec difference in the initial position and only a 12-msec difference in the final position. Thus, preference for initial stressed syllables over noninitial ones is, at least at a descriptive level, apparent in these results too. In conclusion, Gow and Gordon's (1993) findings, coupled with complementary data in Pitt and Samuel's (1990), suggest that the attentional preference for stressed syllables cannot be accounted for by rhythmic predictability alone. Lexical factors must be considered as well in order to explain positional differences.

Direct evidence for the role of stress in lexical processing has also been provided by a number of experiments using words in isolation. For example, Cutler and Norris (1988) had their subjects detect a monosyllabic

word (e.g., *mint*) starting a bisyllabic pseudoword. One version of the pseudoword bore two stressed syllables (e.g., *mintayf* /'mɪn'tef/), whereas the other version had an unstressed second syllable (e.g., *mintef* /'mɪntəf/). The authors believed that, if strong syllables were used to initiate lexical processing, then an item like *mintayf* would trigger two cohorts of candidates, one starting with /'mɪn/ and one starting with /'tɛf/. Therefore, the detection of *mint* in *mintayf* would be hampered because the /t/ would be attributed to the second cohort and not to the first one. Such segmentation interference would not happen with *mintef*, because the unstressed syllable /təf/ would not activate any cohort. This is exactly what Cutler and Norris (1988) found. Detection of *mint* was slower in *mintayf* than in *mintef*. This effect was absent when the target word did not straddle the putative segmental hypotheses (e.g., *thin* in *thintayf* or *thintef*). This finding has been invoked extensively to support the idea that metrical prosody is used to segment the speech stream. This hypothesis has come to be known as the metrical segmentation strategy (MSS).

The listener's aversion to word-initial unstressed syllables is a complementary implication of MSS. Empirical evidence attests to this possibility. For example, when L. A. Taft (1984) presented subjects with sequences such as /'lɛtəs/ and /ɪn'vests/, the preferred interpretation was one that did not allow an unstressed syllable to be word initial. Subjects tended to choose *lettuce* and *in vests* rather than *let us* and *invests*, respectively. Similarly, Cutler and Butterfield (1992) showed that over two thirds of the missegmentations of continuous speech found in corpora of spontaneous "slips of the ear" (speech misperception) and over 90% of experimentally induced lexical juncture misperceptions resulted from the erroneous insertion of a word boundary before a stressed syllable.

From a temporal point of view, MSS suggests that in initial-stressed words, lexical processing fruitfully proceeds in a left-to-right manner, from one stressed syllable to the next. Such a rhythmic operation would be dramatically disrupted if the acoustic correlates to stress were not correctly picked up by the listener. Mis-stressing words is indeed shown to significantly affect recognition (Bansal, 1966; Bond & Small, 1983; Cutler & Clifton, 1984; Slowiaczek, 1987, 1990), which is an indication that stress is used at an early stage of lexical processing to specify a set of potential candidates. Such a view can already be found in Brown (1977), who argued that the presentation of a word with a given stress pattern activated only the words in the lexicon that bore this pattern (see also Connine, Clifton, & Cutler, 1987).

However, these results are at odds with two studies. First, Cutler and Clifton (1984) observed that prior knowledge of the stress pattern of a word did not facilitate lexical decision performed on this word, a result that contradicts the assumption that prosodic information is used directly for word recognition. Second, Cutler (1986) found that the stress pattern of activated candidates bore no relationship to that of the input from which they were generated. In a cross-modal priming experiment, Cutler

showed that the meanings of differently stressed but segmentally identical words (e.g., *DIScount* and *disCOUNT*) were activated equally when either version of the word was presented. Selective priming occurred only when the target was presented 750 msec after the offset of the prime rather than immediately after it. This result suggests that lexical stress, at least in the initial stage of processing, is not taken into account to guide lexical processing. Cutler argued that there was actually little reason for the system to include lexical stress in the access code to the lexicon, given that segmentally identical but suprasegmentally distinct pairs of words are rare in English. However, this conclusion may be restricted to the contrast between primary and secondary stressed syllables—the only distinction examined in Cutler's study—which both bear a full vowel. It is possible that the computation of stress during lexical processing involves only the more salient alternation between full vowels and reduced vowels or schwas.

The distinction between lexical prosody (which assigns primary, secondary, and no stress) and metrical prosody (which has only two levels: stressed and unstressed) might be important for determining the level at which stress intervenes in word recognition. Perceptual differences between these two types of prosody have been reported in a number of studies. For example, Lieberman (1965) found that when subjects were asked to identify the different stress levels of an utterance in which the segmental information had been removed without affecting the suprasegmental contour, they could distinguish between stressed and unstressed syllables (metrical prosody) but not between primary and secondary stressed syllables (lexical prosody). Nakatani and Schaffer (1978) obtained similar results with "reiterant" speech in which the prosody of words was mimicked by substituting a nonsense syllable, such as /ma/, for each syllable in the original sentence (e.g., 'Ma''ma 'ma ma 'mama 'ma for *Mary had a little lamb*).

The possibility that only metrical distinctions are computed in the course of word recognition also finds support in research on read versus spontaneous speech. When McAllister (1991) presented subjects with short segments of words carefully read by naive speakers, recognition of initial-stressed words was not significantly different from recognition of non-initial-stressed words. However, when these words were excised from spontaneous speech recorded from the same speakers, recognition was better for initial-stressed words than for non-initial-stressed words. Apart from the fact that the words in the spontaneous speech condition originated from sentences whereas those in the read speech condition did not, it is likely that the former presented essentially an alternation between full vowels (in stressed syllables) and reduced vowels (in unstressed syllables), whereas the latter, akin to citation form, had a greater number of full unstressed syllables and, thus, was more likely to exhibit the three levels of stress found in lexical prosody. This difference could explain why stress did not have the same impact on the processing of read and spontaneous

speech (see Mehta & Cutler, 1988, for a similar interpretation). The discrimination between finer grained prosodic features, such as secondary stressed vowels and unstressed full vowels, could intervene at some later stage of processing.

The question of whether or not stress (or some of its features) is encoded prelexically concerns primarily the nature of cohort members. Eventually, this debate should allow us to determine how the stress pattern of a given set of activated candidates relates to the prosody of the input from which they are generated. Although insight in this matter is critical to making accurate predictions about recognition latencies, stress encoding issues do not interfere with the hypothesis that stressed syllables assist segmentation and initiate lexical processing. Today, an important body of research supports the possibility of a procedure similar to the MSS (e.g., Bradley, 1980; Cutler, 1976; Grosjean & Gee, 1987; Mattys & Samuel, 1997; McQueen, Norris, & Cutler, 1994; Nakatani & Schaffer, 1978; Nooteboom et al., 1978; Norris, 1994; Norris, McQueen, & Cutler, 1995; Vroomen & de Gelder, 1995). Overall, these studies reveal that the speech processor makes the hypothesis that stressed syllables are the beginning of words.

Wrong Start: The Need for Backtracking and for a Processing Window

With non-initial-stressed words, stress-based models fail to account for how the signal is segmented and how the correct lexical representation is accessed. A number of authors recognize that some special mechanism must be used with this kind of input (Bradley, 1980; Cutler, 1976, 1989; Gow & Gordon, 1993; Grosjean, 1985; Grosjean & Gee, 1987). For example, Gow and Gordon (1993) state that

stressed syllables in second position should cause inappropriate segmentation, treating the second syllable of a target-bearing word as the first syllable of a new word. As this would generally not match any entries in the lexicon, it would delay lexical access until the segmentation problem could be corrected through processes guided by non-stress information. (p. 559)

Similarly, when talking about the initiation of lexical access through stressed syllables, Cutler (1989) argues that

the strategy will not locate the onsets of words beginning with weak syllables—*appear, begin, succeed*, and the like. However, it is at least conceivable that the strategy of treating strong syllables as if they were onsets is supplemented by an ancillary strategy whereby lexical words beginning with weak syllables can, under appropriate circumstances, be successfully accessed via their strong syllables. (p. 353)

Implicitly, stress-based models postulate that the speech processor has the capacity to use backtracking procedures in cases for which lexical access through strong syllables is unsuccessful. Thus, in addition to the traditional left-to-right mechanism triggered by stressed syllables, the listeners would also use right-to-left proce-

dures. This account is far from the conventional sequential account of lexical processing in which lexical candidates are activated in a temporally ordered manner as the acoustic input unfolds. Here, stressed syllables are essentially the milestones of the listener's attention, and although they are processed from left to right as the sentence is heard, they each open a processing window in which the surrounding unstressed syllables are decoded through proactive and retroactive analyses. The basis for such a view of lexical processing has been proposed by Grosjean and Gee (1987). In their model, the speech input is parsed into stressed and unstressed syllables. Strong syllables serve as the access code to the lexicon. The activated candidates, in turn, help to identify the unstressed syllables on either side of the stressed syllable, thus creating both proactive and retroactive processing flows.

There is, so far, little empirical evidence that the speech processing system repairs faulty lexical segmentation by backtracking to information presented earlier in time. However, indications of retroactive procedures in lexical processing are found in Cluff and Luce (1990) and Mattys and Samuel (1997). Cluff and Luce examined subjects' identification of spondees in white noise. Spondees are bisyllabic compound words like *chestnut* or *deadlock* that bear two stressed syllables. The authors presented their subjects with four categories of spondees. In one category, the spondees were made of two "easy" monosyllabic words (high in frequency, with few phonetic neighbors [e.g., *deadline*]). In another, the monosyllables were both "hard" (low in frequency, with many phonetic neighbors [e.g., *beehive*]). In the two other categories, spondees were either "easy-hard" (e.g., *turnpike*) or "hard-easy" (e.g., *hacksaw*). Cluff and Luce found that "easy-easy" spondees produced the highest recognition scores, whereas "hard-hard" produced the lowest ones. They also found that the subjects' performance was influenced by the position of the hard and the easy syllables in the spondees. Specifically, "hard-easy" spondees yielded better identification than did "easy-hard" ones. This result was taken by the authors as an indication of retroactivity; ambiguity in a first hard syllable would be assisted by information coming later in time. The proactive pattern—an easy first syllable helping a hard second syllable—was not shown in the data. These results suggest that, under conditions of lexical uncertainty, retroactive procedures can occur in lexical processing. They also indicate that the speech system has the capacity to maintain the temporally unprocessed part of the input in memory until later information provides disambiguation.

As Cluff and Luce (1990) pointed out, although these results do not directly bring support to stress-based models, they suggest that primacy is given to easy speech segments (or salient ones, such as stressed syllables) that, in turn, assist the processing of harder segments (like unstressed syllables). This view is very similar to that presented by Grosjean and Gee (1987), except that Cluff and Luce found evidence for retroactive procedures only, whereas Grosjean and Gee hypothesized that stressed

syllables assisted identification of unstressed syllables both retroactively and proactively. Cluff and Luce suggested that this asymmetry could originate from the system's preference for solving ambiguity on already presented material rather than predicting upcoming material. The risk of making garden-path interpretations would be lower in the former than in the latter case.

Evidence for retroactive processing is also found in Mattys and Samuel (1997) with the *migration paradigm* (Kolinsky, 1992; Kolinsky, Morais, & Cluytens, 1995). In their experiment, subjects were asked to detect an auditory target (e.g., *contro•versy*) in a pair of items presented simultaneously over headphones. On target-absent trials, one of the items of the pair differed from the target by one vowel (e.g., *kīn•tro•ver•sy*). The other item either carried the missing vowel at a corresponding location (e.g., *bōs•glo•ra•fe*), the experimental condition, or it carried an unrelated vowel (e.g., *bās•glo•ra•fe*), the control condition. The dependent variable was the migration rate of the vowel, computed as the difference between the misperception of the target in the experimental condition and in the control condition.

Mattys and Samuel (1997) showed that, when the items of the pairs were presented dichotically, the migration rate was higher with pseudoword targets than with word targets: lexical targets resisted intrusions better than nonlexical targets did. However, when the items of the pairs were played in both ears at different amplitudes so that they were perceived to be closer to each other than in the dichotic situation, the lexical effect on migration remained only when the mispronounced vowel was in an unstressed syllable, regardless of its location in the word. Thus, when the stressed syllable of a target was intact, the vowels of its unstressed syllables resisted migrations better if the target was a word than if it was a pseudoword. In contrast, when the unstressed syllables of a target were intact, the vowel of its stressed syllable did not resist migrations better if the target was a word than if it was a pseudoword. The results were interpreted by the authors as showing the importance of stressed syllables in accessing the lexicon. In their view, the lexicon is accessed through stressed syllables, and the information that comes back from this stress-based lexical search is used to assist the identification of unstressed syllables (before and after the stressed syllable) and, hence, avoid migrations. These findings are very consistent with the model elaborated by Grosjean and Gee (1987), in that they emphasize the need to include both proactive and retroactive routines in the speech processing system.

It could be argued that stress-based models have difficulty accounting for a robust phenomenon that has been consistently taken as evidence for strict sequentiality—namely, the uniqueness point effect (i.e., early-unique words are recognized faster than late-unique words). In fact, it can be demonstrated that, in theory, the uniqueness point effect is not totally incompatible with stress-based models. First, with initial-stressed words, UP effects are just as logical in sequential models as they are in stress-based models. The cohort initiated on the first

syllable is expected to drop down to one candidate sooner in early-unique words than in late-unique words, therefore producing the classic UP effect. Second, non-initial-stressed words whose UP occurs before the stress in early-unique words and after the stress in late-unique words are also expected to generate a UP effect. Indeed, although lexical access would be triggered at the same location in both cases (i.e., on the non-initial-stressed syllable), only early-unique words would be recognized right after the stressed syllable. Late-unique words would not be recognized until their actual UP. This difference would be due to the fact that the stressed syllable of early-unique words would generate a cohort of candidates in which only one would be compatible with the beginning of the word. The cohort would thus shrink at once. In contrast, the cohort generated by the stressed syllable of late-unique words would need to reach the UP, past the stress, to isolate a single candidate.

In this description of UP effects, there is only one case that cannot be accounted for by stress-based models. When a noninitial stress falls after the UP of both early-unique and late-unique words, stress-based models do not predict any recognition difference. For both types of word, the cohort activated by the stressed syllable would contain only one candidate that matches the input. Recognition would thus be instantaneous in both cases. The several cases described above could certainly be tested empirically.

As was suggested in the section on delayed commitment models, any system that does not proceed strictly sequentially must include some memory store in which the temporarily unprocessed information can be stored and retrieved later. In stress-based models, a copy of the unstressed syllables would be kept in a memory buffer and retrieved during retroactive processing.

To understand how the buffer-supported retroactive mechanism may be implemented computationally, it is important to realize that the processing of unstressed syllables *results from* that of stressed ones and hence might be distorted (at least in the first pass) to fit the lexical hypotheses generated by the strong-syllable-based search. Martin (1972) pointed out such "negligence" toward unstressed syllables in both speech production and perception, while Cutler and Butterfield's (1992) data on slips of the ear revealed how greatly unstressed syllables could be misperceived to conform to the postulated initial-stressed words. In this conceptualization, the stress-based strategy is only a raw mechanism heavily driven by top-down expectations which sometimes generates erroneous lexical hypotheses. Retroactive processing constitutes a way to undo these errors. When the stress-based search does not return any lexical solution, lexical access is attempted using the unstressed syllable (stored in the buffer) immediately preceding the stressed syllable. If this unstressed-syllable search does not yield recognition (as would be the case in the third-syllable-stressed word *panorama*), lexical access is attempted with the

unstressed syllable preceding the one used in the last search. Although such a retroactive repair function can proceed for several steps in this fashion, it theoretically stops when an unstressed-syllable-based search returns a lexical hypothesis compatible with the input. In principle, this point corresponds to the offset of the last word correctly recognized.

To some extent, the proactive/retroactive mechanism resembles visual processing in which the gaze proceeds by saccade and regression. The only difference is that, with speech, attention does not jump back to information previously displayed, but rather jumps back in time, searching the memory buffer. This memory buffer is not simply a temporary store in which the information on unstressed syllables is kept available. It is more dynamic, in the sense that ordered (proactive and retroactive) operations take place inside of the window defined by the memory buffer.⁴ Unstressed syllables on both sides of the stressed syllable are subject to active identification procedures. For example, as hypothesized by Grosjean and Gee (1987) and empirically supported by Mattys and Samuel (1997), top-down lexical information could influence the processing of unstressed syllables inside of the processing window.

We have already seen in the preceding section that there is an abundant literature on two types of memory storage that could satisfy the demands of delayed commitment models: a physical-acoustical buffer and short-term memory. They both seem to qualify as well to perform retroactive operations. Specifically, it has been shown that the physical-acoustical buffer is far less sensitive to temporal confusion for speech sounds than for non-speech sounds. When nonverbal stimuli are repeated in a loop, subjects need at least 300 msec of each stimulus for accurate perception of their temporal order (Warren, Obusek, Farmer, & Warren, 1969), whereas 125 msec are sufficient to complete a similar task with vowel sounds (Thomas, Hill, Carroll, & Bienvenido, 1970). A solid memory trace for temporally ordered speech events is crucial for analyzing the contents of the processing window adequately. How adaptive would a system be that located word boundaries successfully but did not assemble syllables in a correct order? Nootboom (1979) suggested that the auditory information stored in memory usually persists long enough to create "perceptual reversals," in which the perception of a segment of speech is influenced by a later occurring segment. For instance, Remington (1977) demonstrated that in a CVC syllable, although the initial consonant precedes the vowel, recognition of the vowel can occur before recognition of the initial consonant. Even though this reversal is not due so much to the functioning of auditory memory as to the way speech sounds are coarticulated (see Studdert-Kennedy, 1975, for a review), it suggests that the phenomenological experience of the temporal deployment of speech and the actual reality can sometimes show important discrepancies.

THE USE OF TIME AS A LANGUAGE-DEPENDENT FACTOR

Most of the results described in this review come from experiments carried out in English. However, for sequential models (Model I), which by definition rely on a language-independent mechanism, some results can be extended to all languages. Thus, for models like the Cohort model, the use of time during lexical processing is universal; sounds heard first are used first to access the lexicon, regardless of the language. Yet, delayed commitment models (Model II) and stress-based models (Model III) have brought forth evidence that the timing of lexical processing may depend on the structure of the lexicon (e.g., pure sequentiality breaks down with words that do not have a uniqueness point before their end) and on some phonological properties specific to the language (e.g., stressed syllables initiate lexical access in English). In particular, the proponents of stress-based models argue that processing primacy is given to salient information more than to first-coming information. This claim implies that since saliency may vary from one language to another, the use of time during lexical processing may be language dependent.

In English, the need for backtracking via auditory memory stems from a very good, but not perfect, segmentation strategy (the stress-based strategy). Backtracking would probably not be necessary if the speech processor had totally reliable cues to segmentation. This possibility could be tested by examining the temporality of lexical processing when segmentation ambiguities are suppressed. For example, one could compare the ratio of proactive and retroactive strategies, assessed with the migration paradigm, in a sample of continuous speech and in the same sample in which words have been separated by short pauses or any artificial word boundary indicators. If retroactive processing is a by-product of imperfect stress-based segmentation, it should be found to a lesser extent in the interrupted version than in the natural version. In fact, having artificial word boundaries does not seem to reduce retroactive processing a great deal, because when words are presented in isolation (i.e., with clear word onsets and offsets), retroactive effects are still present (Cluff & Luce, 1990; Mattys & Samuel, 1997). However, this result does not rule out the hypothesis that retroactive processing is the consequence of an imperfect metric-based segmentation. The stress-based strategy could be a mandatory process originating from prolonged exposure to English. Presenting English speakers with words in isolation may not be enough to prevent them from using this strategy.

In order to assess the relationship between the timing of lexical processing and the availability of prosodic cues for word boundaries, it may be informative to examine languages in which word boundary cues are totally reliable in a natural way. For example, in Hungarian, as in Finno-Ugric languages in general, primary stress is always assigned to the initial syllable of content words. A secondary stress is assigned to alternating syllables after

it (Archibald, 1993). Provided that the speech processor can exploit this prosodic cue to segment speech, such a fixed-stress system makes segmentation problems theoretically nonexistent. The speech processor would locate every primary stressed syllable in the speech stream and use it to initiate a lexical search. Lexical processing would essentially proceed proactively because each lexical attempt would be successful by definition. No additional strategy like backtracking would be necessary. This hypothesis certainly calls for empirical testing.

What this example suggests is that if we accept that the use of time during lexical processing depends largely on how word boundaries relate to prosodic cues, it may vary from one language to another. A growing number of studies carried out in Dutch (de Gelder & Vroomen, 1994; Vroomen & de Gelder, 1994, 1995), French (Banel & Bacri, 1994; Cutler, Mehler, Norris, & Segui, 1983, 1986, 1992; Dupoux, Pallier, Sebastian, & Mehler, 1997; Mehler, Dommergues, Frauenfelder, & Segui, 1981; Segui, Frauenfelder, & Mehler, 1981), Japanese (Cutler & Otake, 1994; Otake, Hatano, Cutler, & Mehler, 1993), and Spanish and Catalan (Sebastián-Gallés et al., 1992) show that prosody does indeed have a considerable influence on the perception and segmentation of speech. The extent to which prosody affects the temporal aspect of lexical processing of these languages remains to be established empirically. However, the data reported above on retroactive processing in English suggest that such a dependency is clearly a possibility.

CONCLUSION

To conclude, the literature that has allowed us to examine the use of time during lexical processing suggests that segmentation problems dictate how lexical processing unfolds in time. Sequential processing, which for a long time was considered a natural and universal principle, can no longer be viewed as the only mechanism during speech processing. Overcoming the segmentation puzzle by taking advantage of the most efficient word boundary cues available in the signal apparently operates at the expense of the very simple sequential procedure. For example, we saw that metrical segmentation must also include a "repair" function that proceeds retroactively. This is not to say that proactive processing does not play any role in word recognition. Evidence in favor of this process is too compelling to be ignored. Rather, both proactive and retroactive mechanisms seem to be necessary to parse the input successfully.

Similarly, it is likely that the temporal characteristics of both serendipitous and explicit segmentation strategies operate in concert during lexical processing. For example, competition (as it is postulated in TRACE or Shortlist) and metrical segmentation could conceivably combine their effects to lead to a more efficient and faster decomposition of the continuous speech input. This possibility, implemented in the Shortlist model, was successfully modeled with a lexicon of over 25,000 words (McQueen et al., 1994; Norris et al., 1995). Likewise, it

is possible that segmental cues such as glottal stops or allophonic variations and supra-segmental cues such as lexical stress are both used by the speech processor explicitly to help locate word boundaries.

In addition, given that the magnitude of segmentation problems can vary considerably across languages, a unique use of time during lexical processing that is successful universally is very improbable. Rather, the way in which the processor handles the unfolding and evanescent nature of speech should be seen as highly flexible. After all, the goal of an adaptive speech system faced with a continuous input is to maximize its chances to reach successful segmentation by taking advantage of all the tools at its disposal. In that respect, insight into the segmentation procedures specific to different languages should provide critical information about the timing of lexical processing and its relationship with segmentation.

Finally, examinations of the time course of word recognition have so far included little consideration for possible time reversals. As a consequence—or perhaps a cause—few of the paradigms used today are well suited for studying the use of time during lexical processing and segmentation. Techniques for exploring the contents and time-related parameters of auditory memory may provide new information about the temporal nature of speech processing. Similarly, cross-modal priming research offers promising ways of obtaining a cross-section of the lexical candidates activated at any one time from the input.

REFERENCES

- ALLEN, G. D. (1967). Two behavioral experiments on the location of the syllable beat in spoken American English. *Studies in Language & Language Behavior*, 4, 2-179.
- ALTMANN, G., & CARTER, D. (1989). Lexical stress and lexical discriminability: Stressed syllables are more informative, but why? *Computer Speech & Language*, 3, 265-275.
- ARCHIBALD, J. (1993). *Language, learnability, and L2 phonology: The acquisition of metrical parameters*. Dordrecht: Kluwer.
- BANEL, M.-H., & BACRI, N. (1994). On metrical patterns and lexical parsing in French. *Speech Communication*, 15, 115-126.
- BANSAL, R. K. (1966). *The intelligibility of Indian English: Measures of the intelligibility of connected speech, and sentence and word material, presented to listeners of different nationalities*. Unpublished doctoral dissertation, University of London.
- BARD, E. G., SHILLCOCK, R. C., & ALTMANN, G. T. M. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception & Psychophysics*, 44, 395-408.
- BASTIAN, J., EIMAS, P., & LIBERMAN, A. M. (1961). Identification and discrimination of phonemic contrast induced by silent interval [Abstract]. *Journal of the Acoustical Society of America*, 33, 842.
- BOND, Z. S. (1971). Units in speech perception. *Ohio State University Working Papers in Linguistics*, 9, 1-112.
- BOND, Z. S., & GARNES, S. (1980). Misperception of fluent speech. In R. Cole (Ed.), *Perception and production of fluent speech* (pp. 115-132). Hillsdale, NJ: Erlbaum.
- BOND, Z. S., & SMALL, L. H. (1983). Voicing, vowel, and stress mispronunciations in continuous speech. *Perception & Psychophysics*, 34, 470-474.
- BRADLEY, D. C. (1980). Lexical representation of derivational relation. In M. Aronoff & M.-L. Keen (Eds.), *Juncture* (pp. 37-55). Saratoga, CA: Anna Libri.
- BROKX, J. P. L. (1979). *Waargenomen continuïteit in spraak: Het belang van toonhoogte* [Continuity observed in speech: The importance of pitch]. Doctoral dissertation, Eindhoven University of Technology.
- BROWMAN, C. P. (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.
- BURGESS, C., TANENHAUS, M. K., & SEIDENBERG, M. S. (1989). Context and lexical access: Implications of nonword interference for lexical ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 15, 620-632.
- BUTCHER, A. (1973). *Experimente zur Pausenperzeption* (Arbeitsberichte 1). University of Kiel: Institut für Phonetik.
- BUXTON, H. (1983). Temporal predictability in the perception of English speech. In A. Cutler & D. R. Ladds (Eds.), *Prosody: Models and measurements* (pp. 111-121). Berlin: Springer-Verlag.
- CAMPBELL, R., & DODD, B. (1982). Some suffix effects on lipread lists. *Canadian Journal of Psychology*, 36, 509-515.
- CHODOROW, M. S. (1979). Time-compressed speech and the study of lexical and syntactic processing. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 87-111). Hillsdale, NJ: Erlbaum.
- CHOMSKY, N., & HALLE, M. (1968). *The sound pattern of English*. New York: Harper & Row.
- CLUFF, M. S., & LUCE, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception & Performance*, 16, 551-563.
- COLE, R. A. (1973). Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics*, 13, 153-156.
- COLE, R. A., & JAKIMIK, J. (1978). Understanding speech: How words are heard. In G. Underwood (Ed.), *Strategies of information processing* (pp. 67-116). London: Academic Press.
- COLE, R. A., & JAKIMIK, J. (1980a). How are syllables used to recognize words? *Journal of the Acoustical Society of America*, 67, 965-970.
- COLE, R. A., & JAKIMIK, J. (1980b). A model of speech perception. In R. Cole (Ed.), *Perception and production of fluent speech* (pp. 133-163). Hillsdale, NJ: Erlbaum.
- CONNINE, C. M., BLASKO, D. G., & TITONE, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory & Language*, 32, 193-210.
- CONNINE, C. M., CLIFTON, C., JR., & CUTLER, A. (1987). Effects of lexical stress and phonetic categorization. *Phonetica*, 44, 133-146.
- CROWDER, R. G. (1967). Prefix effects in immediate memory. *Canadian Journal of Psychology*, 21, 450-461.
- CROWDER, R. G. (1983). The purity of auditory memory. *Philosophical Transactions of the Royal Society of London: Series B*, 302, 251-265.
- CROWDER, R. G., & MORTON, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics*, 5, 365-373.
- CUTLER, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics*, 20, 55-60.
- CUTLER, A. (1986). Forbear is an homophone: Lexical prosody does not constrain lexical access. *Language & Speech*, 29, 201-220.
- CUTLER, A. (1989). Auditory lexical access: Where do we start? In W. Marslen-Wilson (Ed.), *Lexical representations and process* (pp. 342-356). Cambridge, MA: MIT Press.
- CUTLER, A. (1990). Exploiting prosodic probabilities in speech segmentation. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 105-121). Cambridge, MA: MIT Press.
- CUTLER, A. (1993). Phonological cues to open- and closed-class words in the processing of spoken sentences. *Journal of Psycholinguistic Research*, 22, 109-131.
- CUTLER, A. (1996). Prosody and the word boundary problem. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 87-99). Mahwah, NJ: Erlbaum.
- CUTLER, A., & BUTTERFIELD, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory & Language*, 31, 218-236.
- CUTLER, A., & CARTER, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, 2, 133-142.

- CUTLER, A., & CLIFTON, C., JR. (1984). The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 183-196). Hillsdale, NJ: Erlbaum.
- CUTLER, A., & FOSS, D. J. (1977). On the role of sentence stress in sentence processing. *Language & Speech*, **20**, 1-10.
- CUTLER, A., MEHLER, J., NORRIS, D. G., & SEGUI, J. (1983). A language specific comprehension strategy. *Nature*, **304**, 159-160.
- CUTLER, A., MEHLER, J., NORRIS, D. G., & SEGUI, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory & Language*, **25**, 385-400.
- CUTLER, A., MEHLER, J., NORRIS, D. G., & SEGUI, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, **24**, 381-410.
- CUTLER, A., & NORRIS, D. G. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. T. C. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 113-134). Hillsdale, NJ: Erlbaum.
- CUTLER, A., & NORRIS, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 113-121.
- CUTLER, A., & OTAKE, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory & Language*, **33**, 824-844.
- DE GELDER, B., & VROOMEN, J. (1994). Metrical segmentation and lexical competition: A happy affair? *Dokkyo International Review*, **7**, 176-193.
- DOOLING, D. J. (1972). Some context effects in the speeded comprehension of sentences. *Journal of Experimental Psychology*, **93**, 56-62.
- DORMAN, M. F., RAPHAEL, L. J., & LIBERMAN, A. M. (1979). Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America*, **65**, 1518-1532.
- DUPOUX, E. (1989). *Identification des mots parlés: Détection de phonème et unité prélexicale*. Unpublished doctoral dissertation, EHESS, Paris.
- DUPOUX, E. (1993). The time-course of prelexical processing: The syllabic hypothesis revisited. In G. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: The Second Sperlonga meeting* (pp. 81-114). Hillsdale, NJ: Erlbaum.
- DUPOUX, E., & MEHLER, J. (1990). Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code. *Journal of Memory & Language*, **29**, 316-335.
- DUPOUX, E., PALLIER, C., SEBASTIAN, N., & MEHLER, J. (1997). A de-stressing "deafness" in French? *Journal of Memory & Language*, **36**, 406-421.
- EFRON, R. (1970a). Effects of stimulus duration on perceptual onset and offset latencies. *Perception & Psychophysics*, **8**, 231-234.
- EFRON, R. (1970b). The minimum duration of a perception. *Neuropsychologia*, **8**, 57-63.
- EFRON, R. (1970c). The relationship between the duration of a stimulus and the duration of a perception. *Neuropsychologia*, **8**, 37-55.
- FAIRBANKS, G., GUTTMAN, N., & MIRON, M. S. (1957). Auditory comprehension of repeated high speed messages. *Journal of Speech & Hearing Disorders*, **22**, 20-22.
- FOSS, D. J., & BLANK, M. A. (1980). Identifying the speech codes. *Cognitive Psychology*, **12**, 1-31.
- FOULKE, E. (1969). The perception of time compressed speech. In D. L. Horton & J. Jenkins (Eds.), *Perception of language: Proceedings of a symposium of the Learning Research and Development Center* (pp. 79-119). Columbus, OH: Charles E. Merrill.
- FOX, R. A. (1984). Effects of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 526-540.
- FRAUENFELDER, U. H., SEGUI, J., & DIJKSTRA, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 77-91.
- FRAUENFELDER, U. H., & TYLER, L. K. (1987). The process of spoken word recognition: An introduction. *Cognition*, **25**, 1-20.
- FRY, D. B. (1955). Duration and intensity of physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, **27**, 765-768.
- FRY, D. B. (1958). Experiments in the perception of stress. *Language & Speech*, **1**, 126-152.
- FUJISAKI, H., & KAWASHIMA, T. (1969). On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute*, **28**, 67-73. [Tokyo: University of Tokyo, Faculty of Engineering]
- FUJISAKI, H., & KAWASHIMA, T. (1970). Some experiments on speech perception and a model for the perceptual mechanisms. *Annual Report of the Engineering Research Institute*, **29**, 207-214. [Tokyo: University of Tokyo, Faculty of Engineering.]
- GARDING, E. (1967). *Internal Structure in Swedish*. Lund, Sweden: Gleerups.
- GARVEY, W. D. (1953). The intelligibility of speeded speech. *Journal of Experimental Psychology*, **45**, 102-108.
- GATHERCOLE, S. (1987). Lipreading: Implications for theories of short-term memory. In B. Dodd & R. Campbell (Eds.), *Hearing by eye* (pp. 227-241). Hove, U.K.: Erlbaum.
- GIMSON, A. C. (1989). *An introduction to the pronunciation of English* (4th ed.). London: Edward Arnold.
- GLUCKSBERG, S., KREUZ, R. J., & RHO, S. H. (1986). Context can constrain lexical access: Implications for models of language comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 323-335.
- GOLDINGER, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1166-1183.
- GOODMAN, J., & HUTTENLOCHER, J. (1988). Do we know how people identify spoken words? *Journal of Memory & Language*, **27**, 684-698.
- GOW, D. W., & GORDON, P. C. (1993). Coming to terms with stress: Effects of stress location in sentence processing. *Journal of Psycholinguistic Research*, **22**, 545-578.
- GREENE, R. L., & CROWDER, R. G. (1984). Modality and suffix effects in the absence of auditory stimulation. *Journal of Verbal Learning & Verbal Behavior*, **23**, 371-382.
- GROSJEAN, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, **28**, 267-283.
- GROSJEAN, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception & Psychophysics*, **38**, 299-310.
- GROSJEAN, F., & GEE, J. P. (1987). Prosodic structure and spoken word recognition. *Cognition*, **25**, 135-155.
- HAYES, B. (1981). *A metrical theory of stress rules*. Bloomington: Indiana University Linguistics Club.
- HUGGINS, A. W. F. (1964). Distortion of the temporal pattern of speech: Interruption and alternation. *Journal of the Acoustical Society of America*, **36**, 1055-1064.
- HUGGINS, A. W. F. (1972a). Just noticeable difference for segment duration in natural speech. *Journal of the Acoustical Society of America*, **51**, 1270-1278.
- HUGGINS, A. W. F. (1972b). On the perception of temporal phenomena in speech. *Journal of the Acoustical Society of America*, **51**, 1279-1290.
- HUGGINS, A. W. F. (1975). Temporally segmented speech and "echoic" storage. In A. Cohen & S. G. Neebboom (Eds.), *Structure and process in speech perception* (pp. 209-225). New York: Springer-Verlag.
- HUTTENLOCHER, D., & ZUE, V. (1983). Phonotactic and lexical constraints in speech recognition. *Working Papers of the Speech Communication Group at MIT*, **3**, 157-167.
- JAKIMIK, J. (1979). *The interaction of sound and knowledge in word recognition from fluent speech*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh.
- KLATT, D. H., & STEVENS, K. H. (1973). On the automatic recognition of continuous speech: Implications from a spectrogram reading experiment. *IEEE Transactions on Audio & Electroacoustics*, **21**, 210-217.
- KOLINSKY, R. (1992). Conjunctions errors as a tool for the study of perceptual processes. In J. Alegria, D. Holender, J. Morais, & M. Radreau (Eds.), *Analytic approaches to human cognition* (pp. 133-149). Amsterdam: North-Holland.
- KOLINSKY, R., MORAIS, J., & CLUYTENS, R. (1995). Intermediate representations in spoken word recognition: Evidence from word illusions. *Journal of Memory & Language*, **34**, 19-40.

- KOZHEVNIKOV, V., & CHISTOVICH, L. (1966). *Speech: Articulation and perception* (Translation JPRS 30, 543). Washington, DC: U.S. Department of Commerce, Joint Publications Research Service.
- LEHISTE, I. (1960). An acoustic-phonetic study of internal open juncture. *Phonetica*, **5** (Suppl.).
- LEHISTE, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.
- LEHISTE, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, **51**, 2018-2024.
- LIBERMAN, A. M. (1970a). The grammar of speech and language. *Cognitive Psychology*, **1**, 301-323.
- LIBERMAN, A. M. (1970b). Some characteristics of perception in the speech mode. In P. Hamburg, K. Pribram, & A. J. Standard (Eds.), *Perception and its disorders* (pp. 238-254). Baltimore: Williams & Wilkins.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431-461.
- LIBERMAN, M. Y. (1975). *The intonational system of English*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.
- LIBERMAN, M. Y., & PRINCE, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, **8**, 249-336.
- LIEBERMAN, P. (1965). On the acoustic basis of perception of stress by linguists. *Word*, **21**, 40-54.
- LISKER, L. (1957). Closure duration and the voiced-voiceless distinction in English. *Language*, **33**, 42-49.
- LUCE, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics*, **39**, 155-158.
- MALÉCOT, A., JOHNSTON, R., & KIZIAR, P.-A. (1972). Syllabic rate and utterance length in French. *Phonetica*, **26**, 235-251.
- MARCUS, S. M. (1978). Distinguishing "slit" and "split"—an invariant timing cue in speech perception. *Perception & Psychophysics*, **23**, 58-60.
- MARCUS, S. M., & FRAUENFELDER, U. H. (1985). Word recognition—Uniqueness or deviation? A theoretical note. *Language & Cognitive Processes*, **1**, 163-169.
- MARSLÉN-WILSON, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, **244**, 522-523.
- MARSLÉN-WILSON, W. D. (1984). Function and process in spoken word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 125-149). Hillsdale, NJ: Erlbaum.
- MARSLÉN-WILSON, W. D. (1985). Speech shadowing and speech comprehension. *Speech Communication*, **4**, 55-73.
- MARSLÉN-WILSON, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, **25**, 71-102.
- MARSLÉN-WILSON, W. D., & TYLER, L. (1980). The temporal structure of spoken language understanding. *Cognition*, **8**, 1-71.
- MARSLÉN-WILSON, W. D., & WELSH, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, **10**, 29-63.
- MARSLÉN-WILSON, W. D., & ZWITSERLOOD, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 576-585.
- MARTIN, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, **79**, 487-509.
- MASSARO, D. W. (1972a). Perceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, **79**, 124-145.
- MASSARO, D. W. (1972b). Stimulus information vs processing time in auditory pattern recognition. *Perception & Psychophysics*, **12**, 50-56.
- MASSARO, D. W. (1974). Perceptual units in speech recognition. *Journal of Experimental Psychology*, **102**, 199-208.
- MASSARO, D. W. (1975). Acoustic features in speech perception. In D. W. Massaro (Ed.), *Understanding language: An information processing analysis of speech perception, reading, and psycholinguistics* (pp. 77-124). New York: Academic Press.
- MASSARO, D. W., & IDSON, W. L. (1978). Temporal course of perceived vowel duration. *Journal of Speech & Hearing Research*, **21**, 37-55.
- MATTYS, S. L., & SAMUEL, A. G. (1997). How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *Journal of Memory & Language*, **36**, 87-116.
- MCCALLISTER, J. (1991). The processing of lexically stressed syllables in read and spontaneous speech. *Language & Speech*, **34**, 1-26.
- MCCLELLAND, J. L., & ELMAN, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.
- MCQUEEN, J. M., CUTLER, A., BRISCOE, T., & NORRIS, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Language & Cognitive Processes*, **10**, 309-331.
- MCQUEEN, J. M., NORRIS, D., & CUTLER, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 621-638.
- MEHLER, J., DOMMERGUES, J.-Y., FRAUENFELDER, U., & SEGUI, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning & Verbal Behavior*, **20**, 298-305.
- MEHTA, G., & CUTLER, A. (1988). Detection of target phonemes in spontaneous and read speech. *Language & Speech*, **31**, 135-156.
- MELTZER, R. H., MARTIN, J. G., MILLS, C. B., IMHOFF, D. L., & ZOHAR, D. (1976). Anticipatory coarticulation and reaction time to phoneme targets in spontaneous speech. *Phonetica*, **37**, 159-168.
- MILLER, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, **63**, 81-97.
- MILLER, J. L. (1981). Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 39-74). Hillsdale, NJ: Erlbaum.
- MILLER, J. L. (1987). Mandatory processing in speech perception. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural language understanding* (pp. 309-322). Cambridge, MA: MIT Press.
- MILLER, J. L., & DEXTER, E. R. (1988). Effects of speaking rates on phonetic perception. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 369-378.
- MILLER, J. L., & LIBERMAN, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, **25**, 457-465.
- NAIRNE, J. S., & CROWDER, R. G. (1982). On the locus of the stimulus suffix effect. *Memory & Cognition*, **10**, 350-357.
- NAIRNE, J. S., & WALTERS, V. (1983). Silent mouthing produces modality and suffix-like effects. *Journal of Verbal Learning & Verbal Behavior*, **22**, 475-483.
- NAKATANI, L. H., & DUKES, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, **62**, 714-719.
- NAKATANI, L. H., & SCHAFFER, J. A. (1978). Hearing "words" without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America*, **63**, 234-245.
- NEWMAN, J. E., & DELL, G. S. (1978). The phonological structure of phoneme monitoring: A critique of some ambiguity studies. *Journal of Verbal Learning & Verbal Behavior*, **17**, 359-374.
- NOOTEBOOM, S. G. (1979). The time course of speech perception. In W. J. Barry & K. J. Kohler (Eds.), *"Time" in the production and perception of speech* (Arbeitsberichte 12). University of Kiel: Institut für Phonetik.
- NOOTEBOOM, S. G. (1981). Lexical retrieval from fragments of spoken words: Beginnings versus endings. *Journal of Phonetics*, **9**, 407-424.
- NOOTEBOOM, S. G., BROKX, J. P. L., & DE ROOIJ, J. J. (1978). Contribution of prosody to speech perception. In W. J. M. Levelt & G. B. Flores d'Arcais (Eds.), *Studies in the perception of language* (pp. 75-107). New York: Wiley.
- NORRIS, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, **52**, 189-234.
- NORRIS, D., MCQUEEN, J. M., & CUTLER, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1209-1228.
- ODEN, G. C., & SPIRA, J. L. (1983). Influence of context on the activation and selection of ambiguous word sense. *Quarterly Journal of Experimental Psychology*, **35A**, 51-64.
- OTAKE, T., HATANO, G., CUTLER, A., & MEHLER, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory & Language*, **31**, 218-236.
- OTTEVANGER, I. B. (1984). The detection of mispronunciations and

- the influence of context. In M. P. R. van den Broecke & A. Cohen (Eds.), *Proceedings of the Tenth International Congress of Phonetic Sciences* (Vol. IIb, pp. 486-490). Dordrecht: Foris.
- PALMERI, T. J., GOLDINGER, D. D., & PISONI, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 309-328.
- PISONI, D. B. (1981). Some theoretical issues in speech perception. *Cognition*, **10**, 249-259.
- PITT, M. A., & SAMUEL, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 564-573.
- PITT, M. A., & SAMUEL, A. G. (1995). Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology*, **29**, 149-188.
- PLOMP, R. (1964). Decay of auditory sensation. *Journal of the Acoustical Society of America*, **36**, 277-282.
- PORT, R. (1976). *The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words*. Unpublished doctoral dissertation, University of Connecticut.
- RADEAU, M., & MORAIS, J. (1990). The uniqueness point effect in the shadowing of spoken word. *Speech Communication*, **9**, 155-164.
- RADEAU, M., MORAIS, J., & SEGUI, J. (1995). Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 1297-1311.
- RADEAU, M., MOUSTY, P., & BERTELSON, P. (1989). The effect of the uniqueness point in spoken word recognition. *Psychological Research*, **51**, 123-128.
- RADEAU, M., MOUSTY, P., & PASDELOUP, V. (1994). *L'effet du débit de la parole sur le rôle du point d'unicité*. Paper presented at the XXèmes Journées d'Étude sur la Parole, Trégastel.
- REDDY, R. (1976). Speech recognition by machine: A review. *Proceedings of the IEEE*, **64**, 501-531.
- REMINGTON, R. (1977). Processing of phonemes in speech: A speed-accuracy study. *Journal of the Acoustical Society of America*, **62**, 1279-1290.
- SALASOO, A., & PISONI, B. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory & Language*, **24**, 210-231.
- SAMUEL, A. G. (1987). Lexical uniqueness effects on phonemic restoration. *Journal of Memory & Language*, **26**, 36-56.
- SEBASTIÁN-GALLÉS, N., DUPOUX, E., SEGUI, J., & MEHLER, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory & Language*, **31**, 18-32.
- SEGUI, J., FRAUENFELDER, U., & MEHLER, J. (1981). Phoneme monitoring, syllable monitoring and lexical access. *British Journal of Psychology*, **72**, 471-477.
- SEIDENBERG, M. S., TANENHAUS, M. K., LEIMAN, J. M., & BIENKOWSKI, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, **14**, 489-537.
- SELKIRK, E. O. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry*, **11**, 536-605.
- SELKIRK, E. O. (1984). *Phonology and syntax: The relationship between sound and structure*. Cambridge, MA: MIT Press.
- SHIELDS, J. L., MCHUGH, A., & 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. (1990). Lexical hypotheses in continuous speech. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 24-49). Cambridge, MA: MIT Press.
- SHOCKEY, L. (1983). Phonetic and phonological properties of connected speech. *Ohio State University Working Papers in Linguistics*, **17**.
- SIMON, H. A. (1974). How big is a chunk? *Science*, **183**, 482-488.
- SIMPSON, G. B. (1984). Lexical ambiguity and its role in models of word recognition. *Psychological Bulletin*, **96**, 316-340.
- SLIS, I. H., & VAN NIEROP, P. J. (1970). On the forward masking threshold of vowels in VC-combinations. *Institute for Perception Research Annual Progress Report*, **5**, 68-72.
- SLOWIACZEK, L. M. (1987). *On the role of lexical stress in auditory word recognition*. Unpublished doctoral dissertation, Indiana University, Bloomington.
- SLOWIACZEK, L. M. (1990). Effects of lexical stress in auditory word recognition. *Language & Speech*, **33**, 47-68.
- SLOWIACZEK, L. M., NUSBAUM, H. C., & PISONI, D. B. (1987). Phonological priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 64-75.
- SMALL, L. H., & BOND, Z. S. (1982, November). *Effects of mispronunciation on lexical access in continuous speech perception*. Paper presented at the meeting of the American Speech-Language Hearing Association, Toronto.
- SPOEHR, K. T., & CORIN, W. J. (1978). The stimulus suffix effect as a memory coding phenomenon. *Memory & Cognition*, **6**, 583-589.
- STICHT, T. G., & GRAY, B. B. (1969). The intelligibility of time compressed words as a function of age and hearing loss. *Journal of Speech & Hearing Research*, **12**, 443-448.
- STUDDERT-KENNEDY, M. (1975). Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 243-293). New York: Academic Press.
- SUMMERFIELD, A. Q., & BAILEY, P. (1977). On the dissociation of spectral and temporal cues for stop consonant manner [Abstract]. *Journal of the Acoustical Society of America*, **61**, S46.
- SWINNEY, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning & Verbal Behavior*, **18**, 645-659.
- TABOSSI, P. (1988). Accessing lexical ambiguity in different types of sentential context. *Journal of Memory & Language*, **27**, 324-340.
- TABOSSI, P. (1993). Connections, competitions, and cohorts: Comments on the chapters by Marslen-Wilson; Norris; and Bard and Shillcock. In G. T. M. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing* (pp. 277-294). Hillsdale, NJ: Erlbaum.
- TABOSSI, P., BURANI, C., & SCOTT, D. (1995). Word identification in fluent speech. *Journal of Memory & Language*, **34**, 440-467.
- TAFT, L. A. (1984). *Prosodic constraints and lexical parsing strategies*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.
- TAFT, M., & HAMBLY, G. (1986). Exploring the cohort model of spoken word recognition. *Cognition*, **22**, 259-282.
- TANENHAUS, M. K., LEIMAN, J. M., & SEIDENBERG, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning & Verbal Behavior*, **18**, 427-440.
- THOMAS, I. B., HILL, P. B., CARROLL, F. S., & BIENVENIDO, G. (1970). Temporal order in the perception of vowels. *Journal of the Acoustical Society of America*, **48**, 1010-1013.
- THORNE, J. P., BRATLEY, P., & DEWAR, H. (1968). The syntactic analysis of English by machine. In D. Michie (Ed.), *Machine intelligence 3* (pp. 281-309). New York: Elsevier.
- TYLER, L. K. (1984). The structure of the initial cohort: Evidence from gating. *Perception & Psychophysics*, **36**, 417-427.
- TYLER, L. K., & WESSELS, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception & Psychophysics*, **34**, 409-420.
- VROOMEN, J., & DE GELDER, B. (1994, November). *Lexical competition in spoken word recognition*. Paper presented at the 35th annual meeting of the Psychonomic Society, St. Louis.
- VROOMEN, J., & DE GELDER, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 98-108.
- WARREN, R. M., OUSEK, C. J., FARMER, R. M., & WARREN, R. P. (1969). Auditory confusions: Confusions of patterns other than speech or music. *Science*, **164**, 586-587.
- WINGFIELD, A., & WHEALE, J. L. (1975). Word rate and intelligibility of alternated speech. *Perception & Psychophysics*, **18**, 317-320.
- WORKING GROUP ON SPEECH UNDERSTANDING AND AGING (1988). Speech understanding and aging. *Journal of the Acoustical Society of America*, **83**, 859-895.
- ZWITSERLOOD, P. (1989). The locus of effects of sentential-semantic context in spoken-word processing. *Cognition*, **32**, 25-64.
- ZWITSERLOOD, P., & SCHRIEFERS, H. (1995). Effects of sensory infor-

mation and processing time in spoken-word recognition. *Language & Cognitive Processes*, 10, 121-136.

NOTES

1. This observation, however, applies only to young normal-hearing adults. Research has shown that both the elderly and the hearing-impaired (young and old) exhibit a marked decrement in their understanding of high-rate speech (Sticht & Gray, 1969; Working Group on Speech Understanding and Aging, 1988).

2. Note that this figure includes all content words starting with either a primary or a secondary stressed syllable. Of these two categories, 2.9% start with a secondary stressed syllable (2.6% of the entire sample of lexical words). It may be undesirable to include this category among those supposed to initiate successful lexical access on a metrical basis. Indeed, a system that would trigger lexical access on primary *and* secondary stressed syllables would ipso facto postulate an erroneous word boundary in the middle of a majority of polysyllabic words. Not counting word-initial secondary stress as a segmentation marker would make the system more efficient with a minimal loss of statistical efficiency.

As for the distinction between content words and function words, Cutler (1990) and Cutler and Carter (1987) reported that in a 190,000-word corpus of spontaneous conversation, three quarters of all strong

syllables in the sample were the sole or initial syllable of a content word while over two thirds of the weak syllables were the sole or initial syllable of a function word. For this reason, the metrical segmentation strategy postulated by Cutler and Norris (1988) assumes two separate lists: one of content words and one of function words. The list of content words is accessed by default whenever a strong syllable is encountered in the signal. Weak syllables that are left unused after search in the content word list are looked up in the function word list. When Cutler and Carter tested the efficiency of this algorithm on a minicorpus, they found that 82% of words were assigned to the correct list on the first pass.

3. In their experiments, Pitt and Samuel (1990) used two types of contexts to predict the stress pattern of the test word: a sentential context comparable to Gow and Gordon's (1993) and a word list context. Only the latter yielded significant response time effects.

4. Free speculation about the size of this window could lead one to wonder whether the span of processing inside the window is the same to "the left" and to "the right" of the stressed syllable. Could it be that, as a result of the directionality of processing, the information after the stressed syllable spread farther than the information before?

(Manuscript received May 23, 1996;
revision accepted for publication May 1, 1997.)