Speech Perception and Spoken Word Recognition: Past and Present

Peter W. Jusczyk and Paul A. Luce

Objective: **The scientific study of the perception of spoken language has been an exciting, prolific, and productive area of research for more than 50 yr. We have learned much about infants' and adults' remarkable capacities for perceiving and understanding the sounds of their language, as evidenced by our increasingly sophisticated theories of acquisition, process, and representation. We present a selective, but we hope, representative review of the past half century of research on speech perception, paying particular attention to the historical and theoretical contexts within which this research was conducted. Our foci in this review fall on three principle topics: early work on the discrimination and categorization of speech sounds, more recent efforts to understand the processes and representations that subserve spoken word recognition, and research on how infants acquire the capacity to perceive their native language. Our intent is to provide the reader a sense of the progress our field has experienced over the last half century in understanding the human's extraordinary capacity for the perception of spoken language.**

(Ear & Hearing 2002;23;2–40)

The Beginnings of Speech Perception Research: Some Core Issues

Research on the perception of speech began in earnest during the 1950s. Not surprisingly, the agenda for the field was set by the kinds of issues that preoccupied language researchers of that era. In the field of linguistics, a major goal was to devise rigorous scientific procedures that would yield a correct structural description of a particular language when provided with a detailed corpus of utterances (Bloomfield, 1933; Harris, 1955). This approach, known as taxonomic linguistics, made certain assumptions about how such a description should proceed. The general view was that a language is hierarchically organized at a number of distinctive levels, so that an accurate description of language structure required descriptions of the organization at each level of the hierarchy. Moreover, because the aim was to achieve a scientific description, the structural analysis of a particular language began with what could be directly observed, namely, the acoustic waveforms of utterances. Because of this emphasis on what could be objectively observed, the description provided for a given level was not supposed to depend on the description of any of the higher levels. In principle, then, a description of a language would begin with an account of how acoustic properties map onto phonetic segments (the phonetic level), how the phonetic segments mapped to particular phonemes (the phonemic level), how the phonemes are combined to form morphemes in the language (morphemic level), and eventually to how the morphemes are combined to form sentences (the syntactic level). Thus, although speech perception research is very much an interdisciplinary field, drawing on concepts and methods from physics, engineering, linguistics, and psychology, many of the core problems that drove the early research efforts reflected the view that languages are hierarchically organized.

What core issues helped to shape speech research during its early history? Three issues were at the heart of most early studies: invariance, constancy, and perceptual units. All are crucial for understanding how the acoustic signal is transformed into phonetic segments. We consider each of these in turn.

Invariance • Each element in the periodic table has its own unique atomic number that corresponds to the number of protons that are contained in its nucleus. Hence, if we know how many protons are present, we know the identity of the element. If speech were structured in this way, each phonetic segment should be specified by a unique set of acoustic properties. Unfortunately, the situation is considerably more complex, as Delattre, Liberman, and Cooper (1955) discovered when looking for acoustic invariants for consonants such as [d]. Specifically, they examined the acoustic realizations for [d] produced in combination with different following vowels (e.g., [di], [da], [du]) and found no obvious common acoustic properties specifying [d] in each of these contexts. Instead, the acoustic characteristics of [d] were strongly influenced by the following vowel. This phenomenon is an example of coarticulation, which refers to the fact that the consonant and vowel segments are produced more or less simultaneously, as opposed to sequentially. This fact

Departments of Psychology and Cognitive Science (P.W.J.), Johns Hopkins University, Baltimore, Maryland; and Department of Psychology and Center for Cognitive Science (P.A.L.), University at Buffalo, Buffalo, New York.

0196/0202/02/2301-0002/0 • Ear & Hearing • Copyright © 2002 by Lippincott Williams & Wilkins • Printed in the U.S.A.

alone indicated that the search for invariant features of phonetic segments would be more complicated than our example of elements and atomic numbers.

Although there might not be a single set of acoustic features that identifies a particular phonetic segment in all contexts, perhaps a weaker version of the acoustic invariance hypothesis holds. Might there be equivalence classes of acoustic features that serve to identify one and only one type of phonetic segment when it occurs in different contexts (e.g., one set of features before [a], another set before [u], etc?) However, even this weaker version of the hypothesis is incorrect. Liberman, Delattre, and Cooper (1952) found that an identical burst of acoustic noise placed in front of the vowels [a] and [u] resulted in the perception of the syllables [pi], [ka], and [pu]. So, the same noise is perceived as a [p] in some contexts, but as a [k] in others. These findings undermine any simple notion of identifying phonetic segments on the basis of invariant acoustic properties. Consequently, Liberman and his colleagues developed the Motor Theory of Speech Perception, whereby speech is perceived by reference to the articulatory gestures used to produce it (Liberman, Cooper, Harris, & MacNeilage, 1963; Liberman & Mattingly, 1985; Liberman & Whalen, 2000). An alternative view, Direct Perception, holds that listeners directly perceive articulatory gestures in the speech signal (Fowler, 1986; Fowler & Rosenblum, 1991). Meanwhile, other researchers explored alternative descriptions of the acoustic characteristics of speech sounds in hopes of revealing invariant properties of phonetic segments (Blumstein & Stevens, 1978; Kewley-Port, Reference Note 12; Sawusch, 1992; Searle, Jacobson, & Rayment, 1979; Stevens, 1998; Sussman, McCaffrey, & Matthews, 1991).

Constancy • Besides the variability that exists in the realization of phonetic segments in the context of other phonetic segments, there are additional sources of variability in speech with which the perceptual system must cope. For example, even when different talkers produce the same intended speech sound, the acoustic characteristics of their productions will differ. Women's voices tend to be higher pitched than men's, and even within a particular sex, there is a wide range of acoustic variability (e.g., consider the speaking voices of Mike Tyson and James Earl Jones).

Individual differences in talkers' voices are a natural consequence of differences in the sizes and shapes of their vocal tracts. The length and mass of the vocal folds, as well as the overall length of the vocal tract, affect the typical fundamental frequency (or pitch) of a voice. Factors having to do with the flexibility of the tongue, the state of one's vocal folds, missing teeth, etc., also will affect the acoustic characteristics of speech. In some cases, the production of a particular word by one talker might actually be more similar acoustically to the production of a different word by another talker than it is to the second talker's production of the same word (Ladefoged & Broadbent, 1957). Peterson and Barney (1952) collected data on the production of different English vowels by 76 talkers. Their acoustic analyses indicated that tokens of a particular vowel from some talkers actually overlapped with the productions of a different vowel from other talkers. Although human listeners cope with this type of variability relatively easily (Creelman, 1957; Verbrugge, Strange, Shankweiler, & Edman, 1976), it poses an obstacle to accurate speech recognition by machines, usually requiring a prior training period to achieve a modest level of accuracy in identifying words produced by particular talkers (Bernstein & Franco, 1996).

There is even considerable variability in productions of the same target by a single talker. Simply changing one's speaking register from adult-directed speech to child-directed speech changes not only pitch, but also other acoustic properties of speech (Fernald et al., 1989; Kuhl et al., 1997). Moreover, speaking rate may vary considerably on different occasions, leading to concomitant changes in the acoustic signal. Phonetic distinctions based on the rate of temporal change of certain acoustic features (such as [b] versus [w]) are especially likely to be affected by changes in speaking rate (Liberman, Delattre, Gerstman, & Cooper, 1956; Miller & Liberman, 1979). Thus, listeners cannot rely on absolute differences in the duration of some particular acoustic feature to cue the phonetic distinction.

Other types of speech contrasts are also affected by changes in speaking rate. Summerfield (Reference Note 21) found that differences in the voicing characteristics of voiced (e.g., [d]) and voiceless (e.g., [t]) stops tend to decrease as speaking rate speeds. Furthermore, the nature of the acoustic changes that ensue is not always predictable a priori. Port (Reference Note 18) found that an increase in speaking rate actually decreased the duration of a long vowel to a greater extent than it did a short vowel [I]. Verbrugge and Shankweiler (1977) demonstrated that such durational changes could affect the identification of particular segments. When syllables taken from utterances at fast speaking rates were spliced into slow speaking rate contexts, listeners tended to misidentify long vowels (e.g., [a]) as short vowels (e.g., [Λ]). Not surprisingly, then, how listeners cope with intra- and inter-talker variability garnered the attention of many researchers.

Perceptual Units • Many early studies assumed that the elementary unit of perception was equivalent to that minimal-sized unit that could distinguish two different word forms, namely, the phonetic segment. It was recognized that phonetic segments could contrast minimally with each other according to certain features (Jakobson, Fant, & Halle, 1952). Thus, [b] contrasts with [p] on a voicing feature; with [m] on an oral/nasal manner of articulation feature; with [d] on a place of articulation feature, etc. According to Jakobson et al., "The distinctive features are the ultimate distinctive entities of language since no one of them can be broken down into smaller linguistic units" (p. 3). However, the distinctive features combine into one concurrent bundle, which corresponds to the phonetic segment. Because the phonetic segment was considered to be the minimal speech sound unit, it was only natural that researchers assumed that there would be direct acoustic correlates of such units.

The development of the pattern playback synthesizer made it possible to explore how various elements in the acoustic signal affected speech perception (Cooper, Delattre, Liberman, Borst, & Gerstman, 1952; Cooper, Liberman, & Borst, Reference Note 5). The patterns used to generate speech sounds with this synthesizer are based on spectrographic analyses of speech. The time-by-frequency-by-intensity analysis of a spectrogram shows that acoustic energy is often concentrated in bands at different acoustic frequencies. These bands, called formants, correspond to the natural resonant frequencies of the vocal tract during speech production. When researchers began to seek the acoustic features corresponding to phonetic segments, they soon discovered that there was no way to divide the formants of a consonant-vowel (CV) syllable, such as /di/, into pieces corresponding to each individual segment (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). In particular, no portion of the formant patterns corresponds to the /d/ alone. Rather listeners report hearing either a /di/ or a nonspeech sound. Such observations led Liberman et al. to conclude that even the beginning portions of the acoustic signal carry information about both the consonant and the vowel simultaneously (i.e., these phonetic segments are coarticulated).

The findings suggesting that phonetic segments were not necessarily assigned to distinct portions of the acoustic signal stimulated a great deal of research directed at identifying the elementary unit of perception. Savin and Bever (1970) found that listeners were faster at detecting syllable targets than phoneme targets. Hence, they suggested that syllables were the basic units of speech perception, and that phonemes are derived secondarily from these (see also Massaro, 1972). However, under different circumstances, faster monitoring times were obtained for phonemes (Cutler, Norris, & Williams, 1987; Healy & Cutting, 1976; Mills, 1980; Swinney & Prather, 1980) or for units larger than syllables (McNeill & Lindig, 1973). Yet, criticisms were raised about the methods used in many of these later studies (Mehler, Segui, & Frauenfelder, 1981). It is fair to say that no strong consensus has emerged regarding the basic perceptual unit. Indeed, there are proponents for a range of different sized units including demi-syllables (Fujimura, 1976), context-sensitive allophones (Wickelgren, 1969), syllables (Cole & Scott, 1974; Mehler, Dommergues, Frauenfelder, & Segui, 1981; Studdert-Kennedy, 1974), and even context-sensitive spectra (Klatt, 1979).

How a Bottom-Up View of Perception Affected Speech Research

The emphasis on finding the acoustic correlates of the minimal units of linguistic description quite naturally led researchers to focus on phenomena relating to the perception of phonemes. Because phonemes are the elementary sound units for constructing the words of a language, a general assumption was that perceiving words first required the recovery of the sequence of phonemes in the word. In addition, pragmatic considerations encouraged investigators to solve the problem of phonemic perception. In his eloquent autobiographical narrative, Liberman (1996) notes that a major impetus for the early investigations was to devise a reading machine for the blind that would render printed words as spoken words. It was assumed that this would involve the transduction of printed letters to the corresponding elementary sounds or, in other words, phonemes.

The range of topics treated below, though by no means exhaustive, are the ones that attracted a lion's share of early researchers' attention. All pertain to the goal of understanding how phonemes are extracted from the speech signal.

Categorical Perception • When researchers at Haskins Laboratories began to study the consequences of manipulating information in the acoustic signal to determine its importance for the perception of phonemes such as the stop consonants /b/, /d/, and /g/, they discovered an interesting phenomenon. They found that "a listener can better discriminate between sounds that lie on the opposite sides of a phoneme boundary than he can between sounds that fall within the same phoneme category" (Liberman, Harris, Hoffman, & Griffith, 1957, p. 358). This finding was surprising because in perceiving other types of acoustic signals, listeners can typically discriminate many more distinctions between stimuli than they can provide distinct labels for. Miller (1956) cites findings from Pollack (1952) demonstrating that listeners discriminate about 1200 pitch differences between 100 Hz and 8000 Hz, but only consistently use about seven labels within this pitch range. By comparison, Liberman et al.' s subjects were poor at discriminating among acoustic differences between instances of stop consonants within a phoneme category (i.e., /b/, /d/, or /g/).

In their acoustic characteristics, /b/, /d/, and /g/ differ in the nature of their initial formant transitions. By creating a graded series of formant transitions, Liberman et al. (1957) produced a continuum that spanned the entire range of place of articulation changes from /b/ to /d/ to /g/. When their subjects were asked to label and discriminate among the stimuli in this continuum, they observed sharp labeling boundaries between categories but poor discrimination within categories, a phenomenon called categorical perception. Moreover, categorical perception was also noted for other dimensions of phonetic contrast such as voicing (Liberman, Harris, Kinney, & Lane, 1961; Lisker & Abramson, Reference Note 13) and manner of articulation (Miyawaki et al., 1975). Three consistent characteristics are considered to be the defining features of categorical perception (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). First, plots of adults' labeling of stimuli from such continua show abrupt, steep slopes between adjacent phonemic categories. Second, discrimination of stimuli from within the same phonemic category tends to be poor, whereas discrimination of equal-sized acoustic differences between two stimuli from different phonemic categories is good. Third, peaks in discriminability of stimulus pairs along a given continuum correspond to the category boundaries obtained in labeling the stimuli.

Categorical perception is typically found for contrasts between many different pairs of consonants. By comparison, the perception of vowel contrasts is said to be continuous. Fry, Abramson, Eimas, and Liberman (1962) investigated a continuum that ranged from /I/ to /E/ to /æ/ and found that discrimination among items from within the same vowel category was quite good. Thus, rather than being categorical, the perception of vowel contrasts was similar to what had been reported for the discrimination of nonspeech sounds (Pollack, 1952; 1953).

The view that categorical perception might be unique to speech gained strength from findings that, with similar acoustic differences in nonspeech analogs of speech contrasts, perception of these nonspeech contrasts was continuous, even when similar changes in speech sounds were perceived categorically (Liberman, Harris, Eimas, Lisker, & Bastian, 1961). The view that categorical perception was unique to speech became a core element of arguments for a special mode of perception for speech processing (Liberman, 1970; Liberman et al., 1967).

Speech versus Nonspeech Processing • The claim that speech and nonspeech sound processing are accomplished by different mechanisms or modes of perception was not universally accepted. Many investigations explored whether speech and nonspeech sounds are processed in the same manner by human listeners. Early investigations centering on categorical perception seemed to indicate that categorical perception was restricted to speech (Liberman, Harris, Kinney, & Lane, 1961; Mattingly, Liberman, Syrdal, & Halwes, 1971; Miyawaki et al., 1975). However, when investigators explored the perception of complex nonspeech stimuli, they found evidence of categorical perception. Indeed, the relationships manipulated in these studies (e.g., varying the timing of one component with respect to another) appeared to be analogous to phonetic dimensions such as voicing, prompting claims that general auditory mechanisms underlie the perception of speech contrasts (Miller, Weir, Pastore, Kelly, & Dooling, 1976; Pisoni, 1977). In addition, categorical perception was observed for stimuli with no ready analogs in speech such as musical chords (Blechner, Reference Note 4; Locke & Kellar, 1973; Zatorre & Halpern, 1979) and the flicker-fusion threshold of visual stimuli (Pastore et al., 1977). Such findings, combined with others suggesting that nonhuman mammalian species, such as chinchillas (Kuhl, 1981; Kuhl & Miller 1975; 1978) and monkeys (Kuhl and Padden, 1982; 1983) appeared to show categorical perception, undermined the view that categorical perception is unique to speech processing. (For more extensive discussion of issues, see Harnad, 1987.)

Proponents of a specialized mode of speech processing offered other evidence of differences in speech and nonspeech processing. For example, there are marked differences in the rate at which speech and nonspeech sounds can be processed. Liberman et al. (1967) reported that listeners can process speech at rates of 25 to 30 phonetic segments per second. Other reports have suggested that, with some difficulty, adults are capable of processing 400 words per minute (Orr, Friedman, & Williams, 1965). By comparison, processing 30 nonspeech sounds per second seems to be beyond the temporal resolving capacity of the human ear (Miller & Taylor, 1948). In fact, human listeners cannot assign the correct ordering to a sequence of different nonspeech sounds occurring at a rate of four sounds per second (Warren, 1974; Warren, Obusek, Farmer, & Warren, 1969). Evidently, the kind of overlap in phonetic segments due to coarticulation contributes to listeners' ability to process speech information at higher transmission rates.

Another argument for specialized processing mechanisms derives from context-effects observed in speech processing. Miller and Liberman (1977) demonstrated that the interpretation of the acoustic information relevant to a stop/glide distinction (i.e., [b] versus [w]) varies with speaking rates. They generated a continuum ranging from [ba] to [wa] by gradually changing the duration of formant transitions between the initiation of these sounds and the vowel. Then they examined the consequences of increases and decreases in speaking rates by varying the overall duration of the syllables. Shorter syllables are characteristic of fast speaking rates, whereas longer syllables are characteristic of slower speaking rates. The point along the formant transition duration continuum at which listeners perceived the syllable to shift from [ba] to [wa] varied systematically with speaking rate. For instance, a formant transition duration that indicated [wa] at a fast speaking rate indicated [ba] at a slow speaking rate. Miller and Liberman argued that such effects were the result of specialized perceptual mechanisms that compensate for changes in speaking rates. However, this interpretation was challenged by Pisoni, Carrell, and Gans (1983), who demonstrated similar effects in the perception of nonspeech tone analogs to the Miller and Liberman stimuli. Pisoni et al. argued that a general auditory processing mechanism could account for both the speech and nonspeech findings.

Other investigations hinted at possible differences in speech and nonspeech processing. One experimental manipulation involved presenting a particular portion of a speech syllable in one ear, such as an isolated third formant transition, and the rest of the syllable in the other ear (Liberman, Isenberg, & Rakerd, 1981). Under these circumstances, listeners report hearing both a complete syllable and a chirp or tone corresponding to the isolated third formant transition. Liberman et al. dubbed this phenomenon "duplex perception" and argued that it was proof that these signals are perceived simultaneously as speech and nonspeech, suggesting two different processors. Duplex perception was attributed to the speech mode taking precedence over the auditory mode in interpreting the acoustic signal (Whalen & Liberman, 1987). Various stimulus or procedural manipulations were shown to independently affect the speech and nonspeech percepts (Bentin & Mann, Reference Note 3; Nygaard, 1993; Nygaard & Eimas, 1990). In sum, these findings seem to suggest the existence of separate processors for speech and nonspeech signals. However, the interpretation of findings on duplex perception has been challenged on a number of grounds (Fowler, 1990; Fowler & Rosenblum, 1990; Hall & Pastore, 1992; Nusbaum, Schwab, & Sawusch, 1983; Pastore, Schmeckler, Rosenblum, & Szczesiul, 1983). For instance, Pastore (1983) found duplex perception for musical chords when one note is played in one ear and the other two notes are played to the opposite ear. Moreover, the claim that the speech mode preempts the auditory mode is challenged by demonstrations of comparable phenomena in the processing of nonspeech stimuli such as slamming doors (Fowler & Rosenblum, 1990) and musical chords (Hall & Pastore, 1992). Clearly, more research is required to understand the implications of duplex perception for speech processing mechanisms.

Perhaps, the strongest evidence supporting the existence of different modes of speech and nonspeech perception comes from studies of sinewave speech, which is produced by replacing the formants of syllables with frequency-modulated sinewaves that follow the center frequency of the formants (Bailey, Summerfield, & Dorman, Reference Note 2; Best, Morrongiello, & Robson, 1981; Remez, Rubin, Pisoni, & Carrell, 1981). Listeners process such signals in distinctive ways depending on whether they are told that they are hearing poor quality synthetic speech or a series of beeps or tones (Remez et al., 1981). Because the acoustic information is identical, regardless of the instructions to the subjects, differences in processing must be due to their expectations about the stimuli (i.e., whether they are speech or nonspeech). The findings suggest that listeners may utilize different modes of processing in the two situations. In contrast to studies of duplex perception, listeners do not hear sinewave speech simultaneously as speech and nonspeech. Indeed, once listeners are instructed to treat the stimuli as speech, they cannot return to a nonspeech mode of processing these stimuli (Bailey et al., Reference Note 2).

In short, despite years of research, no firm resolution has yet been achieved regarding whether there is a special mode of perception for processing speech. In particular, neither categorical perception nor duplex perception serve to unequivocally support the presence of a speech-specific processing mode.

Selective Adaptation • Early debates about speech perception mechanisms focused on information in the acoustic signal. However, during the 1970s the emphasis shifted to the structure of the perceptual system itself. This change was prompted by discoveries in other sensory domains suggesting that certain cortical cells might act as feature detectors in responding to input from sensory systems. The chief example was Hubel and Weisel's (1965) discovery that certain cells in the visual cortex respond differentially to basic visual properties, such as the orientation of lines and edges. These findings prompted speculation that feature detectors might play a significant role in explaining how phonemes are extracted from the acoustic signal (Abbs & Sussman, 1971; Liberman, 1970; Stevens, 1972).

Eimas and his colleagues noted similarities in linguistic descriptions of phonetic features and descriptions offered for aspects of visual processing, such as color perception (Eimas, Cooper, & Corbit, 1973; Eimas & Corbit, 1973). Specifically, phonetic feature accounts pointed to binary oppositions between similar speech sounds (e.g., [b] and [p] share all phonetic features, except for voicing, where they have contrasting values, [b] being voiced and [p] voiceless). Similarly, binary opposition is prominent in descriptions of color vision, where the underlying sensory mechanisms are said to be organized as opponent processes: black/white, blue/yellow, and red/green (Hurvich & Jameson, 1969). Prolonged stimulation of one member of an opponent process pair temporarily depresses its sensitivity to the stimulus while leaving the other member of the pair unaffected, changing the point at which stimulation of the two pair members cancel each other out. Thus when shown a previously neutral surface, the nonstimulated member will have a greater responsiveness to the sensory information, producing a negative after image. Eimas and Corbit (1973) hypothesized that the same might hold for speech sounds with opposite values on some phonetic feature dimension, such as voicing. They devised a selective adaptation paradigm to test this possibility. Using a voicing continuum ranging from [ba] to [pa], they determined the phonetic category boundary for each of their subjects, then exposed them repeatedly to a stimulus from one of the two endpoints of the continuum. Afterwards, subjects were tested once again for the location of their phonetic category boundaries. Significant shifts in the locus of the boundaries occurred. Subjects exposed to [ba] adaptors were more likely to label stimuli from the region of the original category boundary as *pa*, whereas those exposed to the [pa] adaptor were more likely to label the same stimuli as *ba*. Eimas and Corbit interpreted these results as evidence for the existence of feature detectors.

Results of another experiment in the same study suggested that these detectors might be specialized for detecting phonetic features. Eimas and Corbit used a new set of adaptors for the same voicing contrast, a voiced sound [da] and a voiceless sound [ta], but they assessed the effects of adaptation on the original [ba]-[pa] stimulus continuum. The [da] adaptor produced the same effect as the original [ba] adaptor; likewise, the [ta] adaptor produced the same effect as the original [pa] adaptor. Eimas and Corbit concluded that selective adaptation effects were not due to a simple response bias to label stimuli from the boundary region as *pa* after repeated exposure to the [ba] endpoint (or vice versa for repeated exposure to the [pa] endpoint). Instead, they interpreted their findings as evidence for a phonetic feature detector of voicing. Subsequently, selective adaptation effects were obtained for other types of phonetic feature contrasts, such as place of articulation (Cooper & Blumstein, 1974). Moreover selective adaptation effects were greater when adaptors were prototypical instances of a phonemic category rather than less good instances (Miller, Connine, Schermer, & Kluender, 1983; Samuel, 1982).

Phonetic feature detectors appeared to provide an account of a variety of speech perception phenomena such as categorical perception. Furthermore, by assuming that different sets of acoustic cues map to the same phoneme detector, listeners' abilities to extract the same phonetic segment from different phonetic contexts despite its acoustic variability could be explained. However, subsequent research showed the phonetic feature detector model to be implausible. First, selective adaptation effects did not transfer across syllable position or vowel context (Ades, 1974; Cooper, 1975; Sawusch, 1977). Thus, adapting to [æd] or to [di] did not produce boundary shifts on a continuum from [bæ] to [dæ]. Such findings ruled out the possibility that the same feature detector detects a particular phonetic property in all contexts, suggesting instead that separate detectors were needed for each context. Second, detectors were shown to respond to acoustic, rather than phonetic properties. Tartter and Eimas (1975) reported that adaptation with nonspeech chirp stimuli was sufficient to produce significant shifts of phonemic category boundaries. Also, Sawusch and Jusczyk (1981) found that when listeners were exposed to an adaptor [spa] that was acoustically more similar to one end of a continuum [ba] yet phonetically similar to the other end of the continuum [pa], adaptation effects followed the acoustic properties of the stimuli. Thus, adaptation effects are more dependent on acoustic than phonetic characteristics of the adaptor.

However, the notion that feature detectors of any sort are involved in speech processing was called into question by Remez (1980). He created a stimulus series with [ba] as one endpoint of a continuum and a nonspeech buzz as the other endpoint. Adaptation with either endpoint resulted in a significant shift of the locus of the category boundary on this continuum. The fact that selective adaptation effects were found for an artificial continuum of this sort, with no known correspondence to any phonetic dimension, cast serious doubt on claims that selective adaptation effects reflect the action of detectors mediating acoustic-phonetic correspondences.

Normalization • In a classic study, Ladefoged and Broadbent (1957) manipulated the perceived dimensions of a talker's vocal tract by raising or lowering the first and second formants of a synthesized version of the sentence "*Please say what this word is*." They then followed this sentence with utterances of four target words: *bit*, *bet*, *but*, and *bat*. Although the acoustic characteristics of the carrier sentence varied across test trials, the acoustic properties of the target words remained unchanged. Listeners' judgments of the target words varied with their perceptions of the talker's voice. Ladefoged and Broadbent concluded that the carrier phrase allowed listeners to calibrate the vowel space of the perceived talker on a particular trial, leading them to adjust their interpretations of the target words accordingly. The view that listeners somehow estimate the typical vowel space of a given talker on the basis of a small speech sample received support in other investigations (Gerstman, 1968; Lieberman, Crelin, & Klatt, 1972; Shankweiler, Strange, & Verbrugge, 1977). However, this view was challenged by demonstrations that providing listeners with examples of a given talker's vowels did not significantly improve their identification of words with similar vowels (Shankweiler et al., 1977; Verbrugge et al., 1976).

An alternative view of normalization assumes that there are invariant acoustic features specifying the identity of particular phonemes (Blumstein & Stevens, 1980; Fant, 1960; Stevens, 1960; Stevens & Blumstein, 1978). This view postulates that listeners focus on the invariant features and ignore indexical information (i.e., information relating to the articulatory characteristics of different individuals): Information having to do with a particular talker's voice is stripped away from the signal, and invariant acoustic cues to phonetic segments are used in recognizing the occurrence of particular words (Nearey, 1989; Syrdal & Gopal, 1986). One suggestion as to how normalization may be accomplished in this way is that properties of the speech signal such as pitch could be used to guide the interpretation of formant frequencies produced by different talkers (Suomi, 1984). Furthermore, Sussman (1986) has proposed that innately specified neuronal cell assemblies may encode both absolute and relative formant frequencies. Connections between these cell assemblies and higher order ones could serve to eliminate information related to vocal tract size, thus allowing the derivation of invariant properties that are necessary to achieve normalization.

This view assumes that listeners rely on a frequency-based recalibration of the signal. However, others have argued that normalization occurs, not by extracting out a static set of acoustic invariants, but by recovering the underlying articulatory dynamics from the speech signal (Verbrugge & Rakerd, 1986). The latter claim is based on research with so-called "silent-centered syllables." These are CVC syllables in which the middle 60% of the signal is removed, leaving only the beginning and ending formant transitions with a silent period in between. Listeners correctly identify the vowels from such patterns, even when a range of tokens from different males and females is used. Verbrugge and Rakerd conclude that "vowels can be characterized by higher-order variables (patterns of articulatory and spectral change) that are independent of a specific talker's vocal tract dimensions" (p. 56).

Although these various accounts of normalization differ in the mechanisms that are inferred to subserve this function, they agree that listeners cope with talker variability by stripping such information from the acoustic signal. In other words, talker-specific information is discarded to perceive the linguistic message. However, talker variability affects the accuracy with which speech sounds are perceived, both in terms of the accuracy (Creelman, 1957) and speed (Mullennix, Pisoni, & Martin, 1989; Summerfield, Reference Note 21; Summerfield & Haggard, Reference Note 22) with which words are recognized. Consideration of the latter findings have led some to question whether the successful recovery and recognition of words from fluent speech requires that talker-specific characteristics be stripped from the acoustic signal (Goldinger, 1996; Houston & Jusczyk, 2000; Jusczyk, 1993; Pisoni, 1992). We will consider this issue further in our discussion of spoken word recognition by adults and infants.

The Development of Speech Perception Capacities: Early Studies

Research on infants provides another kind of window on the nature of speech perception capacities, allowing investigation of these capacities before the point at which experience with a particular language has had a significant influence on their functioning. When such studies began, it was not certain that infants even had any ability to discriminate differences between speech sounds. The issues that motivated early investigations in this area were similar to those that concerned researchers studying the speech perception abilities of adult listeners at that time. What kinds of phonetic distinctions can infants perceive, and what mechanisms underlie their abilities? Are the underlying mechanisms specific to speech perception, or more generally used in auditory processing? What capacities do infants have for coping with variability in speech? In addition to these issues, research with infants provided the opportunity to explore how experience with a particular language affects the development of speech perception capacities.

Infants' Discriminative Capacities • When Eimas and his colleagues began investigating infant speech perception capacities, two questions motivated them (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). First, can infants discriminate a minimal phonetic contrast between two consonants? Second, if so, is their perception of such a contrast categorical? Eimas et al. used the high-amplitude-sucking procedure to test English-learning 1- and 4-mo-olds on voicing contrasts from a continuum that ranged from [ba] to [pa]. The critical dimension varied was voice-onset-time (VOT), which refers to the moment at which the vocal cords begin to vibrate relative to the release of closure in the vocal tract, which for these sounds is the release of the lips. For English-speaking adults, if the vocal chord vibration begins within 25 msec of the release of the lips, the sound is heard as [b], whereas if it is delayed by more than 25 msec, it is perceived as [p] (Abramson & Lisker, Reference Note 1). Eimas et al. presented pairs of syllables that differed by 20 msec in VOT. For one group (Between Category), infants were familiarized with a syllable from one phonemic category (e.g., [ba]), then presented with another syllable from the other phonemic category (e.g., [pa]). For another group (Within Category), both the familiarization syllable, and the subsequent test syllable came from the same phonemic category (i.e., either two different [ba] stimuli or two different [pa] stimuli). Finally, a third group (Control) continued to hear the original familiarization syllable for the entire test session. Relative to the Control group, only infants in the Between-Category group significantly increased their sucking rates in response to the test syllable. Thus, these findings demonstrated that infants could detect VOT differences, but that, like adults, their discrimination is categorical. That is, they only discriminate VOT differences between stimuli from different phonemic categories. Eimas et al. concluded that categorical perception is part of infants' biological endowment for acquiring language.

Eimas et al.'s findings generated considerable interest in the range of infants' speech perception abilities. Initially, many studies explored infants' abilities to discriminate different types of phonetic contrasts. In addition to discriminating voicing contrasts between consonants, infants were also found to perceive consonant differences involving place of articulation (Eimas, 1974; Levitt, Jusczyk, Murray, & Carden, 1988; Moffitt, 1971; Morse, 1972) and manner of articulation (Eimas, 1975; Eimas & Miller, 1980b; Hillenbrand, Minifie, & Edwards, 1979). Although such studies typically contrasted phones in the initial portions of the syllables, infants were also shown to detect some contrasts at the ends of syllables (Jusczyk, 1977) and in the middle of multisyllabic utterances (Jusczyk, Copan, & Thompson, 1978; Jusczyk & Thompson, 1978; Karzon, 1985; Williams, Reference Note 24). Infants were also shown to discriminate vowel contrasts (Kuhl & Miller, 1982; Swoboda, Kass, Morse, & Leavitt, 1978; Swoboda, Morse, & Leavitt, 1976; Trehub, 1973). As with adults, infants' discrimination of vowel contrasts appears to be continuous in that they display some ability to distinguish two different vowel tokens from within the same phoneme category (Swoboda et al., 1976).

Many questions were raised about the role of experience with language in the development of these capacities. To what extent is prior listening experience required for infants to discriminate these contrasts? Two different sorts of investigations suggested that prior experience is not a significant factor in discriminating phonetic contrasts during the first few months of life. First, several investigations explored speech sound contrasts that did not occur in the language spoken in the infant's environment (Lasky, Syrdal-Lasky, & Klein, 1975; Streeter, 1976; Trehub, 1976). These investigations indicated that even when infants had no prior experience with a particular phonetic contrast, they could discriminate it. For example, Streeter (1976) found that Kikuyu-learning 1- to 4-mo-olds distinguished a [ba]-[pa] contrast, despite the absence of this contrast in Kikuyu. A second line of investigation suggesting that discrimination abilities do not depend on a long period of prior exposure to a language involved studies with newborns. Bertoncini, Bijeljac-Babic, Blumstein, and Mehler (1987) showed that infants only a few days old can discriminate phonetic contrasts.

Considering the difficulty that adults have in discriminating contrasts outside of their native language (Miyawaki et al., 1975; Trehub, 1976), infants' abilities are quite remarkable. The findings suggest that infants' capacities for detecting speech sound contrasts are better than those of adults. Thus, the picture that emerged from these early studies is that infants are born with excellent abilities to discriminate phonetic contrasts.

Mechanisms Underlying Infants' Discriminative Capacities • When Eimas et al. (1971) first reported their findings, categorical perception was still believed to be specific to speech processing. Thus, the findings appeared to suggest that infants are innately endowed with specialized speech processing abilities. Indeed, Eimas (1974, 1975) found that infants' discrimination of nonspeech sounds such as isolated formant transitions is not categorical, even though their discrimination of the same information in speech is categorical. Thus, these findings reinforced the view that infants are born with specialized speech processing mechanisms. However, shortly after the first reports of categorical perception of certain nonspeech sound contrasts by adults (Miller et al., 1976; Pisoni, 1977), new studies demonstrated that infants, too, discriminated nonspeech contrasts categorically (Jusczyk, Pisoni, Walley, & Murray, 1980). Jusczyk et al. argued that strong parallels between how infants processed temporal order differences in nonspeech sounds and how they processed VOT differences in speech sounds, pointed to a common auditory processing mechanism underlying the discrimination of both kinds of contrasts (see also Jusczyk, Rosner, Reed, & Kennedy, 1989).

After the discovery that categorical perception is not specific to speech, the grounds for postulating specialized speech processing mechanisms changed. Proponents of specialized processing mechanisms pointed out that even 2-mo-olds compensate for speaking rate differences in their perception of phonetic contrasts (Eimas & Miller, 1980a; Miller & Eimas, 1983). Thus, infants' discrimination of acoustic cues for the stop/glide contrast between [b] and [w] varies with speaking rate changes, much as does adults' perception of this contrast. This finding led Eimas and Miller (1980) to claim that infants possess specialized processing mechanisms to compensate for speaking rate changes. However, Jusczyk, Pisoni, Fernald, Reed, and Myers (Reference Note 11) found that infants displayed the same pattern in responding to nonspeech sounds with similar acoustic characteristics. Thus, they argued that a general auditory processing mechanism underlies infants' discrimination performance with the speech and nonspeech stimuli.

More recently, Eimas and Miller (1992) demonstrated that 3- to 4-mo-olds show a duplex perception effect in speech processing. In their experiment, an isolated third formant transition critical to distinguishing [da] from [ga] was presented to one ear, and the rest of the syllable to the opposite ear. Not only did infants discriminate the patterns corresponding to [da] and [ga] in this situation, but they also discriminated the patterns when the intensity of the third formant transition was greatly attenuated. In contrast, when the attenuated third formant transitions were presented by themselves (without the rest of the syllable in the opposite ear) infants' discrimination of these differences was significantly poorer. This finding parallels one reported by Bentin and Mann (Reference Note 3) with adults (see also Nygaard & Eimas, 1990). Because infants discriminated the syllabic stimuli, even without a perceptible difference in the formant transitions (i.e., when they were attenuated), Eimas and Miller inferred that infants integrated the information from two different distal sources to form a unified phonetic percept that served as the basis for discrimination. Comparable studies have not yet been undertaken with nonspeech analogs of the stimuli used in duplex perception studies with infants, so whether this phenomenon is specific to infants' processing of speech sounds remains to be seen.

The issue of whether infants' basic speech perception capacities are specific for processing language or more widely applicable in auditory processing has not been satisfactorily resolved. In some sense, the focus of the field has moved to investigating how these basic capacities, whether or not they are specific to speech perception, are used in acquiring language.

How Infants Handle Variability in Speech • Given young infants' sensitivity to subtle differences that mark phonetic contrasts, variability from different talkers' productions of the same word could pose problems for their recognition of the word. However, even at a young age, infants appear to cope with variability in different talkers' productions of the same syllables. Kuhl and Miller (1982; Kuhl, 1976) first explored this issue by investigating 1- to 4-mo-olds' abilities to perceive vowel contrasts. In one experiment, they used different tokens of the vowels [a] and [i], which varied in whether they were produced with rising or falling pitch. Initially, infants were exposed to the different tokens of one vowel and then shifted to the different tokens of the other vowel. Despite the fact that the pitch characteristics of the tokens varied, the infants still discriminated the vowel tokens. Kuhl and Miller concluded that even 1-mo-olds had the means to generalize across the different tokens of the same vowel to perceive the phonetic contrast.

Subsequently, Kuhl (1979, 1983) provided even more impressive demonstrations that 6-mo-olds handle variability in productions of the same syllable by different talkers. Infants were initially taught to respond to a vowel contrast between [a] and [i] produced by a single talker. After they had responded appropriately to this distinction, Kuhl (1979) gradually introduced tokens of the vowels produced by other talkers into the stimulus set. The final set had tokens from three different talkers, including males and females. Even with this degree of variability, infants continued to respond to the vowel contrast, leading Kuhl to claim that they have some capacity for perceptual normalization. Kuhl (1983) extended these findings to a more subtle vowel contrast ([a] and [ɔ]) in another study with 6-mo-olds.

Jusczyk, Pisoni, and Mullennix (1992) found that even 2-mo-olds have some capacity to cope with talker variability in perceiving phonetic contrasts. Infants discriminated a contrast between *bug* and *dug*, even when exposed to tokens from 12 different talkers (six male and six female). However, the variability did affect infants' encoding and subsequent recall of the sounds. In particular, when Jusczyk et al. inserted a two-minute delay between exposure to the tokens of one of the words and testing on the tokens of the other word, infants showed no discrimination of the words. This failure to discriminate the tokens from multiple talkers after the delay contrasted with evidence of discrimination shown by another group who were familiarized with only a single token and tested on a different one from the same talker after a two-minute delay. Thus, although 2-mo-olds have some ability to cope with talker differences, dealing with these differences bears a cost in their ability to encode and remember speech information.

As noted earlier, in addition to dealing with talker variability, infants have some capacity to handle variability induced by changes in speaking rate (Eimas & Miller, 1980a; Miller & Eimas, 1983). Thus, although young infants possess excellent capacities for discriminating speech sounds, they also cope with variability introduced by talker and speaking rate differences.

Native Language Influences on Underlying Perceptual Capacities • Under 6 mo, infants seem to discriminate phonetic contrasts that do not occur in their native language. In this respect, young infants differ from adults, who have difficulty in perceiving non-native language contrasts without considerable training (Logan, Lively, & Pisoni, 1989; Miyawaki et al., 1975; Pisoni, Aslin, Perey, & Hennessy, 1982; Strange & Jenkins, 1978; Trehub, 1976). Thus, language experience does eventually affect perceptual capacities. When, in development, do changes arise in discriminating non-native language contrasts?

Werker and Tees (1984) found that sensitivity to non-native speech contrasts begins to decline in the latter half of the first year. They tested English-learning 6- to 8-mo-olds' perception of three phonetic contrasts: an English [ba]-[da] distinction, a Hindi [ta]-[t^a] distinction, and a Nthlakapmx [k'i]-[q'i] distinction. Despite the fact that neither of the latter two contrasts occurs in English, the infants discriminated all of the contrasts. However, only some of the 8- to 10-mo-olds discriminated the two non-native contrasts, although all of them discriminated the English contrast. By 10 to 12 mo, the infants discriminated only the English contrast. Experiments with Hindi-learning and Nthlakapmx-learning 10- to 12-mo-olds showed that these infants had no difficulty discriminating the contrast from their respective native languages. Hence, the decline in English-learners' ability to discriminate the Hindi and Nthlakapmx contrasts seems to be due to their lack of experience with these contrasts. Research by Werker and Lalonde (1988) with a different Hindi contrast further confirmed the developmental timecourse of English-learners' declining sensitivity to such contrasts.

Other investigations provided further evidence of a decline during the second half of the first year in infants' abilities to perceive certain non-native language contrasts. However, the decline in sensitivity was found not to extend universally to all non-native contrasts. Best, McRoberts, and Sithole (1988) found that English learners do not decline in their ability to perceive a contrast between two Zulu click sounds. Nor do English-learners seem to have difficulty with a place of articulation distinction for different ejective sounds in Ethopian, whereas 10- to 12-mo-olds show a significant decline in discriminating a lateral fricative voicing distinction from Zulu (Best, Lafleur, & McRoberts, 1995). Also, Japanese-learning 6- to 8-mo-olds discriminate the English [ra]-[la] contrast, but Japanese 10- to-12-mo-olds do not (Tsushima et al., Reference Note 23). Hence, lack of direct experience is an important factor in the decline of sensitivity to non-native contrasts, but it is certainly not the sole factor.

Other explorations of the role of linguistic input have focused on how such experience might contribute to the development of phonemic categories. It has been claimed that linguistic experience leads infants to develop representations of prototypical instances of native language vowel categories (Grieser & Kuhl, 1989; Kuhl, 1991; Kuhl et al., 1997; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). This claim is based on asymmetries in the perception of certain vowel contrasts (i.e., "perceptual magnet effects"). Specifically, listeners are less likely to discriminate contrasts between a prototypical instance and another exemplar of the category than to discriminate the same exemplar from a more peripheral instance of the category (Kuhl, 1991).

Models of Developmental Changes in Phoneme Perception • The favored explanation for why sensitivity declines for some non-native contrasts, but not others, has to do with how the contrasts relate to the native language (Best, 1995; Eimas, 1991; Flege, 1995; Werker, 1991). Best (1993, 1995) has attempted to account for the course of developmental change with her Perceptual Assimilation Model (PAM). The model is founded on the assumption that, all other things being equal, the better the mapping of non-native phones to distinct phonemic categories in the native language, the easier they should be to discriminate. Thus, the model predicts that non-native contrasts that map to two different native language phonemic categories should remain easy to discriminate. Cases in which both non-native phones fall outside native language categories should prove more difficult to discriminate, whereas ones that map to a single phonemic category should be most difficult to discriminate and, therefore, most likely to decline as the native language is acquired (Best, 1995).

Werker (1991) has proposed a different account of developmental changes in infants' perception of non-native contrasts. Her Model of the Development of Phonemic Perception attributes changes in sensitivity to non-native contrasts to a re-organization of perceptual processes, which is promoted by the beginnings of a phonological system in the receptive lexicon. Although this view is similar to Best's, Werker has also implicated the role of cognitive factors in promoting the reorganization (Lalonde & Werker, 1995), noting correlations between performance on certain cognitive and perceptual tasks and shifts in sensitivity to non-native contrasts. From this perspective, the reorganization of perceptual processes during the learning of native language phoneme categories is seen as just one example of an initial, biologically based categorization ability coming under the mediation of cognitive control. The common factor promoting cognitive control, in these different situations at this point in development, stems from a need to integrate disparate sources of information.

A different account is offered by Kuhl (1993), who proposed the Native Language Magnet Theory (NLM). Its starting assumption is that infants' innate endowment provides the ability to categorize speech sounds into groupings that are separated by natural boundaries. The natural boundaries are not the result of any specialized phonetic processing mechanisms but, rather, stem from general auditory processing mechanisms. At some point before 6 mo, infants develop something beyond categories provided by the auditory mechanisms. For example, the prototype for a vowel category in the native language reflects the distributional properties of different instances of the vowel that infants have heard (Kuhl et al., 1997). This theory attributes changes in the perception of non-native contrasts to changes in the nature of the perceptual space that come about through the development of native language prototypes. Prototypes act as perceptual magnets causing certain boundaries to disappear as the space is reorganized to reflect the native language categories. The disappearance of certain boundaries may allow the developing magnets to pull in sounds that were discriminable by infants at an earlier point in development.

The predictions of NLM about which contrasts may show significant declines in sensitivity are much less specific than those of PAM. In addition, many questions have been raised about the source and robustness of these perceptual asymmetries in adults (Iverson & Kuhl, 1995; Lively & Pisoni, 1997; Sussman & Lauckner-Morano, 1995; Thyer, Hickson, & Dodd, 2000). Moreover, infants display similar asymmetries for vowel contrasts not present in their native language (Polka & Bohn, 1996). Consequently, it is not clear whether these perceptual asymmetries are the result of experience with language or attributable to the characteristic organization of the human auditory processing system.

Spoken Word Recognition: Models and Issues

As we have discussed, early work in the field of speech perception was dominated by research on the discrimination and categorization of phonetic segments. In the 1970s, a new emphasis emerged that focused on the processes and representations responsible for the perception of spoken words. This focus grew from the recognition that a comprehensive theory of spoken language comprehension—especially a theory of how the listener perceives fluent speech—must account for more than the perception of individual consonants, vowels, and syllables. The spoken word became a primary object of scientific inquiry.

Research on the perception of printed words had already established itself as a highly active and theoretically productive area by the 1970s. But the theories that are developed to account for visual word recognition were inadequate as models of spoken language processing (e.g., Forster, 1976). In particular, early models of visual word recognition had little to say about how the unfolding information in an acoustic stream of temporally distributed information is mapped on to one of thousands of representations in memory.

One of the first and most influential models solely devoted to accounting for the process of spoken word recognition was Marslen-Wilson's Cohort theory (Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Welsh, 1978; see also Cole and Jakimik, 1980). Indeed, Marslen-Wilson's empirical and theoretical work throughout the decade of the 1970s did much to establish the field of spoken word recognition. Although others (e.g., George Miller, John Morton, and Richard Warren) had already made important contributions to our understanding of the perception of spoken words and sentences, Marslen-Wilson's

Cohort theory became the focus of and catalyst for a whole generation of speech researchers concerned with the processes and representations subserving the recognition of spoken words. The earliest version of Cohort theory presaged many of the issues that were to occupy research on spoken word recognition for years to come. Indeed, many questions that currently dominate research in the field either have their roots in the early cohort theory or are attempts to correct shortcomings of the original model.

Models of Recognition • To its credit, the field of research devoted to spoken word recognition has been extensively driven by the development, testing, and refinement of its theories. Thus, to understand the nature and substance of the issues that occupy the field today, one must have some appreciation for the models that have inspired current empirical research. We briefly review four current models of spoken word recognition: Cohort, Trace, Shortlist, and the Neighborhood Activation Model. Table 1 provides a summary of the important features of each model.

Cohort. The Cohort model holds a special place in the field because it was the most influential early theory devoted exclusively to explaining the process of spoken (in contrast to visual) word recognition. The model has evolved considerably over the years, and it is difficult to speak exclusively about one single model under the rubric "Cohort." Thus, the reader should bear in mind the caveat that the most recent ancestors of the original Cohort model bear little resemblance to their progenitor.

Cohort theory (Marslen-Wilson & Welsh, 1978) proposes that input in the form of a spoken word activates a set of similar items in memory, referred to as the word-initial cohort. The cohort consists of all spoken words known to the listener that begin with the initial segment or segments of the input word. For example, the word *elephant* may activate in memory the cohort members *echo*, *enemy*, *elder*, *elevator*, and so on. Once activated, the cohort is winnowed based on both bottom-up (acoustic-phonetic) and top-down (syntactic and semantic) information until a single candidate remains, at which time recognition is achieved. In early versions of the theory, activation is a function of an exact match between acoustic-phonetic information at the beginnings of words and representations in memory. According to the theory, acoustic-phonetic information is solely responsible for establishing the cohort, and the recognition system tracks the input so closely that minimally discrepant featural information is sufficient to remove an inconsistent candidate from play (Warren & Marslen-Wilson, 1987, 1988). In addition, both early and more recent versions of the theory propose a specific and restricted competition process. Words in the cohort are not assumed to affect the activation levels of one another: The effect of a competitor on its target arises through its mere presence in the cohort as a candidate for recognition. For example, recognition of the input word *elephant* must wait until the word diverges—or becomes unique—from its competitor, *elevator*. (However, see Marslen-Wilson, Tyler, Waksler, & Older, 1994, for a discussion of inhibition among inflected words sharing the same base form.)

Cohort theory has been very successful in focusing attention on the temporal dynamics of spoken word recognition. In particular, the theory provides an elegant account of the earliness of word recognition, stating that spoken words may be identified well before their offsets if overlapping competitors are not active. The theory also proposes an explicit mechanism for the effects of context on word recognition: Top-down information may speed recognition by eliminating competitors from the cohort. The early theory's strong emphasis on exact match between input and representation, its rejection of sublexical levels of representation (see below), and its lack of computational specificity are among its notable shortcomings. Although many of these concerns have been addressed by a more recent, computationally explicit version of the theory (Gaskell & Marslen-Wilson, 1997, 1999) that adopts a distributed representational format for modeling the mapping of form onto meaning, the newer theory still preserves the notion of lexical competition without lateral inhibition and eschews intermediate sublexical representations between feature and word.

Trace. The Trace model of spoken word recognition (McClelland & Elman, 1986) is an interactive-activation, local connectionist model of spoken word recognition. The Trace model consists of three levels of primitive processing units corresponding to features, phonemes, and words. These processing units have excitatory connections between levels and inhibitory connections among levels. These connections serve to raise and lower activation levels of the nodes depending on the stimulus input and the activity of the overall system.

The hallmark of Trace is its interactivity. By passing activation between levels, the model serves to confirm and accentuate evidence in the input corresponding to a given feature, phoneme, and word. For example, evidence consistent with voicing (as in the consonants /b/, /d/, or /g/) will cause the voiced feature at the lowest level of the model to become active, which in turn will pass its activation to all voiced phonemes at the next level of units, which will in turn activate words containing those phonemes. Moreover, through lateral inhibition among units within a level, winning hypotheses may

TABLE 1. Models of spoken word recognition.

	Cohort	Trace	Shortlist	NAM & PARSYN
Activation	Constrained	Radical	Radical	Radical
Units and levels (arrows indicate direction of information flow)	Words ↑ Features	Words ↑ ↓ Phonemes ↑ ↓ Features	Words ↑ Phonemes	NAM: Word decision units ↑ Patterns PARSYN: words ↑ ↓ Allophone patterns ↑ Allophone input
Lexical competition via lateral inhibition	No	Yes	Yes	NAM: no PARSYN: yes
Sublexical-to-lexical interaction (bottom-up)	Facilitative and inhibitory	Facilitative	Facilitative	NAM: N/A PARSYN: facilitative
Lexical-to-sublexical interaction (top-down)	No	Facilitative	No	NAM: no PARSYN: inhibitory
Distinguishing features	●Focus on time-course of recognition ●Interactivity	●Highly interactive, simple processing units ●Computationally explicit ●Attempts to account for broad range of phenomena	●Lexical competition among a restricted candidate set ●No feedback from lexical to sublexical level ●Focus on lexical segmentation	●NAM: quantitative account of lexical competition ●PARSYN: computational account of probabilistic phonotactics

easily come to dominate other competing units that are also momentarily consistent with the input. Thus, evidence for the word *cat* at the lexical level will cause *cat's* unit to send inhibitory information to similar, competing lexical units (e.g., for *pat*), helping to ensure that the best candidate word will win the competition for recognition.

Trace has been enormously influential, owing primarily to its computational specificity and to the broad range of phenomena for which it attempts to account (Norris, 1994). Simulations of the model are readily available, making direct tests of the behavior of the model relatively easy to conduct. However, Trace incorporates a decidedly questionable architecture in which its system of nodes and connections are duplicated over successive time slices of the input, a rather inelegant (and probably psychologically implausible) means of dealing with the temporal dynamics of spoken word recognition.

Shortlist. Norris' (1994) Shortlist model, like Trace, is a connectionist model of spoken word recognition. In the first stage of the model, a "short list" of word candidates is derived that consists of lexical items that match the bottom-up speech input. In the second stage of processing, this abbreviated list of lexical items enters into a network of word units, much like the lexical level of Trace. Lexical units at this second level of processing compete with one another (via lateral inhibitory links) for recognition.

The Shortlist model is attractive for two primary reasons: First, the model attempts to provide an explicit account of segmentation of words from fluent speech via mechanisms of lexical competition. It is one of the first computationally explicit models that was purposefully designed to simulate effects of subsequent context on spoken word recognition, thereby attempting to account for the process by which individual words are extracted from the speech stream. Second, Shortlist improves on the highly unrealistic architecture of Trace, in which single words are represented by a plethora of identical nodes across time.

Recently, the Shortlist model has attracted considerable attention as the prime example of an autonomous model of recognition. Unlike Trace, Shorlist does not allow for top-down lexical influences on its phoneme units; flow of information between phoneme and word units is unidirectional and bottom-up. Thus, the Shortlist model embodies the notion, which has received some empirical support (Burton, Baum, & Blumstein, 1989; Cutler et al., 1987; McQueen 1991), that the processing of phonemes in the input is unaffected by—or autonomous of—top-down, lexical influences. This central tenet of Shortlist (and its companion model, Merge) has engendered a lively debate in the literature between the autonomist and interactionist positions (see Norris, McQueen, & Cutler, 2000; Samuel, 2000; and the accompanying responses). As yet, no unequivocal support for either side has emerged. However, Shortlist remains a particularly attractive alternative to the Trace model, primarily because of its more plausible architecture and its superiority in accounting for lexical segmentation in fluent speech.

Neighborhood Activation Model and PARSYN. Over the past few years, Luce and colleagues have devoted considerable effort to modeling the processes of activation and competition. According to their Neighborhood Activation Model (NAM; Luce & Pisoni, 1998; Luce, Pisoni, & Goldinger, 1990), stimulus input activates a set of similar sounding acoustic-phonetic patterns in memory. The more similar the pattern is to the input, the higher its activation level. Once the acoustic-phonetic patterns are activated, word decision units tuned to each of the patterns attempt to decide which pattern best matches the input. The word decision units compute probabilities for each pattern based on 1) the frequency of the word to which the pattern corresponds, 2) the activation level of the pattern (which depends on the pattern's match to the input), and 3) the activation levels and frequencies of all other words activated in the system. The word decision unit that computes the highest probability wins, and its word is recognized. In short, word decision units compute probability values based on the acoustic-phonetic similarity of the word to the input, the frequency of the word, and the activation levels and frequencies of all other similar words activated in memory.

NAM predicts that multiple activation has its consequences: Spoken words with many similar-sounding neighbors should be processed more slowly and less accurately than words with few neighbors. That is, NAM predicts effects of neighborhood density arising from competition among multiply activated representations of words in memory. This prediction has been confirmed in many studies: Words in high-density similarity neighborhoods are indeed processed less quickly and less accurately than words in low-density neighborhoods (Cluff & Luce, 1990; Goldinger, Luce, & Pisoni, 1989; Goldinger, Luce, Pisoni, & Marcario, 1992; Luce & Pisoni, 1998; Vitevitch & Luce, 1998; 1999).

Recently, Luce et al. (2000) have instantiated NAM in a more explicit processing model, called PARSYN. (The name PARSYN is a combination of the terms PARadigmatic and SYNtagmatic, which refer to neighborhood activation and phonotactic constraint, respectively.) PARSYN has three levels of units: 1) an input allophone level, 2) a pattern allophone level, and 3) a word level. Connections between units within a level are mutually inhibitory, with one exception: Links among allophone units at the pattern level are facilitative across temporal positions. Connections between levels are facilitative, also with one exception: The word level sends inhibitory information back to the pattern level, essentially quelling activation in the system once a single word has gained a marked advantage over its competitors.

PARSYN is designed to simulate the effects of both neighborhood activation and probabilistic phonotactics on the processing of spoken words. Effects of neighborhood density arise primarily from lateral inhibition at the lexical level. Effects of probabilistic phonotactics arise from activation levels of and interconnections among units at the allophone pattern level: Allophones that occur more frequently have higher resting activation levels. Also, allophones that frequently occur together will excite one another via facilitative links.

In addition, PARSYN was developed to overcome two significant shortcomings of NAM. First, PARSYN is an attempt to better account for the temporal dynamics of recognition, including the ebb and flow of neighborhood activation and phonotactic constraints. Second, PARSYN incorporates a sublexical level of representation missing in the original NAM. We return to the issue of sublexical representations in more detail below.

Summary. Trace, Shortlist, and PARSYN include an account of the competition process that contrasts sharply with both early and late versions of the Cohort model. Each of these three models assumes that multiple form-based representations of words compete directly for recognition. In particular, each model proposes that word units are connected via lateral inhibitory links, enabling a unit to suppress or inhibit the activations of its competitors (for empirical support for this claim, see McQueen, Norris, & Cutler, 1994). The degree to which a unit inhibits its competitors is proportional to the activation level of the unit itself, which is determined in large part by its similarity to the input. Competitor activation is also assumed to be a function of the degree of similarity of the competing words to the input.

Trace, Shortlist, and PARSYN all differ from Cohort theory in positing sublexical levels of representation. However, each model, to varying degrees, suffers from a significant weakness in terms of how it maps input onto these sublexical representations. In particular, the models rely on coding the acoustic-phonetic signal into either abstract phonetic features (in Trace) or phonemes (in Trace and Shortlist) that vary neither as a function of time or context. The input to the models ignores much of the contextual and temporal detail encoded in the signal. Although Trace allows for overlapping features in an attempt to capture effects of coarticulation, the features themselves remain unchanged by the context in which they occur. Whereas Shortlist holds out promise for more realistic input based on the output of a simple recurrent network, the model as implemented makes no use of context-dependent, subphonemic information in lexical processing. Although PARSYN's use of allophonic representations attempts to capture some context-dependency at the sublexical level, it too fails to make full use of the rich source of information embodied in the speech signal itself.

In defense of the models, their failure to capture contextually and temporally conditioned information in the signal is not inherent in their architectures; nothing in the models precludes them from using this information. However, the models' emphasis on processing dynamics over input may provide a somewhat a distorted picture of the activation-competition process. In particular, information in the signal itself may play a much more significant role in lexical discrimination than proposed by current versions of Trace, Shortlist, or PARSYN.

Current Issues in Spoken Word Recognition • We now turn to a discussion of a set of core issues that occupy much attention in current research and theory on spoken word recognition, all of which are related directly or indirectly to the original issues addressed by the Cohort model. These are 1) the nature of lexical activation and competition, 2) the nature of sublexical and lexical representations and their interactions, (3) the problem of lexical embeddedness, 4) segmentation of spoken words from fluent speech, and 5) representational specificity of spoken words.

Activation and Competition in Spoken Word Recognition • Virtually all current models of spoken word recognition share the assumption that the perception of spoken words involves two fundamental processes: activation and competition (see Luce & Pisoni, 1998; Marslen-Wilson, 1989; McClelland & Elman, 1986; Norris, 1994). That is, there is some consensus that the input activates a set of candidates in memory that are subsequently discriminated among. However, details of the activation and competition processes are still in dispute.

Activation. The nature of the activation process itself has been the source of much debate. Models of spoken word recognition fall into roughly two categories regarding their characterization of the activation process (see Table 1). Radical activation models (e.g., Trace, Shortlist, PARSYN) propose that form-based representations consistent with stimulus input may be activated at any point in the speech stream. For example, spoken input corresponding to *cat* may activate *pat* based on the overlapping vowel and final consonant, despite the fact that the two words differ initially. Of course, most radical activation models afford priority to *cat* in recognition process, primarily because of the relative temporal positions of the mismatch and overlap. Furthermore, lateral inhibition at the lexical (and sometimes prelexical) levels typically grants considerable advantage to representations overlapping at the beginnings of words. Nevertheless, radical activation models propose that any consistency between input and representation may result in some degree of activation.

In contrast, constrained activation models propose that form-based representations respond to only specific portions of the input, such as word beginnings (Marslen-Wilson, 1987; 1989; 1990) or strong syllables (Cutler, 1989; Cutler & Norris, 1988). Primary among this class of models is Cohort theory, which states the word-initial information has priority in activating the set of representations (i.e., the cohort) that will subsequently compete for recognition. According to early versions of Cohort theory, stimulus input corresponding to *cat* will not activate the representation for *pat*, owing to the mismatch of initial phonetic information. Although later postperceptual recovery processes may give the appearance that representations mismatching on initial segments enter into the recognition process, prelexical activation is exclusively controlled by overlapping information at the beginnings of words (see Marslen-Wilson, et al., 1996). Moreover, this prelexical activation serves to inhibit mismatching representations. Thus, membership in the activated competitor set (or cohort) is controlled by bottom-up, and not lateral, inhibition.

A fundamental assumption of Cohort theory's constrained activation framework is embodied in the minimal discrepancy hypothesis. According to this hypothesis, minimally inconsistent or discrepant information in the speech input is sufficient to exclude representations from the recognition process (Warren & Marslen-Wilson, 1987; 1988). Thus, word-initial featural information indicating the presence of a /k/, as in *cat*, would be sufficient to inhibit activation of all representations not beginning with /k/ (e.g., *pat*; see Marslen-Wilson, et al., 1996). The minimal discrepancy hypothesis also incorporates the notion of the divergence point. Recall that, according to the Cohort model, the precise moment at which a word is recognized occurs when a single word diverges from all other candidates in the lexicon. As a spoken word is processed in time, competitors are eliminated from the cohort by discrepant bottom-up information until only one candidate remains (Marslen-Wilson, 1989). Crucially, information occurring after the minimal discrepancy between target and competitor is predicted to have no demonstrable effect on lexical activation, even if this postdivergence information is consistent with the competitor.

Evidence in favor of the minimal discrepancy hypothesis comes from a series of gating experiments conducted by Warren and Marslen-Wilson (1987, 1988) in which they examined the role of coarticulatory information in listeners' ability to guess the identities of words based on their fragments. (In the gating paradigm, successively longer portions, or gates, of a spoken word are presented—e.g., *e-*, *el-*, *ele-*, *eleph-*, etc.—and subjects must guess the identity of the word based on each fragment.) They found that listeners were remarkably sensitive to fine acoustic-phonetic detail and need not wait until the ends of segments to correctly guess the identity of a word. Instead, listeners appeared to make "maximally efficient" use of temporally overlapping phonetic information. The implications of Warren and Marslen-Wilson's work are clear: Minimally discrepant information in the speech signal controls perceptual choice, ruling out alternatives (i.e., competitors) at the earliest possible point in time.

Evidence against the minimal discrepancy hypothesis has come, in part, from form-based rhyme priming studies. This research has investigated lexical activation of targets by primes that mismatch on word-initial information but overlap at the end (e.g., *cat* and *pat*). Facilitative effects of rhyme priming would presumably provide evidence that competitors may be activated in the absence of shared word-initial information. Connine, Blasko, and Titone (1993) found facilitative priming effects between rhyming nonword primes and real word targets, suggesting that overlapping word-initial information is not crucial for activation of competitors. (See also Connine, Titone, Deelman, & Blasko, 1997; Marslen-Wilson et al., 1996; Slowiaczek, McQueen, Soltano, & Lynch, 2000.)

The conclusion that competitor activation depends on initial overlap is also contradicted by a series of intra-modal form-based priming studies (Goldinger, Luce, & Pisoni, 1989; Goldinger, Luce, Pisoni, & Marcario, 1992; Luce, Goldinger, Auer, & Vitevitch, 2000). In one of these studies, Luce et al. presented subjects with primes and targets that were phonetically similar but shared no position-specific segments (e.g., *shun-gong*). Luce et al. found that shadowing times were significantly slower for those targets following phonetically related primes than to ones following unrelated primes. This result is consistent with the radical activation account, given that none of the prime-target pairs shared word-initial segments.

Allopena, Magnuson, and Tanenhaus (1998) provide additional support for radical activation models (and hence against the minimal discrepancy hypothesis). Using a paradigm that tracked participants' eye movements as they followed spoken instructions to manipulate objects on a computer screen, Allopena et al. found that rhyming competitors are activated early in the recognition process. When asked to use a mouse to click on a picture of a *beaker*, participants' fixation probabilities indicated that they also considered a picture of a *speaker* to be a likely candidate. The remarkable aspect of this finding is that fixation probabilities to the competitor started increasing before offset of the spoken word, suggesting that the participants' eye movements closely tracked competitor activation. These findings indicate that shared word-initial information is not necessary to activate competitors (the spoken word *beaker* resulted in increased probabilities to fixate on the picture of the *speaker*).

In short, the evidence from both intramodal phonetic priming and eye movement studies casts doubt on the validity of the minimal discrepancy hypothesis. Moreover, results from research on the activation of embedded words in longer carrier words (e.g., *lock* in *hemlock*; Luce & Cluff, 1998; Luce & Lyons, 1999; Shillcock, 1990; Vroomen & de Gelder, 1997) and from a number of other sources (Andruski, Blumstein, & Burton, 1994; Charles-Luce, Luce, & Cluff, 1990; Cluff & Luce, 1990; Connine, Blasko, & Hall, 1991; Goodman & Huttenlocher, 1988; see also Mattys, 1997) also call the hypothesis into question, thus supporting the general claims of radical activation models.

Although the current evidence appears to favor a radical activation account of spoken word recognition, the data point more toward a modification, rather than outright rejection, of the Cohort model's original claims regarding the nature of the activated competitor environment. The evidence still strongly favors a marked left-to-right bias in the processing of spoken words, supporting the spirit of the original cohort model, if not the precise detail (see Luce, 2001). In short, it is fairly certain that under optimal listening conditions, word onsets strongly determine the activation of competitors in memory, and unfolding acoustic-phonetic information over time guides the activation process. It does appear, however, that activation of form-based representations is not exclusively limited to onsets.

Competition. In activation-competition models (such as the original Cohort model), the hallmark of the lexical recognition process is competition among multiple representations of words activated in memory. As a result, the role of competition has been a primary focus of research and theory on spoken word recognition in the last few years (e.g., Cluff & Luce, 1990; Goldinger, Luce, & Pisoni, 1989;

Marslen-Wilson, 1989; McQueen, Norris, & Cutler, 1994; Norris, McQueen, & Cutler, 1995; Vitevitch & Luce, 1998, 1999).

Evidence for competition among form-based lexical representations activated in memory has come from a variety of experimental paradigms. For example, Luce and colleagues (Cluff & Luce, 1990; Luce & Pisoni, 1998) have shown that similarity neighborhood density and frequency, both indices of lexical competition, have demonstrable effects on processing time and accuracy in speeded single-word shadowing, auditory lexical decision, and perceptual identification. Recall that a similarity neighborhood is defined as a collection of words that are similar to a given target word. For example, the target word cat has neighbors such as *pat*, *kit*, *catty*, *cad*, *scat*, and so on. (See Luce & Pisoni, 1998, and Luce et al., 2000, for a discussion of the various similarity metrics that have been employed in the computation of similarity neighborhoods).

Neighborhoods may vary on both the density and frequency of the words that comprise them. Some words (e.g., *cat*) have many high-frequency neighbors, whereas others (e.g., *try*) have fewer, less frequent neighbors. As previously noted, Luce and colleagues have shown that words residing in densely populated similarity neighborhoods, in which lexical competition is predicted to be strong, are processed less quickly and less accurately than words residing in sparsely populated neighborhoods. Moreover, in similarity neighborhoods composed of high-frequency words, competition is more severe than in neighborhoods of low-frequency words, resulting in slower and less accurate processing. (See also Goldinger et al., 1989; Goldinger et al., 1992; Luce et al., 2000.)

Although there is now considerable evidence for competitive effects in spoken word recognition, debate continues over the precise mechanisms underlying lexical competition. As noted above, in models of recognition such as Trace, Shortlist, and PARSYN, lateral inhibition among lexical representations is a fundamental feature of the competitive process. The Cohort model, on the other hand, eschews the notion of lateral inhibition in favor of a competitive process that is modulated primarily by top-down (context-driven) and bottom-up (stimulus driven) facilitation and inhibition (Marslen-Wilson, 1987, 1990; see, however, Marslen-Wilson et al., 1996). More recently, Gaskell and Marslen-Wilson (1997, 1999) have proposed that speech input consistent with many lexical representations (or neighbors) results in more diffusely activated distributed representations, producing effects of lexical competition in the absence of lateral inhibition. In short, although most agree that lexical items activated in memory somehow compete, there is currently some disagreement regarding the exact mechanisms of this lexical interaction.

Summary. Although virtually all current models of spoken word recognition assume multiple activation of, and subsequent competition among, similar sounding representations in memory, the models differ on precisely how representations are activated (constrained or radical) and the means by which they compete (via simple competition or direct lateral inhibition). These differences, although important, should not obscure the now strong consensus in the field that activation-competition models of spoken word recognition—much in the spirit of the original Cohort theory—best capture the fundamental principles of spoken word recognition.

The Nature of Lexical and Sublexical Representations • Research on lexical competition has focused primarily on interactions among representations of words. However, an equally crucial topic concerns the nature and existence of sublexical (or segmental) representations. Some have argued against the existence of any sort of sublexical representation intervening in the mapping between feature and word (Marslen-Wilson & Warren, 1994). Among the proponents of sublexical representations, a lively debate has arisen concerning the nature of the interaction—or lack thereof—between segmental and lexical representations. We first consider the argument against sublexical representations and then discuss recent research examining the consequences of sublexical patterning (or probabilistic phonotactics) for lexical processing. The evidence thus far suggests a role for sublexical representations, although the debate over the interaction of lexical and sublexical units is still very much alive (see Norris et al., 2001; Pit & Samuel, 1995; Samuel, 2000).

Against sublexical representations. Researchers in mainstream speech perception research have long assumed that the speech waveform is recoded into successively more abstract representations, proceeding from acoustic to phonetic feature, from phonetic feature to segment, from segment to syllable, and, ultimately, to lexical representation (Pisoni & Luce, 1987). This central dogma of the field was challenged by Marslen-Wilson and Warren (1994; see also Klatt, 1979 for a similar view), who argued for a direct mapping of feature to word, with no intervening representations. Drawing inspiration from Streeter and Nigro (1979) and Whalen (1984, 1991), Marslen-Wilson and Warren generated a set of cross-spliced words and nonwords in which coarticulatory information signaling phonetic segments mismatched. For example, the initial consonant and vowel of the word *jog* was spliced onto the final consonant from the word *job*, creating a subcategorical mismatch between the vowel of the spliced stimulus and the final consonant. That is, information in the vowel of the spliced stimulus was consistent with the final consonant /g/, not the spliced consonant /b/. Similar nonword stimuli were constructed with mismatching coarticulatory information between the vowel and final consonant (e.g., the consonant and vowel of *smod* was cross-spliced with the final consonant of *smob*).

Marslen-Wilson and Warren's results demonstrated that mismatching coarticulatory information slowed processing only for stimuli that activate lexical representations (i.e., words cross-spliced with other words and words cross-spliced with nonwords). Nonwords cross-spliced with other nonwords (e.g., *smob*) failed to show detrimental effects of subcategorical mismatch on processing times. According to Marslen-Wilson and Warren, mismatching coarticulatory information can only be detected when representations at the lexical level are activated. Thus, the failure to observe subcategorical mismatch for nonwords is presumably a direct consequence of the absence of sublexical representations that can detect the mismatching information in the nonwords cross-spliced with other nonwords.

Later research revealed that conclusions regarding the demise of the segmental representation were premature. Two problems arose, one empirical and one theoretical. First, McQueen, Norris, and Cutler (1999) demonstrated that the asymmetry between cross-spliced words and nonwords could be made to come and go as a function of task demands, a demonstration that calls into question the empirical basis of Marslen-Wilson and Warren's original conclusions. Equally as damning, Cutler, Norris, and McQueen (Reference Note 6) showed that models with a phonemic level of representation could simulate Marslen-Wilson and Warren's original data pattern, thus removing the theoretical underpinnings for their claims that sublexical representations are not operative in spoken word recognition.

Evidence for two levels of processing: probabilistic phonotactics. At present, there is little compelling evidence against intermediate representations in spoken word recognition. Indeed, recent research on probabilistic phonotactics strongly suggests that sublexical representations have demonstrable affects on processing of both words and nonwords.

Probabilistic phonotactics refers to the relative frequencies of segments and sequences of segments in syllables and words. Using estimates of positional probabilities based on a computerized lexicon, Treiman, Kessler, Knewasser, Tincoff, and Bowman (1996) found that participants' performance on rating and blending tasks was sensitive to probabilistic differences among phonetic sequences. Participants in the rating task judged high probability patterns to be more "English-like" than low probability patterns (see also Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997). In the blending task, when asked to combine two sound patterns into a single item, high probability sequences tended to remain intact more often than low probability sequences. (See also Brown & Hildum, 1956; Eukel, 1980.)

Vitevitch, et al. (1997) examined the effects of probabilistic phonotactic information on processing times for spoken stimuli. They used bisyllabic nonwords composed of phonetic sequences that were legal in English but varied in their segmental and sequential probabilities. Using a speeded single-word shadowing task, Vitevitch et al. found that nonwords composed of common segments and sequences of segments were repeated faster than nonwords composed of less common segments and sequences. Taken together, these studies demonstrate that information regarding the legality and probability of phonotactic patterns has demonstrable influences on the representation and processing of spoken stimuli (see also Massaro & Cohen, 1983).

However, a potential anomaly has arisen. The effects of phonotactics demonstrated thus far seem to contradict the predictions of—and evidence for—the class of models that emphasizes the roles of activation and competition in spoken word recognition. In particular, predictions of NAM are in direct contrast to Vitevitch et al.'s work on probabilistic phonotactics. Recall that, according to NAM, spoken words that sound like many other words (i.e., words in dense similarity neighborhoods) should be recognized more slowly and less accurately than words with few similar sounding words (i.e., words in sparse similarity neighborhoods). A contradiction is revealed by the observation that high probability segments and sequences of segments are found in words from high density neighborhoods, whereas low probability segments and sequences of segments are found in words from low density neighborhoods. Thus, NAM predicts that high probability phonotactic stimuli should be processed more slowly than low probability phonotactic stimuli, in contrast to the findings of Vitevitch et al.

To explore these seemingly contradictory results, Vitevitch and Luce (1998; see also Vitevitch & Luce, 1999) presented participants in a speeded auditory shadowing task with monosyllabic words and nonwords that varied on similarity neighborhood density and phonotactic probability. They generated two sets of words and nonwords: 1) high phonotactic probability/high neighborhood density stimuli and 2) low phonotactic probability/low neighborhood density stimuli. Vitevitch and Luce replicated the pattern of results obtained in the Vitevitch et al. study for nonwords: High probability/density nonwords were repeated more quickly than low probability/density nonwords. However, the words followed the pattern of results predicted by NAM. That is, high probability/density words were repeated more slowly than low probability/density words.

Vitevitch and Luce suggested that two levels of representation and processing—one lexical and one sublexical—are responsible for the differential effects of phonotactics and neighborhoods. (The concept of two levels of processing has, of course, a long history in the field. For similar proposals regarding levels of processing in spoken word recognition, see Cutler & Norris, 1979; Foss & Blank, 1980; McClelland & Elman, 1986; Norris, 1994; Radeau, Morais, & Segui, 1995). In particular, Vitevitch and Luce suggested that facilitative effects of probabilistic phonotactics reflect differences among activation levels of sublexical units, whereas effects of similarity neighborhoods arise from competition among lexical representations. Models of spoken word recognition such as Trace, Shortlist, and NAM all propose that lexical representations compete with and/or inhibit one another (see Cluff & Luce, 1990; Goldinger, Luce, & Pisoni, 1989; Marslen-Wilson, 1989; McQueen, Norris, & Cutler, 1994; Norris, McQueen, & Cutler, 1995). Thus, words occurring in dense similarity neighborhoods succumb to more intense competition among similar sounding words activated in memory, resulting in slower processing. Apparently, effects of lexical competition overshadow any benefit these high density words accrue from having high probability phonotactic patterns. On the other hand, because nonwords do not make direct contact with a single lexical unit, and thus do not immediately initiate large-scale lexical competition, effects of segmental and sequential probabilities emerge for these stimuli. That is, in the absence of strong lexical competition effects associated with word stimuli, higher activation levels of sublexical units (i.e., those with higher phonotactic probabilities) afford advantage to high probability nonwords.

Although the research examining neighborhood density and phonotactics strongly suggests the operation of two levels of representation, the Vitevitch and Luce studies did not demonstrate effects of probabilistic phonotactics on real words. Is the effect of phonotactics restricted to nonwords? If so, proponents of direct access models like Cohort (in which there are no sublexical representations) might assert that the available evidence does not support a role for intermediate, sublexical representations in the recognition of real words. Fortunately, subsequent research by Luce and Large (2001) indicates that facilitative effects of probabilistic phonotactics are not restricted to nonwords. By orthogonally manipulating density and phonotactics, thereby unconfounding their effects, Luce and Large demonstrated simultaneous competitive effects of neighborhood density and facilitative effects of probabilistic phonotactics for both words and nonwords.

Summary. At present there is little compelling direct evidence against those models that incorporate intermediate levels of representation (such as Trace, Shortlist, or PARSYN). In addition, research on phonotactics and neighborhood activation supports the hypothesis of (at least) two levels of representation and process in spoken word recognition. One level is sublexical, consisting of facilitative activation among segments and sequences of segments. The other level is lexical, consisting of competitive interactions among multiple word-forms. Models of spoken word recognition, such as NAM and Cohort theory, which lack a sublexical level of representation, cannot account for these effects. However, Trace, Shortlist, and PARSYN, which have two levels of representation and processing, may more accurately account for spoken word recognition effects as a function of neighborhood activation and probabilistic phonotactics.

Lexical Embeddedness

How do listeners recognize a word containing other words? On hearing the word *hemlock*, do listeners entertain hem and *lock* as possible interpretations of the input? Or, is one interpretation—say that of the longest word consistent with the input (e.g., *hemlock*)—preferred over others (e.g., *hem* and *lock*)? The problem of lexical embeddedness has attracted much attention in research on spoken word recognition, primarily because understanding how the processing system deals with embedded words has important consequences for the nature of the activation process (is it radical?), the existence of sublexical units of representation (do they exist?), and the segmentation of words from fluent speech (how does the listener decide where one word ends and another begins?). More specifically, understanding how the system copes with lexical embeddedness bears directly on issues of activation and competition in spoken word recognition. Activation of lexical components (e.g., *hem* and *lock*) of a longer word would be consistent with radical activation models, and evidence for suppression of component lexical items by the longer, carrier word would provide further evidence for lexical competition. In short, questions regarding the role of lexical embeddedness in spoken word recognition may have important implications for theoretical accounts of the nature of lexical activation and processing (see Norris, 1994).

In the early version of Cohort theory, embedded words entered into the cohort only if they coincided with the beginnings of the carrier words (e.g., *chair* in *cherish*). An embedded word occurring later in the carrier word (such as *stress* in *distress*) would not be activated because the later occurring embedded item would not be a member of the cohort. Later versions of the theory (Marslen-Wilson, 1987, 1993) relaxed this constraint. Nonetheless, neither present nor previous versions of this theory actually predict that embedded items should affect processing times. If lexical uniqueness points are held constant, processing should be neither slower nor faster for words containing initial embedded items.

Swinney (1981) has offered a somewhat different proposal that nonetheless makes similar predictions to those of the earlier version of Cohort theory. According to Swinney's "minimal accretion principle," when faced with lexically embedded items, listeners opt for the interpretation that spans the longest word. For example, in processing a potential two-word item such as *kidney*, the carrier word would be the preferred interpretation, and not the two-word utterance, *kid knee*. The principle clearly states that later occurring embedded items should not result in ambiguous parsings of the speech stream if a single interpretation consistent with the longer carrier word is viable. In short, both Swinney's minimal accretion principle and Marslen-Wilson's Cohort theory predict little effect of later occurring embedded words on processing. Although both accounts hypothesize independent activation of initially embedded items, they are silent about the precise effects these items may have on processing of the carrier word.

Proposals contrary to those of Marslen-Wilson and Swinney about lexical embeddedness come in many forms. For example, Cutler (1989) discusses a model of lexical access in which form-based lexical representations (i.e., representations of sound patterns) may be activated at the onset of each syllable. Thus, every syllable that corresponds to a word will result in activation of a lexical item in memory. Cutler and Norris (1988) later modified this radical activation hypothesis by invoking the "metrical segmentation strategy" (MSS), in which lexical hypotheses are generated based on strong syllables. (Strong syllables, not to be confused with stressed syllables, are those containing full vowels.) According to MSS, embedded items coinciding with strong syllables should activate lexical representations regardless of where they occur in the carrier word (see also Vroomen & de Gelder, 1997; Charles-Luce, Luce, & Cluff, 1990; Cluff & Luce, 1990).

The Trace model has much in common with this radical activation view, while at the same time making predictions that are remarkably similar to those of Swinney (1981; see also Frauenfelder & Peeters, 1990). According to Trace, all lexical representations that are consistent with a given portion of the speech signal may be activated at any point in time. Thus, lexically embedded items at both the beginnings and endings of words may be activated. However, due to its architecture, Trace clearly prefers interpretations corresponding to longer words. In a series of simulations, Frauenfelder and Peeters (1990) examined Trace's processing of lexically embedded carrier words. Their simulations confirmed that within Trace, embedded words may indeed be activated at any point in time. However, lateral inhibition among lexical units provides activation advantages to longer carrier words over later occurring embedded words. For example, the embedded word *seed* in the carrier word *precede* will be inhibited by the previously activated carrier word and will subsequently exert little influence on the recognition process. In the case of initially embedded items (e.g., *chair* in *cherish*), however, Frauenfelder and Peeters demonstrated that Trace predicts strong activation for both the carrier and the embedded words. Because of lateral inhibition among multiply activated items (which include the carrier and embedded words), Trace predicts that carrier words with initially embedded items should be processed more slowly than longer words with no embedding. Shortlist makes similar predictions regarding the activation and processing of embedded words.

To date, the empirical work on the effects of embedded words has failed to provide unequivocal support for any of these competing theoretical accounts. Using the cross-modal priming technique, Prather and Swinney (1977) found evidence for activation of embedded words occurring in the first, but not the second, syllables of bisyllabic words. For example, auditory presentation of *boycott* primed a visual target related to *boy*, but not one related to *cot*, a result that motivated the formulation of the minimal accretion principle. Pitt (Reference Note 16) also failed to obtain evidence for the activation of second-syllable embedded words in bisyllabic carrier words (see also Gow & Gordon, 1995), although he obtained evidence of activation of words embedded as the second syllable of nonword carrier items. For example, although the nonword *trolite* primed *dark*, *polite* failed to do so. Pitt proposed that the carrier word inhibits activation of noninitial embedded words (but see Norris, McQueen, and Cutler, 1997, for possible reasons why word-final embedded words may not prime semantic associates in a cross-modal task). Shillcock (1990) obtained a different pattern of results. For items in which the first syllables were not prefixes (e.g., *guitar*), he found evidence for activation of second-syllable embedded words (e.g., *tar*), suggesting that even noninitial embedded words may sometimes be activated (see also Luce & Cluff, 1998; Vroomen & de Gelder, 1997).

The results from the cross-modal priming paradigm are obviously mixed. Overall, the bulk of the evidence suggests that embedded words are activated during recognition, although the precise conditions under which activation is evident have yet to be specified.

The effects of lexical embeddedness on spoken word processing have also been examined using the word spotting task. In this task, participants attempt to detect an embedded word in a longer stimulus item as quickly as possible. McQueen, Norris, and Cutler (1994) found that second-syllable embedded words were harder to detect when the nonword carrier item constituted the beginning of a real word. For example, participants had more difficulty detecting *mess* in *demess* (which is the beginning of the word domestic) than *mess* in *nemess* (which does not begin a real word in English; see also Norris, McQueen, & Cutler, 1995). These results demonstrate that carrier items in which lexically embedded words occur compete with the embedded items for recognition.

Although the cross-modal priming and the word spotting tasks have been used to examine activation and detection of lexically embedded words, neither of these tasks requires participants to respond to the carrier word itself. To determine the direct effects of lexical embedding on the processing of carrier words, Luce and Lyons (1999) compared processing of words with and without lexically embedded items. In one experiment, lexical items were embedded in the initial portions of target words (e.g., *chair* in *cherish*); in a second experiment, embedding occurred in final position (e.g., *tar* in *guitar*). The results from both auditory lexical decision and speeded single-word shadowing tasks demonstrated a clear effect of embedding only for initial position: Carrier words with embedded words in initial position were responded to more quickly than their matched, nonembedded counterparts. No effects were observed for words with embedded words in final position. Luce and Lyons argued that the facilitative effects of initial lexical embedding were the result of feedback loops between lexical and segmental units. That is, segments may pass activation to lexical nodes corresponding to both the carrier word and the initial embedded item. These lexical nodes may then pass activation back to the segment nodes, establishing a feedback loop that would afford activation advantages to words with initial embedded items.

Luce and Lyons' finding that final embedded items fail to show measurable effects on processing times for spoken carrier words is consistent with earlier proposals that the processing system prefers interpretations corresponding to longer words. This preference may result from implicit strategic processing or may be a consequence of lateral inhibition among competitors. Whatever the precise mechanism, it is now becoming clear that, despite evidence for activation of lexically embedded items at the ends of carrier words (Luce & Cluff, 1998; Shillcock, 1990; Vroomen & de Gelder, 1997), the processing of longer carrier words appears to be unaffected by the presence of finally embedded items.

Summary. Although the details of the results from the cross-modal, word spotting, and speeded processing tasks sometimes conflict, the overall data pattern on lexical embeddedness supports a radical activation account of spoken word recognition. However, the failure to observe effects of final embedding in certain studies suggests that the time-course of processing may disfavor pronounced activation of late occurring embedded items, especially when strong lexical hypotheses are already in play.

Segmentation of Spoken Words from Fluent Speech • To this point, we have focused on research devoted primarily to understanding the perception of isolated spoken words. However, since the 1970s (and even earlier; see, e.g., Miller, Heise, & Lichten, 1951), much research has focused on the perception of words in larger units. Indeed, Marslen-Wilson's seminal work on Cohort theory was concerned with the role of sentential context on recognition (see also Cole & Jakimik, 1980). Recent work has also been concerned with the perception of words in larger contexts, and in particular, with the fundamental issue of how spoken words are segmented from fluent speech.

The listener's phenomenal experience of continuous speech is one of a succession of discretely produced words. However, the search for reliable acoustic information marking the beginnings and endings of words has met with little success (Nakatani and Dukes, 1977; Lehiste, 1972). Given that the signal fails to consistently mark the boundaries of words in spoken discourse, a crucial question in research on spoken word recognition concerns the means by which the perceptual system segments the continuous stream of speech. Four traditional solutions to the problem of segmentation have been proposed: phonetic, prosodic, lexical, and phonotactic.

According to the phonetic solution, the speech signal may provide cues to the position-specificity of phonemes (which are sometimes referred to as "context-sensitive allophones" [Church, 1987]), thus providing potential information to the listener regarding beginnings and endings of words. For example, syllable-initial and syllable-final stops are phonetically and acoustically distinct, distinguished in part by degree of aspiration (e.g., Davidson-Nielson, 1974; Dutton, Reference Note 8). In particular, aspirated /t/ is always syllable initial (see Church, 1987). Previous research has demonstrated that some phonetic cues may be used in segmentation. Among the possible cues to word boundaries that have been shown to facilitate segmentation are allophonic variation of /l/ and /r/ (Nakatani and Dukes, 1977, 1979), aspiration of word-initial stops (Christie, 1974), and duration of word initial segments (Christie, 1977). This collection of cues, among others (see Dutton, Reference Note 8), may assist the listener in identifying consonants as pre- or postvocalic, thus indicating whether a given consonant occurs at the beginning or end of a syllable or word. To date, however, the lack of success at identifying all but a restricted set of possible phonetic cues to word boundaries suggests that the phonetic solution may be severely limited.

According to the prosodic solution to the segmentation problem, listeners parse the speech stream by exploiting rhythmic characteristics of their language (Cutler, 1996; Cutler & Norris, 1988; Cutler & Butterfield, 1992; Vroomen, van Zon, & de Gelder, 1996). The most notable hypothesis representing the prosodic solution to the segmentation problem is embodied in Cutler and Norris' (1988) metrical segmentation strategy (MSS). As previously discussed, MSS asserts that lexical access is attempted based on each strong syllable. Cutler and Norris argue that because most strong syllables in English are word-initial (see Cutler & Carter, 1987), a strategy of attempting lexical access at each strong syllable would meet with frequent success.

Many models of spoken word recognition espouse the lexical solution, whereby segmentation is a byproduct of the recognition process. Cole and Jakimik (1980) propose that words in fluent connected speech are recognized one at a time in sequential order. The word-by-word assumption obviates the need for any explicit segmentation process because the recognition of a word makes evident both the end of the just recognized word and the beginning of the next. Likewise, the Cohort model (Marslen-Wilson, 1992; Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Zwitserlood, 1989) characterizes word recognition as a strictly sequential process in which segmentation is accomplished incidentally as each word in the speech stream is recognized. Although both models have intuitive appeal and appear to match our experience of recognizing one spoken word after another, the phenomenon of lexical embeddedness is problematic for approaches that assume that segmentation is a byproduct of recognition (see also Tabossi, Burani, & Scott, 1995). In particular, it is unclear how a word-by-word approach would deal with the presence of a short word embedded in a longer lexical item. For example, in the word *catatonic*, if the end of *cat* signals the beginning of a new word, the listener will be forced to attempt lexical access based on the fragment *-atonic*, resulting in a potential failure of recognition. The pervasiveness of lexical embeddedness suggests that the word-by-word strategy proposed by Cole and Jakimik and Marslen-Wilson may result in an unacceptably high number of misparsings (McQueen & Cutler, 1992; McQueen, Cutler, Briscoe, & Norris 1995; Luce, 1986; Luce & Lyons, 1999; see also Grosjean, 1985; Bard, Shillcock, and Altmann, 1988). It should be noted, however, that recent research (Davis, Marslen-Wilson, & Gaskell, Reference Note 7) suggests that initially embedded words and their nonembedded counterparts may be phonetically distinct (i.e., *car* in *carpet* is not phonetically identical to the individual lexical item, *car*). Thus, acoustic-phonetic information may help alleviate the embeddedness problem by signaling when a potential word is actually part of a longer item.

Other models of spoken word recognition espousing the lexical solution to segmentation emphasize the role of competition among lexical candidates (e.g., Norris, McQueen, & Cutler, 1995; Vroomen & de Gelder, 1995). For example, in Shortlist, segmentation occurs via a competitive process in which the unit most closely matching the entire input receives the most activation.

Recently, research on spoken language has focused on a fourth potential solution to the segmentation problem: probabilistic phonotactics (see Massaro & Cohen, 1983). If listeners are sensitive to variations in the frequencies of segments and their sequences, probabilistic phonotactics may provide useful information for segmentation. Norris, McQueen, Cutler, and Butterfield (1997) have demonstrated that listeners take into account phonotactic information when attempting to detect real words embedded in nonsense words (which, of course, requires the identification of the boundaries of the target words). Norris et al., demonstrated that participants were able to detect words embedded in nonsense words faster and more accurately when the additional segments of the nonsense words formed phonotactically legal syllables (or possible words) than when the additional segments did not constitute well-formed syllables in English. That is, subjects were faster and more accurate at detecting *apple* in *vuffapple*, where *vuff* is itself a phonotactically legal syllable, than *apple* in *fapple*, in which the additional segment *f* does not constitute a legal syllable. These results suggest that listeners are able to make use of phonotactic information on-line in parsing fluent speech. (See also McQueen, 1998; Gaygen & Luce, Reference Note 9.)

Summary. Although the available evidence supports some role for probabilistic phonotactics in the segmentation of spoken words, it has become increasingly clear that segmentation is best viewed as a constraint satisfaction problem in which the perceiver employs various solutions—phonetic, prosodic, lexical, and phonotactic—to determine the beginnings and endings of spoken words in fluent speech.

Representational Specificity • Theories of spoken word recognition have traditionally assumed, either implicitly (e.g., McClelland & Elman, 1986; Luce & Pisoni, 1998; Norris, 1984) or explicitly (e.g., Jackson & Morton, 1984), that lexical items are represented in memory by abstract phonological codes that only preserve information relevant for lexical discrimination. In many current models of word recognition, stimulus variation—arising from factors such as changes in speaking rate and the identity of the talker—is treated as irrelevant information that is discarded early in the encoding process. As noted earlier, the extraction of information that is solely relevant for identification is referred to as normalization, and it is during the normalization phase that representations of stimuli that vary in physical detail but fall within a given perceptual category are equated.

For example, feature-based accounts of speech perception (see Klatt, 1989; Pisoni & Luce, 1987; Marslen-Wilson & Warren, 1994) have proposed that speech sounds and words are processed using the elemental features of linguistic description (e.g., [vocalic], [consonantal], [sonorant]). However, spoken words may differ on many physical dimensions not captured by these features. The normalization process is responsible for winnowing the information in the speech signal and extracting only the featural information that is relevant for identification. This process thereby serves a substantial data reduction function that may ultimately result in considerable economy of process and representation.

Despite the arguments that have been made for abstract lexical representations in memory, recent research (see Goldinger, 1996, 1998, for a review) has suggested that putatively irrelevant "surface" details of words—such as information specific to a given talker—are preserved in some form in memory. These findings regarding specificity effects have led to the proposal (e.g., Goldinger, 1996, 1998) that lexical items are represented in memory by episodic representations that preserve, rather than discard, much of the physical detail of the stimulus.

Research has demonstrated that variation in the surface details of spoken stimuli (usually measured by changes in the identity of the talker, hereafter referred to broadly as changes in "voice") has implications for both identification and memory. Typically, subjects have more difficulty in identifying (Mullennix, Pisoni, & Martin, 1989), recognizing (Church & Schacter, 1994; Goldinger, 1996; Palmeri, Goldinger, & Pisoni, 1993; Schacter & Church, 1992; Sheffert, Reference Note 19; Sheffert, 1998a, 1998b), and recalling (Goldinger, Pisoni, & Logan, 1991; Martin, Mullennix, Pisoni, & Summers, 1989) lists of stimuli composed of words spoken by multiple talkers compared with lists composed of stimuli spoken by a single talker. (See Palmeri, Goldinger, & Pisoni, 1993, for one interesting exception). One explanation for these effects is that normalization processes reduce resources available for encoding and/or rehearsal.

The effects of changes in the surface details of stimuli between study and test in recognition memory experiments have been of particular interest in the literature. For example, Church and Schacter (1994) and Schacter and Church (1992) investigated the effects of talker variation on implicit and explicit memory. They observed effects of talker variation in implicit tasks such as fragment completion and identification of low-pass filtered stimuli. Subjects were more likely to complete a fragment of a word if the fragment was repeated in the same voice. Subjects were also more accurate at identifying low-pass filtered words that were repetitions of previously presented items if the repetition preserved surface characteristics of the stimulus. However, these researchers did not find effects of stimulus specificity in explicit tasks. When subjects performed cued recall or recognition of previously presented items, changes in surface characteristics between study and test had no statistically significant effects on performance.

Goldinger (1996) also conducted a series of experiments examining the effects of voice on memory for spoken words. In one experiment, he presented words in explicit (recognition) and implicit (perceptual identification in noise) tasks with varying delays between study and test. He found significant effects of voice in both recognition and identification, demonstrating that voice effects are not, in fact, restricted to implicit tasks. However, Goldinger found that effects of voice were reduced more by delay between study and test in the explicit task than the implicit task. In another experiment, Goldinger manipulated levels of processing and voice in the study-test implicit-explicit format. His results demonstrated that effects of voice varied with level of processing, such that strongest effects of stimulus specificity were observed in the shallower processing conditions, especially for recognition memory.

Although somewhat varied, the overall results of studies examining the effects of voice on identification and memory are consistent with a number of current theoretical proposals. According to the exemplar-based models (e.g., Hintzman, 1986), a new representation of a stimulus item is stored in memory each time it is encountered, and it is hypothesized that these representations preserve surface information about the stimulus. One advantage of exemplar-based models is that they have the potential for solving the long-standing problem of perceptual normalization in speech perception by dispelling the notion that the ultimate goal of the perceptual process is to map acoustic-phonetic information onto abstract form-based representations of words in memory. In exemplar-based models, the representational currency of the perceptual encoding process is more-or-less true to the details of the stimulus itself. In an application of this general theoretical approach to spoken word recognition, Goldinger (1996, 1998) has proposed an episodic lexicon in which the individual memory traces themselves may encode both abstract and surface information (see also Luce & Lyons, 1998; Luce, Charles-Luce, & McLennan, Reference Note 14), with the degree of stimulus specificity depending crucially on attentional factors during encoding.

Other theoretical approaches have proposed that abstract and specific phonetic representations are stored in separate memory systems (e.g., Schacter, 1990, 1992), in contrast to the exemplar models, which propose a single storage mechanism. Distributed memory models (e.g., Gaskell & Marslen-Wilson, 1997) may also provide a means for encoding specificity without the need for storage of multiple exemplars. Although the precise nature of the representational format is currently unclear, it is incumbent on the next generation of word recognition models to be able to account for the representation and processing of acoustic and phonetic specificity.

Summary. Recent research on representational specificity has demonstrated that listeners preserve much more specific information in memory about spoken words than previously thought. Variations in voice and speaking rate have demonstrable effects on processing. Moreover, these variations appear to be preserved in memory for spoken words. Research on representational specificity has—and will continue to have—an important effect on the way we conceive of representation and process in spoken word perception.

Spoken word recognition: Conclusion. Research and theory on spoken word recognition continues to evolve at a rapid pace, and there is little doubt that substantial progress has been made in our understanding of how the listener maps acoustic-phonetic information onto lexical representations of words in memory. Despite the number of models and unresolved issues, there is a growing consensus among researchers concerning the fundamental principles that characterize spoken word perception. At the very least, we can be cautiously confident that some form of activation-competition model will eventually prove to be an adequate working model of the listener's remarkable ability to recognize the spoken word.

Developing the Capacities to Process Fluent Speech

The early infant studies yielded a wealth of information about the nature and extent of the perceptual capacities and provided the necessary foundation for understanding their role in infants' acquisition of a native language. Nevertheless, centered as they were on the processing of minimal differences between speech sounds, the early studies encouraged the view that infants' acquisition of the sound organization of their native language proceeds in a bottom-up fashion, beginning with the recovery of the elementary units, observing their possible combinations, and building up to larger units such as words. Although development could follow this course, the early findings certainly do not prove it. Moreover, an exclusive focus on infants' perception of phonetic contrasts ignores the basic end of learning to speak and understand a language, namely, the desire to be able to communicate one's thoughts and feelings to others. If, instead, infants are seen as trying to master a system to communicate with others, then the concerns that drive research become ones having to with recovering larger units of organization that bear more directly on recovering a meaningful interpretation of utterances, namely, words, phrases, and clauses. Indeed, the mastery of more fine-grained elements of sound organization, such as phonetic categories and syllables, may conceivably fall out of infants' efforts to learn the larger units that map more directly to meanings. Indeed, several models have been proposed to account for how speech perception capacities develop to meet infants' needs in developing an effective system for learning words and communicating with other speakers of their native language.

Models Relating Speech Perception Capacities to Word Recognition Abilities • Several attempts have been made to relate changes in speech perception capacities to the development of word recognition skills in infants. These models typically assume infants' innate capacities for distinguishing speech contrasts, and focus on how task demands associated with developing efficient word recognition skills influence the organization of these capacities.

The Syllable Acquisition, Representation, and Access Hypothesis (SARAH) model. This model, proposed by Mehler, Dupoux, and Segui (1990), tries to explain the relation among speech perception capacities, word recognition processes, and acquiring a lexicon. The model assumes a strong correspondence between the processes used to acquire a lexicon and those used by adults during lexical access. According to the model, infants initially possess three important components for acquiring a lexicon. The first of these is a syllabic filter that chops continuous speech into syllable-sized segments. Only legal syllable structures such as CV, CVC, V, and CCVC syllables, with talker-specific and speaking rate variables factored out, are output by this filter. The second component is a phonetic analyzer, which provides a description of the syllable in terms of a universal set of phonetic segments and allows it to be mapped to a code that relates to the articulatory gestures required to produce it. The third component is a detector that draws on syllabic representations and other acoustic information to compute word boundary cues.

The model posits that the change from a language general capacity to one attuned to infants' native language depends on two specialized mechanisms. The first of these is unlearning, or selective stabilization. In effect, the system becomes attuned to those phonetic contrasts, syllables, and word boundary detectors that work best for the native language. Mehler et al. are not very clear about just how this happens, but argue that it occurs before the acquisition of a lexicon and suggest that it depends on statistical extraction and parameter setting. The second mechanism is compilation, a process that stores syllabic templates and logogens in long-term memory. The templates are extracted from the input and bootstrap the acquisition of lexical entries. Mehler et al. suggest that infants' lexical entries may differ considerably from those of adults. For instance, infants' lexicons may include items that are not fully segmented, as when a clitic or function word remains attached to a content word. Just how these entries are modified during development is not clear, although Mehler et al. suggest that the joint operation of bottom-up and lexical-morphological indices play some role in this process.

Developmental Model of Adult Phonological Organization (DAPHO). Suomi's (1993) model attempts to deal with both speech perception and production. DAPHO is a developmental model in the sense that it is built up from an earlier model, CHIPHO, that characterizes the child's early speech behavior at the 50-word stage. According to DAPHO, each word meaning in the lexicon is linked to both a motor plan and an auditory prototype, normalized for talker, speaking rate, etc., and containing the essential auditory properties of the word across different contexts. These properties are apprehended directly during word recognition without an intermediate stage of segmental analysis. After the incoming signal passes through a stage of auditory analysis, word boundaries are detected, and each word candidate is matched against the set of prototypes in the lexicon. Novel words are stored as holistic prototypes, whereas familiar words are matched against the existing set of prototypes in the lexicon. Ones which are sufficiently similar to the word candidate are activated, and the best fitting one is selected as a match and its meaning is activated. One point that is not discussed has to do with how the model determines whether an item is novel and should be stored as a new prototype, or should be treated as a familiar item. Presumably, some criterion based on the degree of similarity to existing prototypes is required for this purpose. Suomi assumes that the prototypes stored in the lexicon are initially quite global and include a limited number of salient auditory features. As more items are added to the lexicon, the prototypes become increasingly detailed to distinguish them from other lexical items, but continue to be continuous, holistic descriptions (i.e., not segmental). The assumption about how words are segmented from fluent speech is reminiscent of the top-down approach proposed by Cole and Jakimik (1978). Words are identified in succession starting with the beginning of an utterance, with the completed recognition of one word indicating the beginning of a new word candidate. As the child acquires more and more words, more top-down information is available to facilitate word boundary detection. Unfortunately, this approach is not without problems. In particular, as noted in our earlier discussion of word segmentation by adults, an exclusively top-down approach can be expected to have difficulties with words that appear as syllables embedded within larger words, such as *can* in *candle*, *toucan*, *uncanny*, etc.

Word Recognition and Phonetic Structure Acquisition (WRAPSA) model. This model was proposed as an account of how speech perception capacities evolve to support on-line word recognition in fluent speech (Jusczyk, 1993; 1997). A key assumption of the model is that the language learner develops a scheme for weighting information in speech so as to maximize the likelihood of picking up those contrasts that signify meaningful distinctions among words in the native language. A set of auditory analyzers identifies the spectral and temporal features present in the acoustic signal. The features extracted at this level are provided by the inherent organization of the human auditory system, so they are neutral with respect to the language that is being spoken. Infants rely largely on this type of description during the first few months of life. As learners gain experience with a particular language, the output of the analyzers is weighted to highlight those features that are most critical to making meaningful distinctions in the language. The weighting scheme allows perceivers to focus attention on certain features and de-emphasize others. A pattern extraction process operates on the weighted output, and segments the continuous signal into word-sized units.

The resulting representation of the input signal is global in that it assigns prominent features temporally to syllables, but does not provide an explicit breakdown into phonetic segments. However, these representations do encode information about prosodic properties, such as the relative stress, pitch, and durations of syllables within the word-sized units. The processed input representation is then matched against existing representations of known words that have been stored in the lexicon. If a close match to some item in memory is achieved, recognition occurs, and the meaning is accessed. If no close match is made, the input is reprocessed in an attempt to find a better match, or else the item may be stored as a new item in the lexicon.

One key element of WRAPSA is the assumption that listeners store specific instances of words they hear rather than abstract prototypes of lexical items. In this respect, the model is the developmental counterpart of exemplar-based models of the lexicon proposed for adults (Goldinger, 1996; 1998). Jusczyk (1993) originally based this assumption on the evidence suggesting that adults encode and remember much detail about specific utterances that they have heard (e.g., Craik & Kirsner, 1974; Goldinger et al., 1992; Schacter & Church, 1992). However, recent studies, which we review below, also suggest that infants' memory representations of words may include surface details, relating to talker voice characteristics.

Current Issues in the Acquisition of Spoken Word Recognition Capacities • We now turn to a discussion of a set of core issues that occupy much attention in current research and theory on the acquisition of spoken word recognition capacities. These issues concern the development of 1) lexical segmentation, 2) lexical representation, and 3) knowledge of linguistic units beyond the word. The picture that emerges is one of a developing sensitivity to the sound organization of the native language. A time-line tracing many of the developmental landmarks discussed in the subsequent sections is shown in Table 2.

Segmenting Words • Even in infant-directed speech, few utterances consist of isolated words. In analyzing the input heard by an infant between 6 and 9 mo, van de Weijer (1998) found that, excluding greetings, vocatives, and fillers, only about 7% of the speech consisted of one-word utterances. Thus, to learn the words of their native language, infants must have some ability to segment words from fluent speech. As discussed earlier, pauses rarely occur between words in fluent speech. Moreover, as anyone with the experience of listening to an unfamiliar foreign language can confirm, if you do not already know the language, it is hard to tell where one word ends and another begins. Segmentation of a foreign language is difficult because the most effective word segmentation cues are tailored to the sound structure of a particular language. Given this observation, one might anticipate that infants must learn about the sound orga-

TABLE 2. Some landmarks in developing sensitivity to native language sound organization.

4.5 mo:	Infants show recognition of their own names. Infants show sensitivity to clause boundaries in fluent speech.
6 mo:	**Infants respond appropriately to the words "Mommy" and "Daddy."**
7.5 mo:	Word segmentation begins. English-learners use prosodic stress cues for segmenting words.
8 mo:	Infants display statistical learning abilities.
9 mo:	Sensitivity to frequently occurring native language phonotactics and word stress patterns is evident. Phonotactic cues are used in segmenting words. Infants retain information about words that occur frequently in the input. English-learners display sensitivity to phrase boundaries.
10–12 mo:	Sensitivity to non-native phonetic contrasts declines.
10.5 mo:	Infants can use allophonic cues to segment words. English-learners can segment words with weak initial syllables.
12 mo:	**Infants appear to integrate different types of word segmentation cues.**
16 mo:	Infants show some ability to segment vowel-initial words.
17 mo:	**Lexical competition effects are present and affect word learning.**
24 mo:	Timecourse of infants word processing resembles that of adults, they are slower to respond to targets when distracters share initial consonants.

nization of their language before they make much progress in segmenting words.

In fact, infants acquire considerable information about native language sound organization between 6 and 9 mo. At 6 mo, English-learners are as likely to listen to lists of words from a foreign language, Dutch, as to ones from their native language (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993). However, by 9 mo, English learners listen significantly longer to the English word lists and Dutch infants listen significantly longer to the Dutch lists. Because the prosodic characteristics of English and Dutch words are very similar (Reitveld & Koopmans-van Beinum, 1987; Schreuder & Baayen, 1994), it appears that infants have learned about the segmental characteristics of their language, such as what sounds and sequences of sounds (phonotactic patterns) are permissible in words. Indeed, when such information was removed from the words through low-pass filtering, the infants no longer showed a preference for native language words. Furthermore, by 9 mo, infants are doing more than simply tracking which phonotactic patterns do or do not appear in native language input. They also demonstrate some knowledge of which phonotactic patterns occur frequently, as opposed to infrequently, within words (Jusczyk, Luce, & Charles Luce, 1994).

Similar gains are evident in infants' knowledge of the prosodic characteristics of native language words. For example, the predominant stress pattern of English words has an initial strong syllable (i.e., one that carries prosodic stress) followed by one or more weak syllables (Cutler & Carter, 1987). At 6 mo, English-learners are as likely to listen to lists of bisyllabic words without the predominant stress pattern as to ones with it. However, by 9 mo, English-learners listen significantly longer to words with the predominant strong/weak pattern than to ones with a weak/strong pattern (Jusczyk, Cutler, & Redanz, 1993).

The knowledge that infants between 6 and 9 mo gain about native language sound structure provides potential cues for segmenting words from fluent speech. Thus, as noted above, phonotactic cues (Brent & Cartwright, 1996; Cairns, Shillcock, Chater, & Levy, 1997) and prosodic stress cues (Cutler, 1990; 1994; Cutler & Norris, 1988) have been promoted as useful for segmenting English words. It is striking, then, that infants display the first signs of segmenting around 7.5 mo of age. Jusczyk and Aslin (1995) found that 7.5-mo-olds, but not 6-mo-olds, segment monosyllabic words, such as *cup*, *dog*, *feet*, and *bik*e from fluent speech. In one experiment, they familiarized infants with a pair of words (e.g., *cup* and *dog*) that were spoken as isolated words. Then infants heard four different six-sentence passages, two of which contained one of the familiarization words and two of which did not. Infants listened significantly longer to passages with the familiarization words, suggesting that they recognized the words in fluent speech. In a second experiment, Jusczyk and Aslin found that infants did equally well when familiarized with passages first and then tested on isolated words.

Subsequent research has sought to identify how infants segment words from fluent speech. Jusczyk, Houston, and Newsome (1999) suggested that English learners begin to segment words by relying on prosodic stress cues, leading them to identify strong syllables as word onsets. In a series of experiments using the same methods as Jusczyk and Aslin, they found that 1) English-learning 7.5-mo-olds segment words with strong/weak patterns (e.g., *doctor*), but not weak/strong words (e.g., *surprise*); 2) 7.5-mo-olds mis-segment weak/strong words at the beginnings of strong syllables (e.g., they segment *prize* rather than *surprise*); and 3) it is not until 10.5 mo, that they can segment weak/strong words. One prediction that follows from the claim that English-learners begin segmenting words by relying on prosodic stress cues is that they might be able to segment words in an unfamiliar language, provided that it has the same predominant word stress pattern. In fact, English-learning 9-mo-olds segment strong/weak words (e.g., *pendel*) from Dutch utterances, despite their lack of prior familiarity with the language (Houston, Jusczyk, Kuipers, Coolen & Cutler, 2000).

Although prosodic stress cues might help infants to begin to segment words from a language, such as English or Dutch, these cues are not sufficient, as 7.5-mo-olds' difficulties with weak/strong words illustrates. To succeed with words that do not begin with strong syllables, infants need to draw on other sources of information. As noted above, between 6 and 9 mo, infants demonstrate increased sensitivity to the frequency with which certain phonotactic patterns appear in words. Indeed, it has been suggested that the ability to track recurring patterns in the input may, by itself, be sufficient for segmenting words from fluent speech contexts (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996). For example, Saffran, Aslin, and Newport (1996) presented 8-mo-olds with a 2.5-minute stream of continuous speech containing no other information to word boundaries except the statistical likelihood of one syllable following another (i.e., transitional probabilities). During a subsequent test on sequences from the familiarization stream, infants responded differentially to sequences corresponding to words, as opposed to part-words, suggesting they had segmented the words.

Other investigations have focused on how infants' knowledge of patterns that they have already detected in the input could be used in segmenting words. For example, 9-mo-olds respond differently to phonotactic patterns that are likely within words, as opposed to between words—such as the last phonetic segment in one word and the first segment in the next word (Friederici & Wessels, 1993; Mattys, Jusczyk, Luce, & Morgan, 1999). Such findings suggest that infants have learned about the distribution of such sequences relative to word boundaries. The latter claim is borne out by results indicating that English-learning 9-mo-olds segmented words only from utterance contexts in which phonotactic cues suggest a likely word boundary (Mattys & Jusczyk, 2001b). Thus, by 9 mo, English-learners can use phonotactic cues to guide their segmentation of words from speech. However, their ability to use another possible word segmentation cue in the speech signal seems to require more time to develop. In particular, it has been suggested that listeners could use information about the distribution of different phonetic variants (i.e., context-sensitive allophones) of a particular phoneme to locate word boundaries (Bolinger & Gerstman, 1957; Church, 1987; Hockett, 1955; Lehiste, 1960). Allophones of a given phoneme are often restricted in terms of their positions within words. Consider some of the allophones of the English phoneme /t/. The aspirated allophone [t^h] occurs at the beginning of words, whereas the unaspirated allophone [t] is found at the ends of words. Knowledge of the contexts in which such allophones typically appear could help to identify possible word boundaries. Although 2-mo-olds discriminate distinctions between allophones that are relevant to specifying word boundaries (Hohne & Jusczyk, 1994), it is not until 10.5 mo that English-learners use this information in segmenting words (Jusczyk, Hohne, & Bauman, 1999).

Infants' use of different types of word boundary cues develops significantly between 7.5 and 10.5 mo. Nevertheless, many questions remain about how they discover these particular cues and how they learn to integrate them in segmenting words from fluent speech. Jusczyk, Houston, and Newsome (1999) suggested that the tendency for English-learners to use prosodic stress cues when they begin to segment words may arise from the fact that items that they are likely to hear spoken in isolation frequently (names and diminutives) are typically strong/weak patterns. The same authors argued that breaking the input into smaller chunks based on the occurrence of strong syllables provides a way to relate phonotactic sequences and allophones to potential word boundaries, thus facilitating the discovery of these types of segmentation cues. Presumably, infants learning languages without predominant stress patterns would have to begin with another of the potential word segmentation cues, and then learn to uncover other cues in the signal.

To this point, relatively little is known about how or when infants actually integrate the different types of word segmentation cues. It has been observed that when prosodic stress cues conflict either with statistical cues (Johnson & Jusczyk, 2001) or with phonotactic cues (Mattys et al., 1999), infants 9 mo or younger favor prosodic stress cues in segmenting speech. However, the fact that 10.5-mo-olds segment weak/strong words indicates that by this age, other types of word boundary cues can outweigh prosodic stress cues. Indeed, by 12 mo, infants' segmentation of speech seems to be governed by whether a particular parse yields items that could be possible words in a language (Johnson, Jusczyk, Cutler, & Norris, 2000). Furthermore, given the critical role (discussed above) that competition among lexical items plays in adults' recognition of words during on-line processing, one might expect that as the lexicon develops, lexical competition would become a significant factor in infants' recognition of words. Indeed, Hollich, Jusczyk, and Luce (2000) found evidence of lexical competition effects in a word learning task with 17-mo-olds. Specifically, infants learned a new word from a sparse neighborhood more readily than one from a dense neighborhood.

In concluding this section, we note that the findings discussed have all involved the segmentation of words beginning with consonants. Some recent reports suggest that when words begin with initial vowels (whose onsets tend to be less prominently marked in the speech stream than those of consonants), English-learners are not successful in segmenting these until between 13 and 16 mo of age (Mattys & Jusczyk, 2001a; Nazzi, Jusczyk & Bhagirath, Reference Note 15).

Developing Lexical Representations • Once infants start segmenting words from fluent speech they are in a position to store information about the sound patterns of possible words and begin building a lexicon in which sound patterns are linked to specific meanings. There are indications that infants store information about sound patterns of words that they hear frequently, even when they do not have a specific meaning to link to them. Jusczyk and Hohne (1997) reported that 8-mo-olds familiarized with three stories once a day for 10 days during a 2-wk period retained information about sound patterns of words that occurred frequently in the stories when tested 2 wk later. Infants who had heard the stories listened significantly longer to the lists of words from the stories than to lists of matched foil words, whereas infants who had not heard the stories displayed no preference for either type of list. The findings suggest that not only did infants segment frequently occurring words, they also encoded the sound patterns in memory.

How much information do infants store about the sound patterns of words (i.e., what is the representational specificity of the items that they store)? One suggestion is that initially, infants might only encode sufficient detail about the sound pattern of a word to distinguish it from other items already in the lexicon (Charles-Luce & Luce, 1990; 1995; Jusczyk, 1993; Walley, 1988; 1993). Some support for this view comes from studies with French infants. Hallé and Boysson-Bardies (1994) found that French 11-mo-olds listen significantly longer to words that are likely to be familiar to them than to unfamiliar words. However, Hallé and Boysson-Bardies (1996) subsequently found that French 11-mo-olds also showed a similar preference when the initial consonants of the familiar words were changed to another consonant. The latter findings suggest that infants' representations of the sound patterns of the familiar words might not be fully specified with respect to all phonetic properties. Similarly, Stager and Werker (1997) reported that, despite 14-mo-olds' abilities to discriminate two syllables [bI] and [dI] differing by a single phonetic feature, they could not succeed at a word learning task that required them to attach these two syllables to distinctive objects. Thus, the findings suggest that even though infants detect fine distinctions between different speech sounds, they might not encode (or be able to encode) the same degree of detail into their representations of words.

This view of the kind of detail encoded into early representations of words is challenged by a growing body of evidence suggesting that infants' representations of words are detailed from the outset. Indeed, Jusczyk and Aslin (1995) reported that 7.5-mo-olds familiarized with isolated repetitions of *tup*, did not listen significantly longer to a passage containing the word *cup* in each sentence. They interpreted this as an indication that infants' representations of such words might be very detailed, including enough detail to distinguish an item that differs only by a single phonetic feature in its initial consonant.

Other findings raise the possibility that infants might actually store individual exemplars of words in memory, or at least that their representations may be so detailed as to include talker specific information (Houston, Jusczyk, & Tager, Reference Note 10; Houston & Jusczyk, 2000). For example, Houston and Jusczyk (2000) familiarized infants with a pair of words produced by one talker, then tested infants recognition of the same words in passages produced by a different talker. When the talkers of different sex were used, English-learners showed no ability to generalize until about 10.5 mo of age. Furthermore, with the introduction of a 24-hr delay between familiarization and testing, even 10.5-mo-olds displayed no ability to generalize across talkers, even those of the same sex (Houston et al., 1998). By comparison, when the talker's voice was unchanged between familiarization and test, even 7.5-mo-olds recognized the words after the 24-hr delay. Thus, in line with WRAPSA's assumption that listeners store specific instances of words rather than abstract prototypes, infants' representations of sound patterns of words in the lexicon may include surface details, relating to talker voice characteristics.

Studies in which infants are required to respond to the appropriate meaning of a particular sound pattern also suggest that infants' representations of words include considerable phonetic detail. In a task with 24-mo-olds, Swingley, Pinto, and Fernald (1999) recorded the latency of infants' looking responses to a visual target after hearing its name. By systematically varying the distracter item paired with the target, they found that infants responded more slowly when the distracter shared an initial consonant with the target item (e.g., *dog* and *doll*) than when it did not (e.g., *doll* and *ball*). This finding suggests that infants' representation of these items included detailed information about the initial consonants. To determine whether such phonetic detail is stored only when infants know two different words that begin in the same way, Swingley and Aslin (2000) conducted another study. They compared 18- to 23-mo-olds' responses to correct pronunciations of a target versus mispronunciations that differed by a single phonetic feature (e.g., *baby* versus *vaby*). The latencies of infants' looking times to the correct object were significantly delayed for mispronunciations versus correct pronunciations of the targets. Swingley and Aslin argued that infants' representations of the target words must have already contained sufficient detail to distinguish them from close mispronunciations. Results from an investigation by Plunkett, Bailey, and Bryant (Reference Note 17) with 18- to 24-mo-olds are a little harder to interpret. On the one hand, they found some indication that words that infants have known for a longer period of time seem to contain more phonetic detail than newly learned words. On the other hand, no significant correlations were found between phonetic detail in the representations and increases in either age or vocabulary size.

Research exploring the nature of infants' representations of the sound patterns of words began only a few years ago. Many interesting findings have been reported, suggesting certain parallels to adults' representations of words. Nevertheless, much remains to be learned about the nature of these early representations and their development as infants' lexicons grow.

Learning about Larger Units of Linguistic Organization • Recovering the underlying meaning of an utterance depends crucially on identifying the relationships that exist among its words. At first glance, these issues might seem to be outside the realm of speech perception. However, it is critical that the listener detects the correct organizational units in an utterance rather than some arbitrary groupings of adjacent words. Consider an utterance such as, "I read the recent book *Tom Clancy wrote. Moby Dick*, it was not." A listener who treated the underlined utterance fragments as a unit would arrive at an incorrect interpretation of the utterance. Furthermore, beginning with arbitrary groupings of words would guarantee failure for someone trying to discover the grammatical organization of utterances in a language. Thus, a critical task for language learners is to detect major constituents and how they are related in utterances. Accomplishing this requires that learners parse the signal into the right sized units.

There is considerable evidence that the boundaries of units, such as clauses, are acoustically marked in both adult-directed (Lehiste, 1973; Nakatani & Dukes, 1977; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991) and child-directed speech (Bernstein Ratner, 1986; Fisher & Tokura, 1996b). Typically, pitch declines, final syllables lengthen, and pauses are likely to occur at clause boundaries. Infants are sensitive to the occurrence of these cues. For example, Hirsh-Pasek, Kemler Nelson, Jusczyk, Wright Cassidy, Druss, and Kennedy (1987) found that English-learning 7- and 10-mo-olds preferred listening to speech passages in which pauses were inserted at clause boundaries, as opposed to in the middle of clauses (see also Morgan, 1994). Furthermore, infants actually seem to use clausal units in encoding and retrieving speech information. For instance, 6-mo-olds are better able to detect clausal units embedded in continuous speech than comparable nonclausal units (Nazzi, Kemler Nelson, Jusczyk, & Jusczyk, 2000). Even 2-mo-olds remember speech information better when it is packaged in clausal units, as opposed to comparable, but nonclausal, word sequences (Mandel, Jusczyk, & Kemler Nelson, 1994; Mandel, Kemler Nelson, & Jusczyk, 1996). Thus, infants show some ability to divide utterances into units corresponding to clauses.

The situation with respect to smaller units of organization, such as phrases, is more mixed. Languages differ considerably in how they organize phrases within clauses. Hence, it is unlikely that there would be any universal properties that mark phrasal units in all languages. Languages that rely heavily on word order to convey syntactic information are likely to group words from the same phrase together. However, languages that allow for freer word orders might have elements from the same phrase in different portions of an utterance. Still, for a language such as English, phrasal units are often marked by acoustic cues (Beach, 1991; Price et al., 1991; Scott, 1982; Scott & Cutler, 1984). When such information is available in child-directed speech, English-learning 9-mo-olds not only detect it (Jusczyk, Hirsh-Pasek, Kemler Nelson, Kennedy, Woodward, & Piwoz, 1992), but also use it in encoding and remembering information contained in such units (Soderstrom, Jusczyk, & Kemler Nelson, Reference Note 20).

Nevertheless, even within languages that rely on word order, such as English, syntactic phrases are not consistently marked in the input that the child receives (Fisher & Tokura, 1996a). Because the markers present in speech tend to relate to prosodic phrases, they do not always pick out the same types of syntactic phrases. Indeed, English-learning 9-mo-olds display listening preferences that accord with the prosodic, rather than with the syntactic, organization of the utterances (Gerken, Jusczyk, & Mandel, 1994). However, even the division of utterances into units corresponding to prosodic phrases still seems to be helpful in learning to identify syntactic phrases by facilitating the discovery of syntactic units such as grammatical morphemes (Jusczyk, 1999). In English, certain morphemes, such as function words, typically occur only at particular locations inside phrasal units. For instance, *the* marks the beginning of a noun phrase, and is extremely unlikely to occur as the last word of a phrasal unit. Hence, grouping the input into prosodic phrases and noting regularities in how certain morphemes are distributed within such phrases may help in delineating their syntactic roles.

In conclusion, infants are sensitive to the information in speech that marks important units of linguistic organization. Although sensitivity to such information is not sufficient to ensure the acquisition of grammatical structure, the ability to group words into units corresponding to clauses and phrases is an important step towards discovering the specific syntactic organization of the native language.

Summary and Conclusions

Early research on the speech perception capacities of adults and infants alike tended to focus on issues concerning phonetic perception. As such, these earlier investigations most often were directed at the nature of the mechanisms responsible for our abilities to identify and extract individual phonemes from the speech signal. Although these issues continue to be of importance in speech research, the primary focus of research with infants and adults has shifted somewhat toward understanding how our speech perception capacities are used in segmenting and recognizing words in fluent speech. Thus, although the primary focus of earlier research asked about what speech perception capacities are, current investigations center more on how such capacities are used in understanding and learning language.

Acknowledgments:

Preparation of the present manuscript was supported by a Senior Scientist Award from NIMH (01490) and a Research Grant from NICHD (15795) to PWJ and a Research Grant from NIDCD (0265801) to PAL. The authors would like to thank Ann Marie Jusczyk, Elizabeth Johnson, and Jan Charles-Luce for helpful comments on a previous version of the present manuscript. The views expressed here are those of the authors, and not necessarily those of the funding agencies.

Address for correspondence: Paul A. Luce, Ph.D., 364 Park Hall, Department of Psychology, University of Buffalo, NY 14260.

Received March 1, 2001; accepted July 18, 2001

References

Abbs, J. H, & Sussman, H. M. (1971). Neurophysiological feature detectors and speech perception: A discussion of theoretical implications. *Journal of Speech and Hearing Research*, *14*, 23–36.

Ades, A. E. (1974). How phonetic is selective adaptation? Experiments on syllable and vowel environment. *Perception & Psychophysics*, *16*, 61–67.

Allopena, P. D, Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419–439.

Andruski, J. E, Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, *52*, 163–187.

Aslin, R. N, Saffran, J. R., & Newport, E. L. (1998). Computation of probability statistics by 8-month-old infants. *Psychological Science*, *9*, 321–324.

Bard, E. G., Shillcock, R. C., & Altmann, G. T. M. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effect of subsequent context. *Perception & Psychophysics*, *44*, 395–408.

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue-trading relations. *Journal of Memory and Language*, *30*, 644–663.

Bentin, S., & Mann, V. (1990). Masking and stimulus intensity effects on duplex perception: A confirmation of the dissociation between speech and nonspeech modes. *Journal of the Acoustical Society of America*, *88*, 64–74.

Bernstein, J., & Franco, H. (1996). Speech recognition by computer. In N. J. Lass (Ed.), *Principles of Experimental Phonetics*, (pp. 408–434). St. Louis: Mosby.

Bernstein Ratner, N. (1986). Durational cues which mark clause boundaries in mother-child speech. *Phonetics*, *14*, 303–309.

Bertoncini, J., Bijeljac-Babic, R., Blumstein, S. E., & Mehler, J. (1987). Discrimination in neonates of very short CV's. *Journal of the Acoustical Society of America*, *82*, 31–37.

Best, C. T. (1993).Emergence of language-specific constraints in perception of native and non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life.* (pp. 289–304). Dordrecht: Kluwer.

Best, C. T. (1995). Learning to perceive the sound patterns of English. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in Infancy Research*, (Vol. 9, pp. 217–304). Norwood, NJ: Ablex.

Best, C. T, Lafleur, R., & McRoberts, G. W. (1995). Divergent developmental patterns for infants' perception of two nonnative contrasts. *Infant Behavior and Development*, *18*, 339–350.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of the perceptual re-organization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 345–360.

Best, C. T, Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, *29*, 191–211.

Bloomfield, L. (1933). *Language*. New York: Holt.

Blumstein, S. E., & Stevens, K. N. (1978). Acoustic invariance for place of articulation in stops and nasals across syllabic context. *Journal of the Acoustical Society of America*, *62*, S26.

Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America*, *67*, 648–662.

Bolinger, D. L., & Gerstman, L. J. (1957). Disjuncture as a cue to constraints. *Word*, *13*, 246–255.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*, 93–125.

Brown, R. W., & Hildum, D. C. (1956). Expectancy and the perception of syllables. *Language*, *32*, 411–419.

Burton, M. W., Baum, S. R., & Blumstein, S. E. (1989). Lexical effects on the phonetic categorization of speech: The role of acoustic structure. *Journal of Experiment Psychology: Human Perception and Performance*, *15*, 567–575.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology*, *33*, 111–153.

Charles-Luce, J., & Luce, P. (1995). An examination of similarity neighborhoods in young children's receptive vocabularies. *Journal of Child Language*, 22.

Charles-Luce, J., & Luce, P. A. (1990). Similarity neighborhoods of words in young children's lexicons. *Journal of Child Language*, *17*, 205–215.

Charles-Luce, J., Luce, P. A., & Cluff, M. S. (1990). Retroactive influences of syllable neighborhoods. In G. Altmann (Ed.), *Cognitive Models of Speech Perception: Psycholinguistic and Computational Perspectives.* Cambridge, MA: MIT Press.

Christie, W. (1977). Some multiple cues for juncture in English. *General Linguistics*, *17*, 212–222.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 521–533.

Church, K. (1987). Phonological parsing and lexical retrieval. *Cognition*, *25*, 53–69.

Clements, G. N., & Keyser, S. J. (1983). *CV Phonology: A Generative Theory of the Syllable*. Cambridge, MA: MIT Press

Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken bisyllabic words. *Journal of Experiment Psychology: Human Perception and Performance*, *16*, 551–563.

Cole, R. A., & Scott, B. (1974). Toward a theory of speech perception. *Psychological Review*, *81*, 348–374.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of words have a special status in auditory word recognition? *Journal of Memory and Language, 32*, 193–210.

Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, *37*, 463–480.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America*, *24*, 597–606.

Cooper, W. E. (1975). Selective adaptation to speech. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, & D. B. Pisoni (Eds.), *Cognitive Theory*, (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cooper, W. E., & Blumstein, S. A. (1974). A "labial" feature analyzer in speech perception. *Perception & Psychophysics*, *15*, 591–600.

Craik, F. I. M. ,& Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, *26*, 274–284.

Creelman, C. D. (1957). Case of the unkown talker. *Journal of the Acoustical Society of America*, *29*, 655.

Cutler, A., (1989). Auditory lexical access: Where do we start? In W.D. Marslen-Wilson (Ed.), *Lexical Representation 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. (1994). Segmentation problems, rhythmic solutions. *Lingua*, *92*, 81–104.

Cutler, A. (1996). Prosody and the word boundary problem. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax* (pp. 87–99). Mahwah, NJ: Lawrence Erlbaum Associates.

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 and Language*, *2*, 133–142.

Cutler, A., Mehler, J., Morris, D., & Segui, J. (1987). Phoneme identification and lexicon. *Cognitive Psychology*, *19*, 141–177.

Cutler, A., Norris, D., & Williams, J. N. (1987). A note on the role of phonological expectation in speech segmentation. *Journal of Memory and Language*, *26*, 480–487.

Cutler, A., & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 113–121.

Davidson-Nielson, N. (1974). Syllabification in English words with medial sp, st, sk. *Journal of Phonetics, 2*, 15–45.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, *27*, 769–773.

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, *16*, 513–521.

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception & Psychophysics*, *18*, 341–347.

Eimas, P. D. (1991). Comment: Some effects of language acquisition on speech perception. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the Motor Theory of Speech Perception* (pp. 111–116). Hillsdale, NJ: Erlbaum.

Eimas, P. D., Cooper, W. E., & Corbit, J. D. (1973). Some properties of linguistic feature detectors. *Perception & Psychophysics*, *13*, 247–252.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, *4*, 99–109.

Eimas, P. D., & Miller, J. L. (1980a). Contextual effects in infant speech perception. *Science*, *209*, 1140–1141.

Eimas, P. D., & Miller, J. L. (1980b). Discrimination of the information for manner of articulation. *Infant Behavior & Development*, *3*, 367–375.

Eimas, P. D., & Miller, J. L. (1992). Organization in the perception of speech by young infants. *Psychological Science*, *3*, 340–345.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 303–306.

Eukel, B. (1980). Phonotactic basis for word frequency effects: Implications for lexical distance metrics. *Journal of the Acoustical Society of America*, *68*, s33.

Fant, C., G. M. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B. d., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*, 477–501.

Fisher, C., & Tokura, H. (1996a). Prosody in speech to infants: Direct and indirect acoustic cues to syntactic structure. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax* (pp. 343–364). Mahwah, NJ: Erlbaum.

Fisher, C. L, & Tokura, H. (1996b). Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development*, *67*, 3192–3218.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues* (pp. 229–273). Timonium, MD: York Press.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New Approaches to Language Mechanisms.* Amsterdam: North Holland.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3–28.

Fowler, C. A. (1990). Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America*, *88*, 1236–1249.

Fowler, C. A., & Rosenblum, L. D. (1990). Duplex perception: A comparison of monosyllables and slamming of doors. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 742–754.

Fowler, C. A., & Rosenblum, L. D. (1991). Perception of the phonetic gesture. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the Motor Theory* (pp. 33–59). Hillsdale, NJ: Erlbaum.

Frauenfelder, U. & Peeters, G. (1990). Lexical segmentation in TRACE: An exercise in simulation. In G. T Altmann (Ed.), *Cognitive Models of Speech Processing* (pp. 50–86). Cambridge, MA: MIT Press.

Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge and its use in infant speech perception. *Perception & Psychophysics*, *54*, 287–295.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, *5*, 171–189.

Fujimura, O. (1976). Syllables as concatenated demisyllables and affixes. *Journal of Acoustical Society of America*, *59*, S55.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and cognitive Processes*, *12*, 613–656.

Gaskell, M. G., & Marslen-Wilson, W. D. (1999). Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science*, *23*, 439–462.

Gerken, L. A., Jusczyk, P. W., & Mandel, D. R. (1994). When prosody fails to cue syntactic structure: Nine-month-olds' sensitivity to phonological vs syntactic phrases. *Cognition*, *51*, 237–265.

Gerstman, L. (1968). Classification of self-normalized vowels. *EEE Transactions on Audio and Electroacoustics* (ACC-16), 78–80.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–279.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, *28*, 501–518.

Goldinger, S. D., Luce, P. A., Pisoni, D. B., & Marcario, J. K. (1992). Form-based priming in spoken word recognition: The roles of competitive activation and response biases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1210–1237.

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 152–162.

Goodman, J. C., & Huttenlocher, J. (1988). Do we know how people identify spoken words? *Journal of Memory and Language*, *27*, 684–689.

Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception & Psychophysics*, *38*, 299–310.

Grieser, D., & Kuhl, P. K. (1989). The categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, *25*, 577–588.

Hall, M. D., & Pastore, R. E. (1992). Musical duplex perception: Perception of figurally good chords with subliminally distinguishing tones. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 752–762.

Hallé, P., & Boysson-Bardies, B. D. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior & Development*, *19*, 463–481.

Hallé, P. A., & de Boysson-Bardies, B. (1994). Emergence of an early receptive lexicon: Infants' recognition of words. *Infant Behavior and Development*, *17*, 119–129.

Harnad, S. (Ed.) (1987). *Categorical Perception: The Groundwork of Cognition*. Cambridge: Cambridge University Press.

Harris, Z. S. (1955). From phoneme to morpheme. *Language*, *31*, 190–222.

Healy, A. F., & Cutting, J. E. (1976). Units of speech perception: Phoneme and syllable. *Journal of Verbal Learning and Verbal Behavior*, *15*, 73–83.

Hillenbrand, J. M., Minifie, F. D., & Edwards, T. J. (1979). Tempo of spectrum change as a cue in speech sound discrimination by infants. *Journal of Speech and Hearing Research*, *22*, 147–165.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.

Hintzman, D. L., & Caulton, D. A. (1997). Recognition memory and modality judgments: A comparison of retrieval dynamics. *Journal of Memory and Language*, *37*, 1–23.

Hintzman, D. L., & Curran, T. (1997). Comparing retrieval dynamics in recognition memory and lexical decision. *Journal of Experimental Psychology: General*, *126*, 228–247.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Wright Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, *26*, 269–286.

Hockett, C. F. (1955). *A Manual of Phonology*. (Vol. 21). Baltimore: Waverly Press.

Hohne, E. A., & Jusczyk, P. W. (1994). Two-month-old infants' sensitivity to allophonic differences. *Perception & Psychophysics*, *56*, 613–623.

Hollich, G., Jusczyk, P. W., & Luce, P. A. (2000). Infant sensitivity to lexical neighborhoods during word learning. *Journal of the Acoustical Society of America*, *108*, 2481.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1570–1582.

Houston, D. M., Jusczyk, P. W., Kuipers, C., Coolen, R., & Cutler, A. (2000). Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin & Review*, *7*, 504–509.

Hubel, D. H., & Wiesel, T. N. (1965). Binocular interaction in striate cortex of kittens reared with artificial squint. *Journal of Neurophysiology*, *28*, 1041–1059.

Hurvich, L. M., & Jameson, D. (1969). Human color perception. *American Scientist*, *57*, 143–166.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, *97*, 553–562.

Jackson, A., & Morton, J. (1984). Facilitation of auditory recognition. *Memory and Cognition*, *12*, 568–574.

Jakobson, R., Fant, C. G. M., & Halle, M. (1952). *Preliminaries to Speech Analysis*. Cambridge, MA: MIT Press.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44,* 548–567.

Johnson, E. K., Jusczyk, P. W., Cutler, A., & Norris, D. (2000). 12-month-olds show evidence of a possible word constraint. *Journal of the Acoustical Society of America*, *108*, 2481.

Jusczyk, P. W. (1977). Perception of syllable-final stops by two-month-old infants. *Perception and Psychophysics*, *21*, 450–454.

Jusczyk, P. W. (1993). From general to language specific capacities: The WRAPSA Model of how speech perception develops. *Journal of Phonetics*, *21*, 3–28.

Jusczyk, P. W. (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.

Jusczyk, P. W. (1999). Narrowing the distance to language: One step at a time. *Journal of Communication Disorders*, *32*, 207–222.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology*, *29*, 1–23.

Jusczyk, P. W., Copan, H., & Thompson, E. (1978). Perception by two-month-olds of glide contrasts in multisyllabic utterances. *Perception & Psychophysics*, *24*, 515–520.

Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Development*, *64*, 675–687.

Jusczyk, P. W., Friederici, A. D., Wessels, J., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, *32*, 402–420.

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, *24*, 252–293.

Jusczyk, P. W., & Hohne, E. A. (1997). Infants' memory for spoken words. *Science*, *277*, 1984–1986.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*, *61*, 1465–1476.

Jusczyk, P. W., Houston, D., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*, 159–207.

Jusczyk, P. W., Luce, P. A., & Charles Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*, 630–645.

Jusczyk, P. W., Pisoni, D. B., & Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month old infants. *Cognition*, *43*, 253–291.

Jusczyk, P. W., Pisoni, D. B., Walley, A. C., & Murray, J. (1980). Discrimination of the relative onset of two-component tones by infants. *Journal of the Acoustical Society of America*, *67*, 262–270.

Jusczyk, P. W., Rosner, B. S., Reed, M., & Kennedy, L. J. (1989). Could temporal order differences underlie 2-month-olds' discrimination of English voicing contrasts? *Journal of the Acoustical Society of America*, *85*, 1741–1749.

Jusczyk, P. W., & Thompson, E. J. (1978). Perception of a phonetic contrast in multisyllabic utterances by two-month-old infants. *Perception & Psychophysics*, *23*, 105–109.

Karzon, R. G. (1985). Discrimination of a polysyllabic sequence by one- to four-month-old infants. *Journal of Experimental Child Psychology*, *39*, 326–342.

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

Klatt, D. (1989). Review of selected models of speech perception. In W. Marslen-Wilson (Ed.), *Lexical Representation and Process* (pp. 169–226). Cambridge, MA: Bradford.

Kuhl, P. K. (1976). Speech perception in early infancy: the acquisition of speech sound categories. In S. K. Hirsh, D. H. Eldridge, I. J. Hirsh, & S. R. Silverman (Eds.), *Hearing and Davis: Essays Honoring Hallowell Davis* (pp. 265–280). St. Louis: Washington University Press.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, *66*, 1668–1679.

Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America*, *70*, 340–349.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, *6*, 263–285.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*, 93–107.

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (pp. 259–274). Dordrecht: Kluwer.

Kuhl, P. K., Andruski, J. E., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units addressed to infants. *Science*, *277*, 684–686.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science*, *190*, 69–72.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, *63*, 905–917.

Kuhl, P. K., & Miller, J. D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception & Psychophysics*, *31*, 279–292.

Kuhl, P. K., & Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception & Psychophysics*, *32*, 542–550.

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America*, *73*, 1003–1010.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experiences alter phonetic perception in infants by 6 months of age. *Science*, *255*, 606–608.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of Acoustical Society of America*, *29*, 98–104.

Lalonde, C. E., & Werker, J. F. (1995). Cognitive influences on cross-language speech perception in infancy. *Infant Behavior and Development*, *18*, 459–475.

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, *20*, 215–225.

Lehiste, I. (1960). *An Acoustic-Phonetic Study of Internal Open Juncture*. New York: S. Karger.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, *51*, 2018–2024.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa*, *7*, 107–122.

Levitt, A., Jusczyk, P. W., Murray, J., & Carden, G. (1988). The perception of place of articulation contrasts in voiced and voiceless fricatives by two-month-old infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 361–368.

Liberman, A. M. (1970). Some characteristics of perception in the speech mode. In D. A. Hamburg (Ed.), *Peceptions and Its Disorders: Proceedings of the Association for Research in Nervous and Mental Diseases* (pp. 238–254). Baltimore: Williams & Wilkins.

Liberman, A. M. (1996). Some assumptions about speech and how they changed. In A. M. Liberman (Ed.), *Speech: A Special Code* (pp. 1–44). Cambridge, MA: Bradford Books/MIT Press.

Liberman, A. M., Cooper, F. S., Harris, K. S., & MacNeilage, P. F. (1963). *A motor theory of speech perception*. Paper presented at the Proceedings of the Speech Communication Seminar, Stockholm: Royal Institute of Technology, Speech Transmission Laboratory.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. G. (1967). Perception of the speech code. *Psychological Review*, *74*, 431–461.

Liberman, A. M., Delattre, P. C., Gerstman, L. J., & Cooper, F. S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. *Journal of Experimental Psychology*, *52*, 127–137.

Liberman, A. M., DeLattre, P. D., & Cooper, F. S. (1952). The role of selected stimulus variables in the perception of unvoiced stop consonants. *American Journal of Psychology*, *65*, 497–516.

Liberman, A. M., Harris, K. S., Eimas, P. D., Lisker, L., & Bastian, J. (1961). An effect of learning on speech perception: The discrimination of durations of silence with and without phonetic significance. *Language and Speech*, *54*, 175–195.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358–368.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. L. (1961). The discrimination of relative-onset time of the components of certain speech and non-speech patterns. *Journal of Experimental Psychology*, *61*, 379–388.

Liberman, A. M., Isenberg, D., & Rakerd, B. (1981). Duplex perception of cues for stop consonants: Evidence for a phonetic mode. *Perception & Psychophysics*, *30*, 133–143.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Science*, *4*, 187–196.

Lieberman, P., Crelin, E. S., & Klatt, D. H. (1972). Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee. *American Anthropologist*, *74*, 287–307.

Lively, S. E., & Pisoni, D. B. (1997). On prototypes and phonetic categories: A critical magnet effect in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 1665–1679.

Locke, S., & Kellar, L. (1973). Categorical perception in a nonlinguistic mode. *Cortex*, *9*, 353–367.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1989). Training Japanese listeners to identify /r/ and /l/. *Journal of the Acoustical Society of America*, *85*, 137–138.

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics*, *39*, 155–158.

Luce, P. A. (2001). The minimal discrepancy hypothesis in spoken word recognition. Unpublished manuscript.

Luce, P. A., & Cluff, M. S. (1998). Delayed commitment in spoken word recognition: Evidence from Cross-modal Priming. *Perception & Psychophysics*, *60*, 484–490.

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, *62*, 615–625.

Luce, P. A., & Large, N. (2001). Phonotactics, neighborhood density, and entropy in spoken word recognition. *Language and Cognitive Processes, 16,* 565–581.

Luce, P. A., & Lyons, E. A. (1998). Specificity of memory representations for spoken words. *Memory and Cognition*, *26*, 708–715.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*, 1–36.

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. Altmann (Ed.), *Cognitive Models of Speech Perception: Psycholinguistic and Computational Perspectives* (pp. 122–147). Cambridge, MA: MIT Press.

Mandel, D. R., Jusczyk, P. W., & Kemler Nelson, D. G. (1994). Does sentential prosody help infants to organize and remember speech information? *Cognition*, *53*, 155–180.

Mandel, D. R., Kemler Nelson, D. G., & Jusczyk, P. W. (1996). Infants remember the order of words in a spoken sentence. *Cognitive Development*, *11*, 181–196.

Marslen-Wilson, W. D. (1987). Parallel processing in spoken word recognition. *Cognition*, *25*, 71–102.

Marslen-Wilson, W. D. (1989). Access and integration: Projecting sound onto meaning. In W. D. Marslen-Wilson (Ed.), *Lexical Access and Representation* (pp. 3–24). Cambridge, MA: Bradford.

Marslen-Wilson, W. D., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, *101*, 3–33.

Marslen-Wilson, W., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1376–1392.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*, 1–71.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63.

Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 576–585.

Massaro, D. W. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, *79*, 124–145.

Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception & Psychophysics*, *34*, 338–348.

Mattingly, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. (1971). Discrimination in speech and non-speech modes. *Cognitive Psychology*, *2*, 131–157.

Mattys, S. (1997). The use of time during lexical processing and segmentation: A review. *Psychonomic Bulletin & Review*, *4*, 310–329.

Mattys, S. L., & Jusczyk, P. W. (2001a). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 644–655.

Mattys, S. L., & Jusczyk, P. W. (2001b). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*, 91–121.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Word segmentation in infants: How phonotactics and prosody combine. *Cognitive Psychology*, *38*, 465–494.

McNeill, D., & Lindig, K. (1973). The perceptual reality of the phoneme, syllables, words, and sentences. *Journal of Verbal Learning and Verbal Behavior*, *12*, 419–430.

Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, *20*, 298–305.

Mehler, J., Segui, J., & Frauenfelder, U. (1981). The role of syllable in language acquisition and perception. In T. F. Myers, J. Laver, & J. Anderson (Eds.), *The Cognitive Representation of Speech* (pp. 295–333). Amsterdam: North-Holland.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

McClelland, J. L., & Rumelhart, D. E. (1981). And interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality and word-final ambiguity. *Journal of Experiment Psychology: Human Perception and Performance*, *17*, 433–443.

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. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experiment Psychology: Human Perception and Performance*, *25*, 1363–1389.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–96.

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, *41*, 329–335.

Miller, G. A., & Taylor, W. G. (1948). The perception of repeated bursts of noise. *Journal of the Acoustical Society of America*, *20*, 171–182.

Miller, J. D., Weir, C. C., Pastore, L., Kelly, W. J., & Dooling, R. J. (1976). Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, *60*, 410–417.

Miller, J. L., Connine, C. M., Schermer, T., & Kluender, K. R. (1983). A possible auditory basis for internal structure of phonetic categories. *Journal of the Acoustical Society of America*, *73*, 2124–2133.

Miller, J. L., & Eimas, P. D. (1977). Studies on the perception of place and manner of articulation: A comparison of the labial-alveolar and nasal-stop distinctions. *Journal of the Acoustical Society of America*, *61*, 835–845.

Miller, J. L., & Eimas, P. D. (1983). Studies on the categorization of speech by infants. *Cognition*, *13*, 135–165.

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.

Mills, C. B. (1980). Effects of context on reaction time to phonemes. *Journal of Verbal Learning and Verbal Behavior*, *19*, 75–83.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception & Psychophysics*, *18*, 331–340.

Moffitt, A. R. (1971). Consonant cue perception by twenty-to-twenty-four-week old infants. *Child Development*, *42*, 717–731.

Morgan, J. L. (1994). Converging measures of speech segmentation in prelingual infants. *Infant Behavior & Development*, *17*, 387–400.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*, 165–178.

Morton, J. Word recognition. (1979). In J. Morton & J. D. Marshall (Eds.), *Psycholinguistics 2: Structures and Processes* (pp. 107–156). Cambridge, MMA.: MIT Press.

Morse, P. A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology*, *13*, 477–492.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, *85*, 365–378.

Nakatani, L., & Dukes, K. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, *62*, 714–719.

Nazzi, T., Kemler Nelson, D. G., Jusczyk, P. W., & Jusczyk, A. M. (2000). Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy*, *1*, 123–147.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, *85*, 2088–2113.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.

Norris, D., McQueen, J., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Brain and Behavioral Sciences*, *23*, 299–325.

Norris, D., McQueen, J. M., Cutler, A., and Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, *34*, 191–243.

Nusbaum, H. C., Schwab, E. C., & Sawusch, J. R. (1983). The role of the "chirp" identification in duplex perception. *Perception & Psychophysics*, *33*, 323–332.

Nygaard, L. C. (1993). Phonetic coherence in duplex perception: Effects of acoustic differences and lexical status. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 268–286.

Nygaard, L. C., & Eimas, P. D. (1990). A new version of duplex perception: Evidence for phonetic and nonphonetic fusion. *Journal of the Acoustical Societyof America*, *88*, 75–86.

Orr, D. B., Friedman, H. L., & Williams, J. C. C. (1965). Trainability of listening comprehension of speech discourse. *Journal of Educational Psychology*, *56*, 148–156.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 309–328.

Pastore, R. E., Ahroon, W. A., Buffuto, K. A., Friedman, C. J., Puleo, J. S., & Fink, E. A. (1977). Common factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 686–696.

Pastore, R. E., Schmeckler, M. A., Rosenblum, L., & Szczesiul, R. (1983). Duplex perception with musical stimuli. *Perception & Psychophysics*, *33*, 469–473.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*, 175–184.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America*, *61*, 1352–1361.

Pisoni, D. B. (1992). Some comments on talker normalization. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech Perception, Production, and Linguistic Structure* (pp. 143–151). Tokyo: IOS Press.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 297–314.

Pisoni, D. B., Carrell, T. D., & Gans, S. J. (1983). Perception of the duration of rapid spectrum changes: Evidence for context effects with speech and nonspeech signals. *Perception & Psychophysics*, *34*, 314–322.

Pisoni, D. B., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, *25*, 21–52.

Pitt, M. A., & Samuel, A. G. (1995). Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology*, *29*, 149–188.

Polka, L., & Bohn, O.-S. (1996). Cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America*, *100*, 577–592.

Pollack, I. (1952). The information in elementary auditory displays. *Journal of the Acoustical Society of America*, *24*, 745–749.

Pollack, I. (1953). The information in elementary auditory displays II. *Journal of the Acoustical Society of America*, *25*, 765–769.

Prather, P., & Swinney, D. (1977). *Some effects of syntactic context upon lexical access*. Presented at a meeting of the American Psychological Association, San Francisco, August 26, 1977.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, *90*, 2956–2970.

Reitveld, A. C. M., & Koopmans-van Beinum, F. J. (1987). Vowel reduction and stress. *Speech communication, 6*, 217–229.

Remez, R. E. (1980). Susceptability of a stop consonant to adaptation on a speech-nonspeech continuum: Further evidence against feature detectors in speech perception. *Perception & Psychophysics*, *27*, 17–23.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional cues. *Science*, *212*, 947–950.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.

Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, *31*, 307–314.

Samuel, A. G. (2000). Some empirical tests of Merge's architecture. *Language and Cognitive Processes,* in press.

Savin, H. B., & Bever, T. G. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*, *9*, 295–302.

Sawusch, J. R. (1977). Processing of place information in stop consonants. *Perception & Psychophysics*, *22*, 417–426.

Sawusch, J. R. (1992). Auditory metrics for speech perception. In M. E. H. Schouten (Ed.), *The Processing of Speech: From the Auditory Periphery to Word Recognition* (pp. 315–321). Berlin: Mouton de Gruyter.

Sawusch, J. R., & Jusczyk, P. W. (1981). Adaptation and contrast in the perception of voicing. *Journal of Experimental Psychology: Human perception and performance*, *7*, 408–421.

Schacter, D. L. (1990). Perceptual representation systems and implicit memory: Toward a resolution of the multiple memory systems debate. *Annals of the New York Academy of Sciences*, *608*, 543–571.

Schacter, D. L. (1992). Understanding implicit memory: A cognitive neuroscience approach. *American Psychologist*, *47*, 559–569.

Schacter, D. L., & Church, B. A. (1992). Auditory priming: Implicit and explicit memory for words and voice. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 915–930.

Schreuder, R., & Baayen, R. H. (1994). Prefix stripping re-revisited. *Journal of Memory and Language*, *33*, 357–375.

Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America*, *71*, 996–1007.

Scott, D. R., & Cutler, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior*, *23*, 450–466.

Searle, C. L., Jacobson, J. Z., & Rayment, S. G. (1979). Phoneme recognition based on human audition. *Journal of the Acoustical Society of America*, *65*, 799–809.

Shankweiler, D. P., Strange, W., & Verbrugge, R. R. (1977). Speech and the problem of perceptual constancy. In R. Shaw & J. Bransford (Eds.), *Perceiving, Acting, and Knowing: Toward an Ecological Psychology* (pp. 315–345). Hillsdale, NJ: Erlbaum.

Sheffert, S. M. (1998a). Contributions of surface and conceptual information on spoken word and voice recognition. *Perception & Psychophysics 60,* 1141–1152.

Sheffert, S. M. (1998b). Voice-specificity effects on auditory word priming. *Memory & Cognition, 26,* 591–598.

Shillcock, R. (1990). Lexical hypotheses in continuous speech. In G. T Altmann (Ed.), *Cognitive Models of Speech Processing* (pp. 24–49). Cambridge, MA: MIT Press.

Slowiaczek, L. M., McQueen, J. M., Soltano, E. G., & Lynch, M. Phonological representations in prelexical speech processing: Evidence from form-based priming. *Journal of Memory and Language*, *43*, 530–560.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*, 381–382.

Stevens, K. N. (1960). Toward a model for speech recognition. *Journal of the Acoustical Society of America*, *32*, 47–55.

Stevens, K. N. (1972). The quantal nature of speech. In J. E. E. David & P. B. Denes (Eds.), *Human Communication: A Unified View*. New York: McGraw-Hill.

Stevens, K. N. (1998). *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *64*, 1358–1368.

Strange, W., & Jenkins, J. J. (1978). Role of linguistic experience in the perception of speech. In R. D. Walk & H. L. Pick (Eds.), *Perception and Experience* (pp. 125–169). New York: Plenum.

Streeter, L. A. (1976). Language perception of 2-month old infants shows effects of both innate mechanisms and experience. *Nature*, *259*, 39–41.

Studdert-Kennedy, M. G. (1974). The perception of speech. In T. A. Sebeok (Ed.), *Current trends in linguistics (Vol. 12)* (pp. 2349–2385). The Hague: Mouton.

Studdert-Kennedy, M. G., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. *Psychological review*, *77*, 234–249.

Suomi, K. (1984). On talker and phoneme identification conveyed by vowels: A whole spectrum approach to the normalization problem. *Speech Communication*, *3*, 199–209.

Suomi, K. (1993). An outline of a developmental model of adult phonological organization and behavior. *Journal of Phonetics*, *21*, 29–60.

Sussman, H. M. (1986). A neuronal model of vowel normalization and representation. *Brain and Language*, *28*, 12–23.

Sussman, H. M., McCaffrey, H. A., & Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America*, *90*, 1309–1325.

Sussman, J. E., & Lauckner-Morano, V. J. (1995). Further tests of the "perceptual magnet effect" in the perception of [i]: Identification and change/no change discrimination. *Journal of the Acoustical Society of America*, *97*, 539–552.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in young children. *Cognition*, *76*, 147–166.

Swingley, D., Pinto, J., & Fernald, A. (1999). Continuous processing in word recognition at 24-months. *Cognition*, *71*, 73–108.

Swinney, D. (1981). Lexical processing during sentence comprehension: Effects of higher order constraints and implications for representation. In T. Myers, J. Laver, & J. Anderson (Eds.), *The Cognitive Representation of Speech* (pp.201–209). The Netherlands: North Holland Publishing Company.

Swinney, D. A., & Prather, P. (1980). Phonemic identification in a phoneme monitoring experiment: The variable role of uncertainty about vowel contexts. *Perception & Psychophysics*, *27*, 104–110.

Swoboda, P., Kass, J., Morse, P. A, & Leavitt, L. A. (1978). Memory factors in infant vowel discrimination of normal and at-risk infants. *Child Development*, *49*, 332–339.

Swoboda, P., Morse, P. A., & Leavitt, L. A. (1976). Continuous vowel discrimination in normal and at-risk infants. *Child Development*, *47*, 459–465.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, *79*, 1086–1100.

Tabossi, P., Burani, C., & Scott, D. (1995). Word identification in fluent speech. *Journal of Memory and Language*, *34*, 440–467.

Tartter, V. C., & Eimas, P. D. (1975). The role of auditory and phonetic feature detectors in the perception of speech. *Perception & Psychophysics*, *18*, 293–298.

Thyer, N., Hickson, L., & Dodd, B. (2000). The perceptual magnet effect in Australian English vowels. *Perception & Psychophysics*, *62*, 1–20.

Trehub, S. E. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology*, *9*, 91–96.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, *47*, 466–472.

van de Weijer, J. (1998). *Language Input for Word Discovery*. Unpublished Ph.D., University of Nijmegen, Nijmegen.

Verbrugge, R. R., & Rakerd, B. (1986). Evidence of talker-independent information for vowels. *Language and Speech*, *29*, 39–57.

Verbrugge, R. R., & Shankweiler, D. (1977). Prosodic information for vowel identity [Abstract]. *Journal of the Acoustical Society of America*, *61*, S39..

Verbrugge, R. R., Strange, W., Shankweiler, D. P., & Edman, T. R. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America*, *60*, 198–212.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, *40*, 47–62.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in spoken word perception. *Psychological Science*, *9*, 325–329.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*, 374–408.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1996). Phonotactic and metrical influences on the perception of spoken and nonsense words. *Language and Speech*, *40*, 47–62.

Vroomen, J., & de Gelder, B. (1995). Metrical Segmentation and Lexical Inhibition in Spoken Word Recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 98–108.

Vroomen, J., & de Gelder, B. (1997). Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 710–720.

Walley, A. C. (1988). Spoken word recognition by young children and adults. *Cognitive Development*, *3*, 137–165.

Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review*, *13*, 286–350.

Warren, R. M. (1974). Auditory temporal discrimination by trained listeners. *Cogntitive Psychology*, *6*, 237–256.

Warren, P. & Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics*, *41*, 262–275.

Warren, P. & Marslen-Wilson, W. (1988). Cues to lexical choice: Discriminating place and voice. *Perception & Psychophysics*, *31*, 21–30.

Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. (1969). Auditory sequence: Confusion of patterns other than speech or music. *Science*, *164*, 586–587.

Werker, J. F. (1991). The ontogeny of speech perception. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the Motor Theory of Speech Perception* (pp. 91–109). Hillsdale, NJ: Erlbaum.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, *24*, 672–683.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.

Whalen, D. H., & Liberman, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science*, *237*, 169–171.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics*, *35*, 49–64.

Whalen, D. H. (1991). subcategorical phonetic mismatches and lexical access. *Perception & Psychophysics*, *50*, 351–360.

Wickelgren, W. A. (1969). Context-sensitive coding in speech recognition, articulation, and development. In K. N. Leibovic (Ed.), *Information Processing in the Nervous System* (pp. 201–209). Berlin: Springer-Verlag.

Zatorre, R. J., & Halpern, A. R. (1979). Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Perception & Psychophysics*, *26*, 384–395.

Reference Notes

1. Abramson, A. S., & Lisker, L. (1967). *Discriminability along the voice continuum: Cross language tests.* Paper presented at the Proceedings of the Sixth International Congress of Phonetic Sciences, Prague.
2. Bailey, P. J., Summerfield, A. Q., & Dorman, M. F. (1977). *On the identification of sine-wave analogs of certain speech sounds* (Status Report on Speech Research SR 51–52): Haskins Laboratories, New Haven, CT.
3. Bentin, S., & Mann, V. A. (1983). *Selective effects of masking on speech and nonspeech in the duplex perception paradigm* (SR 76). New Haven, CT: Haskins Laboratories Report on Speech Research.
4. Blechner, M. J. (1977). *Musical skill and the categorical perception of harmonic mode.* Unpublished Unpublished doctoral dissertation, Yale University.
5. Cooper, F. S., Liberman, A. M., & Borst, J. M. (1951). The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proceedings of the National Academy of Sciences*, *37*, 318–325.
6. Cutler, A., Norris, D., & McQueen, J. M. (2000). Tracking Trace's troubles. Proceedings of Spoken Word Access Procedures, Nijmegen, The Netherlands: Max Planck Institute for Psycholinguistics.
7. Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2000). Lexical segmentation and ambiguity: Investigating the recognition of onset-embedded words. *Proceedings of SWAP*. Nijmegen, The Netherlands: Max Planck Institute for Psycholinguistics.
8. Dutton, D. (1992). Allophonic variation and segmentation in spoken word recognition. Doctoral Dissertation State University of New York at Buffalo.
9. Gaygen, D., & Luce, P. A. (Submitted). Troughs and Bursts: Probabilistic phonotactics and lexical activation in the segmentation of spoken words in fluent speech.
10. Houston, D., Jusczyk, P. W., & Tager, J. (1998). Talker-specificity and the persistence of infants' word representations. In A. Greenhill, M. Hughes, H. Littlefield, & H. Walsh

(Eds.), *Proceedings of the 22nd Annual Boston University Conference on Language Development*, (Vol. 1, pp. 385–396). Somerville, MA: Cascadilla Press.

11. Jusczyk, P. W., Pisoni, D. B., Fernald, A., Reed, M., & Myers, M. (1983). *Durational context effects in the processing of nonspeech sounds by infants.* Paper presented at the Meeting of the Society for Research in Child Development, Detroit, MI.
12. Kewley-Port, D. (1983). *Time-varying features as correlates of place of articulation in stop consonants* (Research on Speech Perception: Technical Report 3): Indiana University, Bloomington, IN.
13. Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics, *Proceedings of the Sixth International Congress of Phonetic Sciences*. Prague: Academia.
14. Luce, P. A., Charles-Luce, J., & Mclennan, C. (1999). Representational specificity of lexical form in the production and perception of spoken words. *Proceedings of the 1999 International Congress of Phonetic Sciences*, 1889–1892.
15. Nazzi, T., Jusczyk, P. W., & Bhagirath, K. (1999). *Infants' segmentation of verbs from fluent speech.* Paper presented at the Biennial Meeting of the Society for Research in Child Development, Albuquerque, NM.
16. Pitt, M. (1994) Lexical competition: the case of embedded words. Poster presented at the 34th annual meeting of the Psychonomic Society, Washington, D. C.
17. Plunkett, K., Bailey, T., & Bryant, P. E. (2000). *Phonological representation and word recognition.* Paper presented at the International Conference on Infant Studies, Brighton, UK.
18. Port, R. F. (1976). The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words. Unpublished Unpublished doctoral dissertation, University of Connecticut.
19. Sheffert, S. M. (1995). *Implicit and explicit memory for words and voices*. Unpublished doctoral dissertation. University of Connecticut.
20. Soderstrom, M., Jusczyk, P. W., & Kemler Nelson, D. G. (2000). Evidence for the use of phrasal packaging by English-learning 9-month-olds. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development*, (Vol. 2, pp. 708–718). Somerville, MA: Cascadilla Press.
21. Summerfield, Q. (1975). Acoustic and phonetic components of the influence of voice changes and identification times for CVC syllables., The Queen's University of Belfast, Belfast Ireland.
22. Summerfield, Q., & Haggard, M. P. (1973). *Vocal tract normalisation as demonstrated by reaction times* (Report of Speech Research in Progress 2): The Queen's University of Belfast, Belfast, Ireland.
23. Tsushima, T., Takizawa, O., Sasaki, M., Siraki, S., Nishi, K., Kohno, M., Menyuk, P., & Best, C. (1994). Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6–12 months: Language specific developmental changes in speech perception abilities. Paper presented at the International Conference on Spoken Language Processing, Yokohama, Japan.
24. Williams, L. (1977). *The effects of phonetic environment and stress placement on infant discrimination of place of stop consonant articulation.* Paper presented at the Second Annual Boston University Conference on Language Development, Boston, Mass.