The Handbook of Phonetic Sciences

Second Edition

Edited by

William J. Hardcastle, John Laver, and Fiona E. Gibbon

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

14 Cognitive Processes in Speech Perception

JAMES M. MCQUEEN AND
ANNE CUTLER

1 Introduction

The recognition of spoken language involves the extraction of acoustic-phonetic information from the speech signal, and the mapping of this information onto cognitive representations. To develop accurate psycholinguistic models of this process, we need to know what information is extracted from the signal, and when and how it is integrated with stored knowledge.

The central knowledge store for speech perception is the mental lexicon, that is, our stored representations of words. The utterances that we hear may be new to us, but they are made up of known words; by recognizing the words and parsing their sequence we are able to understand what has been said. Word recognition, we argue, is therefore at the heart of speech perception.

The cognitive processes we will discuss, therefore, concern the relationship between lexical processing (the recognition of words) and prelexical processing (getting from the acoustic-phonetic input to the representations of words). Models of spoken-word recognition have been distinguished principally by their characterization of this relationship, in particular their proposals regarding the directionality of flow of information between the prelexical and lexical levels. The role of the lexicon in prelexical speech processing is discussed in section 2.

In every language, the vocabulary contains tens or hundreds of thousands of words. Since all these words are made up from just a few phonemes (across languages, the average phoneme repertoire size is around 30; Maddieson, 1984), it is necessarily the case that words resemble one another. Successful speech recognition thus entails discriminating the phonemic contrasts that distinguish a word from the other words that resemble it – for instance, deciding that we have heard *word*, and not *bird*, *ward*, or *work*. The processing of segmental information in word recognition is discussed in section 3.

One of the most salient and important facts about speech perception is our ability to understand speech from talkers we have never heard before, and to perceive the same phoneme despite acoustically very different realizations (e.g., by a child's

voice versus an adult male's). This ability has been characterized as a process of "normalization," in which an underlying commonality is extracted from the surface variation. The relevant research is discussed in section 3.2 (and see also Johnson, 2005, for a review of models of the normalization process). The role played by abstract underlying representations of the kind assumed in such accounts has, however, been called into question in recent years because of accumulated evidence both that individual speech perception episodes contribute to the knowledge that is stored about speech, and that they can influence later processing (see, e.g., Johnson & Mullennix, 1997). Speech perception theory is currently in an exciting state of transition to a new generation of models in which a role for abstract representations is combined with a role for veridical representations of speech episodes or exemplars. The evidence which shows that both types of representation are involved in speech perception is also summarized in section 3.2.

An account of the cognitive process of speech perception would, finally, be incomplete if it considered only segmental and lexical information, for the suprasegmental dimensions in which the speech signal varies also contribute significantly to listeners' processing decisions. This is documented in section 4.

The recognition of speech is one of humankind's most useful and significant achievements; underlying it is cognitive processing of enormous complexity but also admirable efficiency. Modeling this process has occupied psycholinguists and speech scientists for decades, and, as our concluding summary in section 5 demonstrates, the search for the ultimately accurate model is not over yet.

2 Lexical Information

We examine the role of lexical information in speech processing by contrasting interactive models with autonomous models. Interactive models hold that lexical information influences prelexical processing. We will focus on one particular model of this class, TRACE (McClelland & Elman, 1986; McClelland, 1991; see left panel of Figure 14.1). This interactive-activation model has three levels of processing containing, respectively, featural representations, phonemes, and words. Units within a level compete with each other via lateral inhibitory connections. Units at lower levels activate the units at higher levels with which they are consistent via facilitatory connections. Thus, during word recognition, activation of a feature node leads to activation of consistent phoneme nodes, which in turn activate word nodes. Importantly, higher-level units also facilitate lower-level units. Activated word units boost the activation of their constituent phonemes: this top-down facilitation instantiates the claim that lexical information influences prelexical processing.

We will contrast the TRACE model with an autonomous model which holds that lexical information is not involved in prelexical processing. The Merge model (Norris et al., 2000; see right panel of Figure 14.1) also has three levels of processing: a prelexical level, a lexical level, and a level at which explicit decisions are made about speech sounds. Units within each of the latter two levels inhibit each other. Prelexical units facilitate lexical and decision-level units with which they

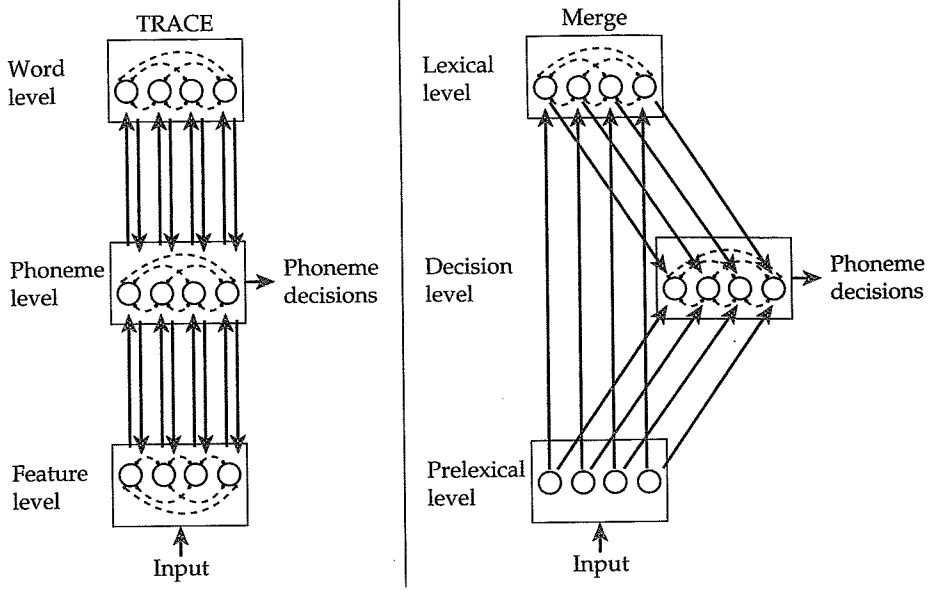

Figure 14.1 Sketches of the processing architecture of, on the left, the interactive TRACE model (McClelland & Elman, 1986), and, on the right, the autonomous Merge model (Norris et al., 2000). Excitatory connections between nodes at different levels are shown with solid lines and arrows; bidirectional inhibitory connections between nodes within levels are shown with dashed lines and closed circles. Not all connections are shown.

are consistent, and lexical units also facilitate decision-level units. There are only these feedforward connections in Merge. Critically, there is no feedback from the lexical to the prelexical level, so the lexicon cannot influence prelexical processing. Merge is linked to the Shortlist model of spoken-word recognition (Norris, 1994; Norris & McQueen, 2008), which is also based on the assumption that there is no lexical feedback. The prelexical and lexical levels in Merge (i.e., those involved in word recognition) are interchangeable with those two levels in Shortlist, while the decision units (and the feedforward connections to them) are not part of the word-recognition process: they are used only when listeners have to make explicit phonetic decisions. Below, we will describe how the TRACE and Merge models, as instances of interactive and autonomous theories, account for lexical involvement in various phonetic tasks.

2.1 *Lexical effects*

2.1.1 Monitoring Phoneme monitoring is sensitive to phonetic factors. Foss and Gernsbacher (1983) found an effect of vowel length: the longer the vowel,

the longer the reaction time (RT) to the preceding target consonant. Another factor is the phonological similarity of the target phonemes to preceding phonemes (Newman & Dell, 1978; Dell & Newman, 1980). Detection of target phonemes in sentences is slower when the word preceding the target-bearing word begins with a phoneme closely related to the target. Several studies, however, have failed to find lexical effects. Foss et al. (1980) found that monitoring was no faster for words than for nonwords, and that the frequency of occurrence of the target-bearing word did not influence RT. Segui et al. (1981) also failed to find an RT advantage for word responses over nonword responses, and Segui and Frauenfelder (1986) found no frequency effect when subjects were required to monitor only for word-initial phonemes ("standard" phoneme monitoring).

These results support the claim that phoneme monitoring is based on prelexical processing which is open to the influence of phonetic information but not lexical information. But there are some studies which have demonstrated lexical effects. Segui and Frauenfelder (1986), for instance, did obtain a word-frequency effect when subjects were required to monitor not just for word-initial targets, but for targets which could appear anywhere in the words ("generalized" phoneme monitoring). Rubin et al. (1976) also found a word/nonword effect: subjects were faster to detect e.g. /b/ in *bat* than in *bal*. Cutler et al. (1987) examined word/nonword effects in a series of experiments. Lexical effects were found to come and go. Responses to targets in words were faster than those to targets in nonwords only when task monotony was reduced.

Lexical effects thus appear to be present only in some phoneme-monitoring experiments. Stemberger et al. (1985) took this variability as support for interactive models like TRACE. Lexical influences were taken to result from top-down facilitation from word nodes increasing the level of activation of target phoneme nodes, thus speeding responses to targets in words relative to nonwords. Where there were no lexical effects, it was assumed that responses were being made from the phoneme-node level, with lexical feedback switched off through some kind of attentional process. But the presence of lexical effects, and their variability, can equally well be explained by autonomous models (Cutler et al., 1987; Norris et al., 2000). In Merge, lexical effects in phoneme monitoring are due to the feedforward influence of the lexical level on decision nodes, and their absence, just as in the TRACE account, is assumed to reflect the fact that the lexical influence, due for example to attentional factors, has been switched off.

Both models can therefore account for the lexical effect, and its variability, in phoneme monitoring. In another task, rhyme monitoring, where subjects detect words and nonwords which rhyme with a prespecified cue, responses are faster to words than to nonwords, and responses are faster to high- than to low-frequency rhyming words (McQueen, 1993). Again both models can explain these lexical effects.

2.1.2 Phonemic restoration If the medial /s/ of the word *legislatures* is replaced with a cough, listeners report hearing a cough and the complete *legislatures*, with the absent phoneme perceptually restored (Warren, 1970). Low-level factors influence the effect: if the replacing noise is acoustically similar to the removed phoneme, the illusion is more likely to occur (Warren & Obusek, 1971; Samuel, 1981a, 1981b); and there is more restoration for fricatives and stops (which are more noise-like) than for liquids, vowels and nasals (Samuel, 1981a, 1981b).

Samuel (1981a) found that several lexical factors influenced the extent of the illusion: there was more restoration for longer than for shorter words; there was a more reliable illusion in words than in phonologically legal nonwords; and presenting an intact version of the target word before the target word also increased restoration. Samuel (1987) found further that there was more restoration for items with several possible restorations (e.g., **egion*: *legion* or *region*) than for items with a unique restoration (e.g., **esion*: *lesion*). He also found that there was more phonemic restoration in words which become unique early, moving left to right through the word (e.g., *boysenb*rry*) than in words which became unique late (e.g., *indel*ble*). Samuel (1996) showed that, as with lexical involvement in phoneme monitoring, lexical effects in phonemic restoration are variable; to use his words, they are "real but fragile." Samuel explained these results in terms of lexical feedback.

An autonomous account of these data, however, is once again also possible. If the illusion is due to attention being focused on lexical information, then the lexical effects can be explained without recourse to top-down connections. In Merge's terms, lexical influences in restoration occur when listeners are using the connections from the lexical level to the decision level. Just as with the monitoring tasks, the evidence for lexical involvement in phoneme restoration reviewed so far does not allow us to distinguish between the two models.

2.1.3 Phonetic categorization In the phonetic categorization task, with a continuum of sounds from /d/ to /t/ in the contexts *deep–teep* and *deach–teach*, for example, a lexical effect would be shown by an increased proportion of /d/ responses in the ambiguous region of the continuum when the voiced endpoint formed a word (*deep*), and an increased proportion of /t/ responses when the unvoiced endpoint formed a word (*teach*). This effect was originally demonstrated by Ganong (1980), and replicated by Fox (1984). In TRACE, this effect is once again accounted for by top-down connections. In Merge, the effect once again reflects the integration of prelexical and lexical information at the decision level.

Connine and Clifton (1987) found both a lexical shift and an RT advantage for word responses relative to nonword responses in the boundary region. They further showed that the lexical effect was not due to postperceptual bias: it was not equivalent to an effect obtained using monetary reward to bias subjects' responses. Lexical effects have also been reported by Burton et al. (1989), who found that the categorization of a word-initial continuum depended on the acoustic-phonetic quality of the continuum, and by Miller and Dexter (1988), who showed that lexical involvement in categorization (as in phoneme monitoring and the phonemic restoration illusion) is not mandatory.

McQueen (1991) and Pitt and Samuel (1993) have found lexical effects for phonemes in word-final position (e.g., for an /ʃ/–/s/ continuum in contexts such

as *fish–fiss* and *kish–kiss*). McQueen (1991) also replicated Burton et al.'s (1989) finding that lexical effects in the categorization task only appear when the materials are of poor acoustic quality. Pitt and Samuel (1993), however, have shown that poor stimulus quality is not a necessary condition for a lexical effect: lexical shifts were obtained with high-quality materials in both word-initial and word-final categorization. Critically, however, the basic lexical effect in this task is consistent with both types of model.

2.2 *Test cases*

Both models can account for lexical effects in several tasks. Are there any test cases which might allow us to distinguish between the models? Can we establish whether or not lexical information is used in prelexical processing? Several attempts have been made to contrast divergent predictions of the TRACE and Merge models.

2.2.1 Inhibitory lexical effects in phoneme monitoring Frauenfelder et al. (1990) presented evidence from the phoneme-monitoring task which challenges the interactive position. TRACE predicts that activation of a lexical candidate will both boost the activation of its constituent phonemes by top-down facilitation and inhibit the activation of nonconstituent phonemes because of phoneme-to-phoneme inhibition. As this study showed, there are strong facilitatory effects on the detection of targets (such as /p/ in *olympiade*), which occur after the word becomes unique, relative to matched nonwords (e.g., *arimpiako*). In TRACE terms, this could be due to top-down facilitation of /p/ from the word node. If this were the case, detection of /t/ in *vocabutaire* should be inhibited relative to detection of /t/ in a matched nonword such as *socabutaire*, because of top-down facilitation of /l/ from the activated *vocabulaire* node followed by inhibition of other phoneme nodes by the /l/ node. No such inhibition was found.

Mirman et al. (2005) have recently shown, however, that lexically induced delays in phoneme monitoring do occur, but only if the target phoneme and the lexically consistent phoneme are phonetically similar (e.g., /t/ detection in *arsenit* was delayed because /t/ is similar to the lexically consistent /k/ in the base word *arsenic*, but /t/ detection in *abolit* was not delayed, presumably due to greater dissimilarity between /t/ and the /ʃ/ of *abolish*). Mirman et al. present TRACE simulations showing that this kind of lexical inhibition only arises in the model if there is some bottom-up support for the lexically consistent sound (i.e., when it is acoustically similar to the target). As Norris et al. (2000) argue, Merge predicts that there should be lexical inhibition in phoneme monitoring when there is perceptual support for two competing phonemes. Thus, while the absence of lexical inhibition in phoneme monitoring was for many years a problem for the interactive account, more recent research has shown that this is not in fact a test case that distinguishes between interactive and autonomous models.

2.2.2 The time course of lexical effects in categorization McQueen (1991) showed that the lexical effect in categorization of word-final ambiguous fricatives, in contexts such as *fish–fiss* and *kish–kiss*, was larger for faster responses. This finding was replicated by Pitt and Samuel (1993) and by McQueen et al. (2003). Previous research (Fox, 1984; Miller & Dexter, 1988; Pitt & Samuel, 1993) had shown that lexical effects in categorization of ambiguous word-initial sounds (e.g., /d/ and /t/ in *deep–teep* vs. *deach–teach*) build up over time, but McQueen et al. (2003) also showed that, at least under some circumstances, lexical effects in word-initial categorization can also decrease over time. In TRACE, the lexical feedback loop becomes stronger over time, and bottom-up evidence thus tends to be overwritten by lexical knowledge. TRACE thus wrongly predicts that lexical effects should build up gradually over time, for word-initial (McClelland & Elman, 1986) and word-final categorization (McClelland, 1987). In Merge, in contrast, there is no feedback loop, and lexical knowledge does not overwrite bottom-up evidence. So lexical effects can reach an asymptote, and indeed die away over time (if lexical input to the decision nodes is switched off, the bottom-up evidence, as represented at the prelexical level, can re-assert itself at the decision level). These time-course analyses thus favor autonomous accounts.

2.2.3 Compensation for coarticulation One type of result appears to support interactive models. Mann and Repp (1981) showed that stops midway between /t/ and /k/ were more often categorized as /k/ after /s/, but as /t/ after /ʃ/. The perceptual system thus appears to compensate for fricative–stop coarticulation. Elman and McClelland (1988) replicated this effect for ambiguous word-initial stops following fricative-final words such as *Christmas* and *foolish*, and, most importantly, they showed that the effect occurred when the word-final fricatives were replaced with an ambiguous fricative. When the final /s/ in *Christmas* was replaced with an ambiguous sound /?/, midway between /s/ and /ʃ/, there were again more /k/ responses to the ambiguous stops. With *fooli?*, there were more /t/ responses.

Elman and McClelland (1988) claimed that this effect was strong evidence in favor of interactive models like TRACE. Lexical information appears to be influencing a compensation process that can be assumed to be operating prelexically. This seems to be direct evidence against the autonomous assumption that there is no lexical feedback. But Pitt and McQueen (1998) argued that there was an alternative account of these data. In English, /s/ is more likely than /ʃ/ after schwa (as in *Christmas*), and /ʃ/ is more likely than /s/ after /ɪ/ (as in *foolish*). Pitt and McQueen then showed, first, that if vowel-fricative transitional probabilities were controlled, there was no lexical effect with ambiguous fricatives in stop categorization, and second, that if vowel–fricative transitional probabilities were manipulated in nonwords, those probability biases on ambiguous fricative identification did lead to a consequent shift in stop categorization. Pitt and McQueen thus argued that if the prelexical level were sensitive to transitional probabilities, the Elman and McClelland results would be consistent with an autonomous model. Studies in the next round of this debate have claimed to show that lexical influences in fricative–stop compensation for coarticulation can be found when vowel–fricative transitional probabilities are controlled (Magnuson et al., 2003; Samuel & Pitt, 2003). The Magnuson et al. result, however, appears to be due to a bias induced

by their practice trials (McQueen et al., 2009), and Samuel and Pitt's findings may reflect longer-range transitional probabilities than those concerning the vowel–fricative sequence alone (McQueen, 2003). This issue is not yet resolved (McClelland et al., 2006; McQueen, Norris, et al., 2006), and appears to depend on increasingly subtle experimental manipulations. There has to date been no completely convincing demonstration of lexical involvement in compensation for coarticulation.

Another aspect of Pitt and McQueen's (1998) data, replicated by McQueen et al. (2009), is problematic for interactive models. Listeners were asked to identify the fricatives (as /s/ or /ʃ/) as well as the stops (as /t/ or /k/). A lexical effect was found in the fricative judgments (just as in the Ganong, 1980, and McQueen, 1991, studies, listeners judged more ambiguous sounds to be lexically consistent than to be lexically inconsistent). Yet in the very same trials there was no lexical effect in the stop judgments. If the lexical effect on the fricatives were due to feedback, as in the TRACE account, then that feedback ought to have had an effect on the prelexical compensation for coarticulation mechanism, and there thus ought also to have been a lexical effect on the stops. This dissociation challenges TRACE. It is consistent with Merge, however, since lexical effects on fricative decisions reflect the influence of the lexicon on the decision level and thus not on the prelexical level where the compensation mechanism is assumed to operate.

2.2.4 Selective adaptation Samuel (1997, 2001) used a logic similar to Elman and McClelland (1988), and again tested for lexical influences on a prelexical process (selective adaptation rather than compensation for coarticulation). In selective adaptation, judgments about speech sounds change through repeated exposure to one sound (e.g., after hearing /da/ repeatedly, listeners report more stimuli on a /da–ta/ continuum to be /ta/; Eimas & Corbit, 1973). The locus of this adaptation effect appears to be prelexical (Samuel & Kat, 1996). Samuel (1997) used sounds replaced with noise as adaptors (capitalizing on the phoneme restoration illusion), and Samuel (2001) used ambiguous sounds as adaptors (capitalizing on the Ganong, 1980, effect). In both cases these ambiguous adaptors appeared in lexical contexts. Adaptation effects were found which were similar to those that would be observed with unambiguous lexically consistent sounds. In line with interactive models such as TRACE, lexical knowledge thus appeared to be influencing the prelexical adaptation process. As we discuss below, however, these results appear to be consistent with the autonomous view that the lexicon does not influence on-line prelexical processing.

2.3 *On-line feedback versus feedback for learning*

As our review of these test cases reveals, there is no clear winning theory. Some experiments favor the interactive view embodied in TRACE, some favor the autonomous view in Merge, and some test cases have proven not to distinguish between the models. Norris et al. (2000) argued, however, that even if the data are not definitive, there are important theoretical arguments to consider. They pointed out that feedback, as instantiated in TRACE, cannot be of any benefit to

word recognition, and can be harmful to phoneme recognition. The best a word-recognition system can do is recognize the words that are most consistent with the input. Thus, if processing is optimal, feedback cannot improve on the decisions made at the lexical level (all it does is copy the decisions made at the lexical level onto the prelexical level). Feedback can help with phoneme recognition (e.g., when a sound is ambiguous), but can potentially create phonemic hallucinations, where lexical knowledge overwrites perceptual evidence. Norris et al. thus argued that, in the absence of clear experimental data in support of interactive models, autonomous models such as Merge should be preferred simply because there is no good reason to postulate feedback connections.

There is one critical exception to this argument. Feedback for perceptual learning would be of benefit in speech perception. If listeners could adjust their prelexical representations over time, using lexical knowledge, then they could learn how to interpret a talker speaking in an unusual way (e.g., a talker with a speech impediment, or someone with a regional or foreign accent). If, for example, a talker with a lisp produces an ambiguous /s/ sound in words where an /s/ (and not an /f/) is expected (e.g., at the end of *platypu-*), then listeners could use this lexical knowledge to adjust their category boundary between /f/ and /s/, facilitating subsequent recognition of speech by that talker. Norris et al. (2003) found support for this prediction in a Dutch perceptual-learning experiment. After exposure to only 20 /s/-final words ending with an ambiguous /fs/ sound (e.g., *radij?*, based on *radijs*, 'radish'), listeners' category boundaries shifted to include more /f/-like sounds in the /s/ category. Another group of listeners heard the same sound in 20 /f/-final lexical contexts (e.g., *olij?*, based on *olijf*, 'olive'), and learned to include more sounds in their /f/ category. Control conditions showed that lexical knowledge was required for this kind of perceptual learning. Subsequent research with this lexical retuning paradigm has shown that the learning can be, but need not be, talker-specific (Eisner & McQueen, 2005; Kraljic & Samuel, 2005, 2006, 2007), and that it is stable over at least 12 hours (Eisner & McQueen, 2006).

Norris et al. (2003) argued, therefore, that there is lexical feedback in perceptual learning. The prelexical level appears to be flexible enough to be able to adjust its operation over time, using stored lexical knowledge, as it encounters different talkers. This kind of feedback is logically distinct from the kind of feedback instantiated in TRACE, where lexical knowledge modulates, on-line, the perceptual process. Thus, while the feedback mechanism in TRACE may offer an explanation for both on-line and learning effects (McClelland et al., 2006), there is no necessity that both types of effect be explained by the same mechanism. A perceptual learning mechanism using lexical feedback could therefore be added to Merge, without there being any effects of that feedback loop on on-line perception (Norris et al., 2003). Furthermore, this kind of model could explain the Samuel (1997, 2001) results. The selective adaptation paradigm requires repeated exposure to stimuli, and, as Norris et al. argued, may thus involve the same kind of perceptual retuning as they observed (see also McQueen, Norris, et al., 2006; Vroomen et al., 2007).

In our chapter in the first edition of this book we concluded that the lexicon does not influence prelexical processing. This conclusion has stood the test of

time, at least with respect to on-line processing. It now appears, however, that there is lexical involvement in prelexical perceptual learning. Perhaps most importantly, while on-line feedback is of no benefit to word recognition, retuning of perception over time does benefit speech perception. There is thus only feedback of the type that helps listeners.

3 Segmental Information

We now consider two key questions about the role of segmental information in the cognitive processes underlying speech perception. First, we ask whether, during word recognition, segmental information is processed serially, or in a cascaded fashion. Our focus is on word recognition, since, as we have already argued, it is only through recognizing words that the listener can understand spoken messages. It is entirely uncontroversial that segmental information must play a central role in spoken-word recognition. The primary way in which listeners determine whether they have heard *word* or *bird* is through identifying that the first sound is a /w/ and not a /b/. It is also uncontroversial that a wide variety of acoustic cues are used in segment perception (see Raphael, 2005). The question we must ask, then, is how these cues modulate word recognition. One possibility is that they do so in a serial manner: First, segments are identified on the basis of these cues (at the prelexical stage of processing), and then, second, words are recognized. Alternatively, acoustic-phonetic information may flow in cascade through the recognition system, such that it influences lexical-level processes without any prior definitive categorization of the input into segmental categories. We consider these two alternatives in section 3.1.

But there is an even more fundamental question about the uptake of segmental information. So far, we have assumed that there is a prelexical stage of processing, where the speech input is recoded (in either a serial or cascaded fashion) into linguistically abstract segmental categories prior to and for lexical access. But there may be no such prelexical abstraction process. We consider arguments for and against abstraction in section 3.2.

3.1 *Cascaded processing of segmental information*

Multiple lexical hypotheses are activated during the word-recognition process (Swinney, 1981; Marslen-Wilson, 1987, 1990; Zwitserlood, 1989; Shillcock, 1990; Gow & Gordon, 1995; Tabossi et al., 1995; Vroomen & de Gelder, 1997; Allopenna et al., 1998). We can thus ask whether segmental information is passed serially or in cascade to the lexical level, by measuring whether lexical activation changes as a function of subsegmental differences in the input. According to a serial model, subsegmental ambiguities should, if possible, be resolved prelexically and thus should not affect lexical activation levels. But in a cascaded model subsegmental differences should be passed on to the lexical level and influence the degree of activation of different lexical hypotheses.

Andruski et al. (1994) reduced the Voice Onset Time (VOT) of the initial stop of, for example, *king*. In English, VOT is a major acoustic-phonetic cue to the distinction between voiceless stops (e.g., /k/, with long VOTs) and voiced stops (e.g., /g/, with short VOTs). In a cross-modal priming task, responses to semantically related targets (e.g., *queen*) were faster after *king* than after an unrelated word. This priming effect became smaller as VOT was reduced (i.e., as the /k/ became more like a /g/). This suggests that subsegmental detail influences lexical activation (the degree of activation of *king* was reduced as the /k/ was shortened), in keeping with the cascaded model. Further evidence is provided by McMurray et al. (2002) and Utman et al. (2000).

Other evidence in favor of cascaded models comes from research examining the effects of mispronunciations. The phonetic similarity between an intended word (e.g., *cabinet*; Connine et al., 1997) and a mispronounced nonword (e.g., *gabinet* vs. *mabinet* vs. *shuffinet*) influences how disruptive the mispronunciation is to lexical access. The greater the similarity between the mismatching sound and the intended sound, the more strongly the intended word appears to be activated (see also Connine et al., 1993; Marslen-Wilson et al., 1996; Ernestus & Mak, 2004). In the domain of research on continuous speech processes (such as place assimilation in English, Gow, 2002; liaison in French, Spinelli et al., 2003; and /t/ reduction in Dutch, Mitterer & Ernestus, 2006) it appears that fine-grained phonetic detail (e.g., the duration or spectral structure of segments) modulates the degree of lexical activation, again as predicted by the cascaded account.

A considerable body of evidence suggests further that, once words consistent with the input speech have been activated, they compete with each other (Goldinger et al., 1989, 1992; Cluff & Luce, 1990; Slowiaczek & Hamburger, 1992; McQueen et al., 1994; Norris et al., 1995; Vroomen & de Gelder, 1995; Vitevitch & Luce, 1998, 1999; Luce & Large, 2001; Gaskell & Marslen-Wilson, 2002). Accessed words compete with each other until one word dominates the others; this one word can then be recognized. This competition process is instantiated, in different ways, in current models of spoken-word recognition: the Neighborhood Activation Model (Luce & Pisoni, 1998), TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994; Norris & McQueen, 2008) and the Distributed Cohort Model (DCM; Gaskell & Marslen-Wilson, 1997).

The lexical competition process provides a litmus test about how segmental information flows through the speech recognition system. If fine-grained phonetic detail modulates the competition process, then it must have been passed forward to the lexical level. Several of the effects just reviewed have been shown to interact with lexical factors. Van Alphen and McQueen (2006) have shown, for example, that the influence of VOT variability on lexical activation in Dutch depends on the lexical competitor environment (i.e., whether the voiced and voiceless interpretations of the stops are both words, as in, e.g., the English pair *bear–pear*, or both nonwords, as in English *blem–plem*, or one word and one nonword, as in *blue–plue* and *brince–prince*), and Marslen-Wilson et al. (1996) also found that segmental mismatch effects were modulated by lexical factors.

Clear demonstrations of this interaction between subsegmental and lexical information come from a series of experiments with cross-spliced stimuli (Streeter & Nigro, 1979; Whalen, 1984, 1991; Marslen-Wilson & Warren, 1994; McQueen et al., 1999; Dahan et al., 2001). Cross-splicing the initial consonant and vowel of *jog* with the final consonant of *job*, for example, produces a stimulus which sounds like *job*, but which contains (in the vocalic portion) acoustic evidence for a /g/. These phonetically mismatching stimuli are more difficult to process, but the extent to which they interfere depends on whether the entire sequence is a word or nonword (e.g., *job* vs. *shob*) and on whether its components derive from words (e.g., *jog*) or nonwords (e.g., *jod*). It thus seems very clear that there is cascade of segmental information to the lexical level.

3.2 *Segmental abstraction in speech perception*

There are two ways in which segmental information might cascade to the lexicon. One possibility is that phonetic segments are extracted explicitly, in some prelexical level of representation, with a classification of the speech signal into linguistically abstract "units of perception" (McNeill & Lindig, 1973; Healy & Cutting, 1976). Units which have been postulated include acoustic-phonetic features (Eimas & Corbit, 1973; Marslen-Wilson, 1987; Marslen-Wilson & Warren, 1994; Stevens, 2002), phonemes (Foss & Blank, 1980; McClelland & Elman, 1986; Norris, 1994; Norris et al., 2000), context-sensitive allophones (Wickelgren, 1969), syllables (Cole & Scott, 1974; Mehler, 1981), and articulatory gestures (Liberman & Mattingly, 1985). Any of these units could operate in cascade, passing information continuously forward to the lexicon. Alternatively, segmental information could be used implicitly, in a direct and continuous mapping of the signal onto the lexicon with no explicit intermediate classification into prelexical units. Klatt (1979, 1989) suggested a template-matching process, where spectral information, as analyzed by the peripheral auditory system, is mapped directly onto a lexicon of spectral templates of diphone sequences. The models of Goldinger (1996, 1998), Johnson (1997), Pierrehumbert (2002), and Hawkins (2003), though differing in many respects, share the assumption that the lexicon consists of episodic memory traces of particular tokens of words, stored with all their acoustic detail (e.g., including talker- and situation-specific details).

The class of models with abstract prelexical representations provide a ready solution to the invariance problem. It is well known that the acoustic cues to segments are far from invariant. They vary greatly depending on a large number of factors, including: coarticulation (the realization of segments depends upon both preceding and following phonological context; Fowler, 1984; see Farnetani & Recasens, this volume); speech rate (e.g., temporal cues such as VOT change depending on speed of articulation, requiring rate-dependent processing; Miller, 1981; Gordon, 1988); and variation between speakers due to differences in sex, age, and dialect (see Ní Chasaide & Gobl, this volume). Some authors have argued that this variation is dealt with by the extraction of acoustic cues which are invariant (Stevens & Blumstein, 1981); others that the variation is lawful, and can be exploited by the listener (Elman & McClelland, 1986). In either case, however, it is clear that the perceptual system must be able to deal with this variability through some kind of normalization process. The same physical signal must be interpretable as different segments, and different signals must be interpretable as the same segment (Repp & Liberman, 1987). If normalization takes place prelexically, prior to lexical access, then only abstract phonological knowledge would need to be stored in the mental lexicon.

In a study of speech rate normalization, for example, Miller et al. (1984) demonstrated that subjects labeled more ambiguous consonants, midway between /b/ and /p/ and embedded in the continuum *bath–path*, in a contextually congruent manner (i.e., more *bath* responses in a bathing context), but only when subjects were explicitly told to attend to the sentence context. These effects were absent in a speeded response condition, which focused the subjects' attention on the target words. Speaking rate was also varied, resulting in shifts in the category boundary between /b/ and /p/, but the task-demand manipulation did not influence this rate-dependent boundary placement. Miller and Dexter (1988) also used the phonetic categorization task to examine effects of lexical status and speaking rate. They found that under speeded response conditions, there was no tendency to label ambiguous initial consonants in a lexically consistent manner (e.g., as /b/ in a *beef–peef* continuum and as /p/ in a *beece–peace* continuum). Listeners could not ignore the rate manipulation, however: even under speeded-response instructions they based their decision on the early portion of the syllable, treating it as if it was physically short (the /b/–/p/ boundary shifted to a smaller VOT for fast responses). These studies neatly demonstrate that rate normalization (unlike the use of lexical knowledge) is a mandatory feature of speech processing. The analysis of acoustic information specifying speech rate appears to be essential for accurate lexical access (Miller, 1987). Such results thus support the view that rate normalization is a prelexical process that necessarily modulates word recognition.

Other, more recent evidence on the need for prelexical abstraction comes from the lexical retuning paradigm reviewed earlier. Norris et al. (2003) argued that adjustments to prelexical phonetic categories would be of benefit to speech perception because, once an adjustment had been made, it could be applied to the recognition of any words containing the adjusted sound. McQueen, Cutler, et al. (2006) tested whether there was indeed generalization of learning to the processing of words that were spoken by the trained talker but that had not been heard during the training phase. Instead of using the phonetic categorization test task (as in, e.g., the Norris et al. study described in section 2.3), they used a cross-modal identity-priming task. The experiment was again in Dutch, and the test phase contained minimal pairs such as *doof–doos*, 'deaf–box'. In a training phase identical to the Norris et al. study (i.e., with no minimal-pair words), listeners were encouraged to learn that an ambiguous /fs/ sound, "[?]", was either /f/ or (for a second group) /s/. In the subsequent test phase, the pattern of priming effects indicated that the listeners in the first group tended to hear [do?] as *doof*, while listeners in the second group tended to hear it as *doos*. This demonstration

of generalization of learning to previously unheard words confirms that the locus of the learning effect is prelexical. It also underlines the major benefit of prelexical abstraction: Once something about how a talker realizes a speech segment has been learned, and that knowledge is stored prelexically, then it can automatically be used to assist in the recognition of all words containing that segment that that talker might produce. Further evidence of lexical generalization of perceptual learning has come from experiments examining adjustments to vocoded speech (Davis et al., 2005) and to an artificial dialect (Maye et al., 2008).

Findings from subliminal priming studies (Kouider & Dupoux, 2005), identification tasks (Nearey, 1990), phonological priming (Radeau et al., 1995; Slowiaczek et al., 2000), and studies on phoneme-sequence learning (Onishi et al. 2002) also support prelexical abstraction. Yet more evidence comes from the second-language literature. Listeners have considerable difficulty learning new phonemic categories (Logan et al., 1991; Strange, 1995), and are influenced by the phonemic categories of their native language while listening to a nonnative language (Best, 1994; Pallier et al., 2001; Weber and Cutler, 2004; Cutler et al., 2006). These findings suggest that, once abstract prelexical categories have been acquired, they necessarily influence speech recognition, even in a second language.

There is therefore considerable evidence for prelexical abstraction. According to an extreme version of the abstractionist position, details about speaking rate, talker, etc., could be discarded prior to lexical access. But there is also considerable evidence that episodic details of words (e.g., how individual talkers produced specific words) are preserved in long-term memory, as measured, for example, in recognition memory experiments (Martin et al., 1989; Goldinger et al., 1991; Palmeri et al., 1993; Church & Schacter, 1994; Goldinger, 1996; Luce & Lyons, 1998): Words are recognized better as having already occurred in the experiment if they are repeated by the same talker. There are also effects of talker-specific detail when participants have to repeat words that they hear (Goldinger, 1998): Repetitions tend to become more like the way the input talker produces them. All of these results show that talker-specific detail cannot be thrown away during word recognition, and thus that extreme abstractionist models are not tenable. But extreme episodic models – in which all acoustic-phonetic detail, including talker-specific attributes, is passed on to the lexicon without normalization – are equally untenable. Such models have the disadvantage that they (unlike abstractionist models) have no ready solution to the invariance problem, and they cannot account for the experimental evidence on abstraction.

What appears to be required, therefore, is a hybrid model in which there is prelexical abstraction, but in which episodic details are not thrown away. On this view, talker-specific features in the speech signal (and other situational details) may be stored in nonlinguistic long-term memory (i.e., not in the mental lexicon), just like other episodic memories (e.g., for faces, colors, or odors), but may also be used in the word-recognition process. That is, just as there appears to be prelexical normalization for speaking rate, there may also be prelexical talker normalization. Indeed, there is evidence that prelexical processing involves adjustments based on talker variability. As already noted, for example, perceptual

learning about unusual segments can be talker-specific (Eisner & McQueen, 2005; Kraljic & Samuel, 2005, 2007). Mullennix et al. (1989) showed that listeners could identify words more easily in lists spoken by a single talker than when the same word-lists were spoken by 15 different talkers, and that this effect was more marked when the speech signal was physically degraded. Nygaard et al. (1994; see also Nygaard & Pisoni, 1998) found, in addition, that familiarity with voices (after extensive training in associating those novel voices with names) made it easier to recognize new words spoken by those voices. Adjustment to talker differences thus appears to occur at a prelexical level, just like rate normalization.

Mullennix and Pisoni (1990) have further shown that talker normalization, again like rate normalization, is mandatory. Subjects could not ignore voice variability when categorizing unambiguous initial phonemes in lists of words spoken by one or several talkers, nor could they ignore variability in those initial consonants when categorizing the words as being spoken by either a male or female speaker. Asymmetries in this interference suggested that extraction of phonetic and speaker information are independent but closely related processes: phonetic decisions appear to be at least partially contingent upon the process of talker normalization, and vice versa.

Segmental information is thus extracted prelexically, recoded into linguistically abstract representations, and used in word recognition. The evidence summarized in section 3.1 suggests that this process operates in cascade. The evidence just reviewed suggests further that this abstraction process is not destructive: acoustic-phonetic details about rate, voice, and so on are used by the normalization process, but are then stored rather than discarded.

4 Suprasegmental Information

As Lehiste (1970) pointed out, it is difficult to carve out a domain of speech research dealing just with suprasegmentals. All speech is realized in time with every subcomponent having a measurable duration, fundamental frequency (f_0) and amplitude. Segment identification thus depends on computations involving f_0, duration (for instance, of vocal tract closure), or amplitude (for instance, of frication in a given frequency range; see Johnson, 2005; Raphael, 2005, for reviews). In quantity languages, there are contrasts between long and short versions of the same segment (Estonian vowels have three levels of duration, for example). Nevertheless, the durational, amplitude, and f_0 patterns of speech also encode structural information at higher levels, and listeners exploit this information in the process of recognizing words (section 4.1), parsing prosodic structure (section 4.2), and segmenting the continuous speech stream (section 4.3).

4.1 *Suprasegmental cues to lexical identity*

Just as durational contrasts at the segment level allow listeners to distinguish between words, so do durational contrasts at the syllable level; for instance, the

greater length of a syllable in isolation than in a polysyllabic sequence (*speed* by itself is longer than *speed* in *speedy* or *speediness*; Lehiste, 1971) allows listeners to know whether they are hearing *ham* or *hamster*, *dock* or *doctor* (Davis et al., 2002; Salverda et al., 2003). This usefully allows listeners to avoid accidentally recognizing words which are only spuriously present in the speech signal, embedded in longer words (such as *ham* in *hamster*). However, there are also ways in which word pairs which are segmentally identical may contrast suprasegmentally.

In languages with lexical tone, such as Mandarin, Vietnamese, or Yoruba, syllables are realized with a f_0 pattern which is phonemically contrastive. Word recognition in such languages depends on processing this f_0 information; to all intents and purposes the tones function at the segmental level, so that a given vowel with a rise–fall tone and the same vowel with a level tone may be regarded as different segments. The available experimental evidence on spoken-word recognition in tone languages indeed supports a parallelism between segmental processing and the processing of tonal information.

For instance, lexical information can affect tone categorization in just the same way as it affects segment categorization. In the segment categorization study of Ganong (1980), listeners' category boundaries between /t/ and /d/ shifted to produce more /d/ responses preceding *-eep* but more /t/ responses preceding *-each*; similarly, in a tone categorization study by Fox and Unkefer (1985), listeners' category boundaries between two tones of Mandarin Chinese shifted as a function of which endpoint tone produced a real word given the syllable the tone was produced on. (This was of course only true when the listeners were Mandarin speakers; English listeners showed no such shift.) Lexical priming studies in Cantonese also suggest that the role of a syllable's tone in word recognition is analogous to the role of the vowel (Yip, 2001; Lee, 2007). The f_0 cues to tone are realized over vocalic segments, but the consequence of this is that vowel identity is apprehended more rapidly than the tone information encoded in the same portion of the speech signal (Cutler & Chen, 1997; Ye & Connine, 1999); thus listeners can detect the difference between two CV syllables with the same onset and the same tone but a different vowel more rapidly than they can detect the difference between two CV syllables with the same onset and the same vowel but a different tone. Suprasegmental contrasts between lexical items in other languages pattern across syllable sequences. In pitch accent languages such as Japanese, words exhibit one of a small number of permissible patterns across syllables (thus the accent pattern of *Toyota* is HLL and of *Mitsubishi* LHLL; in each case the syllable labeled H is accented). Although pitch accent patterns are defined across words, however, their realization in the f_0 contour is easily apprehensible for listeners. Listeners can tell from which of two words differing in accentual structure a given syllable has been extracted (e.g, *ka* from *baka* HL vs. *gaka* LH; Cutler & Otake, 1999), incorrect accent patterns delay word identification (Minematsu & Hirose, 1995), and accent patterns are used to distinguish between competing word candidates in spoken-word recognition, so that, for example, listeners hearing *na-* from *nagasa* HLL know that it cannot be the beginning of *nagashi* LHH (Cutler & Otake, 1999; Sekiguchi & Nakajima, 1999).

Like pitch accent, word stress patterns in lexical stress languages are realized across polysyllabic sequences. There is now an extensive literature on the realization and perception of lexical stress, which has recently been analyzed in detail by Cutler (2005). Listeners can use lexical stress patterns in word recognition, too, so that Dutch listeners can accurately tell from which of two words differing in stress a given syllable has been extracted (e.g, *voor* from initially stressed *voornaam* 'first name' vs. finally stressed *voornaam* 'respectable'; Cutler & Donselaar, 2001), incorrect stress in Dutch words inhibits word recognition (van Leyden & van Heuven, 1996), and Dutch listeners can use stress to distinguish between competing word candidates in spoken-word recognition, so that hearing *domi-* from initially stressed *dominee* 'minister' is evidence that it cannot be the beginning of finally stressed *dominant* 'dominant' (Donselaar et al., 2005). The same results are observed in Spanish: *espi-* from finally stressed *espiral* 'spiral' is perceived as evidence against *espiritu* 'spirit' with stress on the second syllable (Soto-Faraco et al., 2001).

However, the same experiment in English, testing for example *admi-* from initially stressed *admiral* versus *admiration* with stress on the third syllable, produces a much weaker effect (Cooper et al., 2002). There is also considerable evidence that mis-stressing effects in English are quite weak unless vowel quality is also changed (Bond & Small, 1983; Cutler & Clifton, 1984; Small et al., 1988; Slowiaczek, 1990), and cross-splicing English vowels with different stress patterns likewise produces unacceptable results only if vowel quality is changed (Fear et al., 1995). In Fear et al.'s study, listeners heard tokens of, say, *autumn*, which has primary stress on the initial vowel, and *audition*, which has an unstressed but unreduced vowel, with the initial vowels exchanged; they rated these tokens as insignificantly different from the original, unspliced, tokens.

The difference in the patterns of results across lexical stress languages can be ascribed to the relative usefulness of stress in distinguishing between words. It is true that studies of English vocabulary structure show that stress pattern information could be of use in word recognition; thus a partial phonetic transcription which includes stress pattern information applies to a smaller candidate set of words than one which does not (Aull, 1984; Waibel, 1988), and an automatic recognition algorithm operating at this level of phonetic specification performs significantly better with stress pattern information than without (Port et al., 1988). But the equivalent effects in Spanish and Dutch are very much larger (Cutler & Pasveer, 2006). This is because of the widespread reduction of vowels in unstressed syllables in English. Cognate examples of the cross-language asymmetry abound. For example, stressed word-initial *com-* in English could be the beginning of *comedy*, or of *comma*, *compliment*, etc., but *comedy*'s morphological relative *comedian* has the reduced vowel schwa in its initial syllable. In Spanish *comedia* 'comedy' (stress on the second syllable) and *comico* 'comedian' (initial stress) have the same vowel in the first syllables; in Dutch, *komedie* 'comedy' (stress on the second syllable) and *komediant* 'comedian' (stress on the fourth syllable) have the same vowels in their first and in their second syllables. English listeners can instantly distinguish between *comedy* and *comedian* as word candidates on the basis of the

vowel in the first syllable, whereas Spanish or Dutch listeners can only achieve such early discrimination by paying attention to the syllable's stress.

Thus listeners are usually able to distinguish between words in English using segmental information alone; suprasegmental information can be ignored with relatively little penalty. In Dutch, which has much less vowel reduction, and in Spanish, which does not allow vowel reduction at all, the penalty for ignoring suprasegmental information would be much more severe. This explains the different strength of the effects in the word recognition experiments. It explains why Dutch listeners outperform native English listeners at telling from which of two words differing in stress a given English syllable has been extracted (e.g., *mus-* from *music* versus *museum*; Cooper et al., 2002). The lesson from these studies is that quite small cross-language differences in how a given level of structure is realized (e.g., the likelihood of unstressed syllables manifesting vowel reduction) can radically affect the value of the structural information for rapid distinguishing between words, and thus determine the degree to which listeners make use of acoustic cues to that structure in speech.

4.2 Prosodic structure

The suprasegmental patterns of speech encode the prosodic structure in utterances. Even though prosodic and syntactic structure are independently determined (Shattuck-Hufnagel & Turk, 1996), it has long been known that listeners derive syntactic boundaries, and discourse boundaries too, from phrase-final lengthening and from the f_0 contour (see Cutler et al., 1997, for a review). More recently it has also become clear that the fine structure of segments is influenced by prosodic boundary placement, such that, for example, segments at the onset of a prosodic constituent are strengthened (Keating et al., 2003).

Listeners make use of such effects to parse speech into words. Cho et al. (2007) examined potentially ambiguous sequences such as *bus tickets*. There are competing English words in which there is no boundary after the /s/ (*bust*, *busty*), and their question was: could prosodic strengthening assist listeners in recognizing the word *bus* in this potentially ambiguous context? They compared utterances such as *When you get on the bus, tickets should be shown to the driver* (in which prosodic strengthening should enhance the /t/) versus *John bought several bus tickets for his family* (no strengthening). They found that the word *bus* was indeed more easily recognized in the phrase *bus tickets* taken from the former context than in the same phrase taken from the latter context. Likewise, Christophe et al. (2004) found that French word sequences such as *chat grinchaux* ('grumpy cat') were likely to be confused with the accidentally embedded *chagrin* if they formed part of a single phrase, but not if a phrase boundary occurred after *chat*.

Listeners are capable of using differences in the duration of segments within words to distinguish between alternative candidates (for instance, the difference in duration of syllable-final /l/ in Italian *silvestre* 'sylvan' versus syllable-initial /l/ in *silencio* 'silence' is available even without the following /v/ or /ɛ/ which disambiguates; Tabossi et al., 2000). Listeners are also capable of using the same sort of differences to distinguish between alternative phrases (for instance, the

/s/ duration distinguishes Dutch *een spot* 'a spotlight' from *eens pot* 'jar once'; Shatzman & McQueen, 2006a). The differences which are used vary across languages. Thus, in both English and Dutch, listeners make use of prosodic strengthening of /t/ to infer the onset of a prosodic constituent, but the way the strengthening is realized is exactly opposite in these languages (Cho & McQueen, 2005). In English, a strengthened /t/ has a longer VOT (which enhances the contrast with short-VOT /d/). In Dutch, strengthened /t/ has a shorter VOT (because the strengthening consists in a longer closure, enhancing the contrast with /d/ which in Dutch is prevoiced).

Even the durational differences which allow listeners to distinguish between stand-alone words and accidentally embedded words such as *hamster* versus *ham* embedded in *hamster* (Davis et al., 2002; Salverda et al., 2003; see also Shatzman & McQueen, 2006b) are modulated by prosodic structure. Thus *taxi* (/taksi/) is embedded in both the sequence *pak de tak sinaasappels* 'grab the branch of oranges' and *pak de tak citroenen* 'grab the branch of lemons'; but in the first, the syllable /si/ after *tak* is stressed, while in the second it is unstressed just as is the second syllable of *taxi*. Salverda et al. made two versions of a sentence containing, for example, *taxi*, by cross-splicing the *tak* syllable either from the above *sinaasappels* context or from the *citroenen* context. They found that the competing shorter word *tak* was more available in the former case. Moreover, they found that what primarily influenced the relative availability of the words was the duration of the initial syllable in the ambiguous portion; *ham*, *tak*, etc. are longer than the same syllables in *hamster*, *taxi*, etc., but the isolated words are also longer when they are followed by a stressed rather than an unstressed syllable. By manipulating the duration of this syllable, Salverda et al. found that they could influence what listeners considered as the most likely word at that moment.

4.3 Rhythmic structure and the recognition of continuous speech

Speech is a continuous signal without consistent demarcation of the words which make it up. Listeners must extract the component words from each speech signal in order to understand the speaker's message. One of the ways they do this is by exploiting the relationship between the rhythmic structure of speech and word-boundary location.

In English, and similar languages such as Dutch, rhythmic structure is stress-based, and segmentation of continuous speech can be usefully based on an assumption that strong syllables are most likely to be word-initial. Evidence that English- and Dutch-speaking listeners do actually operate with such an assumption comes partly from studies of word boundary misperceptions, in which listeners most commonly err by assuming strong syllables to be word-initial and weak syllables to be noninitial (Cutler & Butterfield, 1992; Vroomen et al., 1996). Further evidence is to be found in studies with the word-spotting task, in which real words embedded in nonsense bisyllables are harder to detect if detecting them requires processing segments from two consecutive strong syllables, i.e., across the canonical point of speech segmentation (Cutler & Norris, 1988; McQueen et al.,

1994; Norris et al., 1995; Vroomen & de Gelder, 1995; Vroomen et al., 1996). Syllable strength is here encoded in terms of vowel quality; recall that Fear et al.'s (1995) study described in section 4.1 showed that this is the most important feature of syllable strength in English. Cutler and Norris (1988), for example, compared detection of the word *mint* in *mintayf* and *mintef*; both bisyllables had initial stress, but they differed in the vowel which occurred in the second syllable. The embedded word *mint* was much harder to detect when the second vowel was strong, as in *mintayf*. The explanation is that the strong syllable *-tayf* triggered speech segmentation, so that the /t/ was momentarily considered to be the beginning of a new word rather than the end of *mint*; recovering from this interfering segmentation delayed recognition of *mint*.

In other languages with different rhythmic patterns, segmentation procedures also exploit rhythmic structure: syllabic rhythm in French (Cutler et al., 1986, 1992; Kolinsky et al., 1995) and Korean (Kim et al., 2008), moraic rhythm in Japanese (Otake et al., 1993; Cutler & Otake, 1994) and Telugu (Murty et al., 2007). In fact, rhythm allows a single, universally valid description of otherwise very different segmentation procedures used across languages (see Cutler, 1994, for further detail of this proposal).

Rhythmic structure allows listeners to predict accentual patterning as well. The initial phonemes of nonsense words are detected more rapidly when sentence rhythm predicts that the syllables containing the target will be accented (Shields et al., 1974). Pitt and Samuel (1990) presented acoustically constant versions of disyllabic minimal stress pairs at the ends of auditory lists in which all the disyllabic items had the same stress pattern; detection of a phoneme in these words was again faster when the syllable containing the target phoneme was predicted to be stressed, suggesting that listeners used the predictive information to attend selectively to stressed syllables. Likewise, listeners direct attention to words bearing sentence accent; detection of the initial phoneme of an acoustically constant word token is faster when the word occurs in a prosodic context consistent with sentence accent falling at that point than when it occurs in a context consistent with lack of accent (Cutler, 1976; Cutler & Darwin, 1981). Listeners can derive sufficient information to perform this attentional focus when f_0 variation has been removed (Cutler & Darwin, 1981), although when dimensions of prosodic information conflict such that, for example, timing predicts accent where f_0 predicts no accent, listeners refrain from deriving predictive information from prosody at all (Cutler, 1987).

5 Conclusions

Knowledge about the cognitive processes in speech perception has advanced considerably since we wrote our chapter in the first edition of this book. This development can be seen most clearly, perhaps, in the way the questions that are asked about this field have changed. In the 1990s, binary questions were being asked: Does lexical knowledge control prelexical decisions, or not? Is prelexical processing serial or is it cascaded? Is speech perception episodic or abstractionist? The questions we now have to ask are more nuanced. With respect to the issue of feedback of lexical knowledge, for example, recent research has shown that we have to distinguish between feedback in on-line processing and feedback in perceptual learning. A simple yes/no answer to whether there is feedback or not is no longer appropriate. Furthermore, given that it has been established that phonetic information cascades continuously to the lexical level, we now have to ask more specific questions about the nature of the information that is passed forward to lexical processing. Might its role depend on its informational value, as our review of cross-linguistic differences on uptake of lexical stress information suggests? Finally, we now have to consider how speech recognition can be episodic and abstractionist at the same time. All of these developments are indicative of a very active field of enquiry; they show that real progress is being made in our understanding of the cognitive aspects of speech perception.

The field is flourishing in another way. Even though the questions have become more refined, there are just as many being asked, and some that have still not been answered. For instance, we do not yet know what the "units of perception" are (McNeill & Lindig, 1973; Healy & Cutting, 1976). It seems clear that there is a prelexical level of processing that mediates between auditory (i.e., not speech-specific) processing and lexical processing, that the prelexical level involves abstraction and normalization, and that it operates in cascade. Furthermore, while it appears that this level is impervious to the immediate influence of lexical feedback, lexical knowledge can be used to retune prelexical processing over time. But we do not yet know what the unit(s) of representation are at the prelexical level. An important issue for future research will be to specify whether linguistic abstraction of segmental information prior to lexical access involves, for example, featural, allophonic, phonemic, or syllabic representations.

Critically, however, the way in which suprasegmental information is extracted prelexically will also have to be specified, and an account will have to be developed for the way in which segmental and suprasegmental information is integrated in modulating the word-recognition process. One possibility is that there are indeed two processing channels, one extracting segmental material (e.g., a mechanism computing the current sequence of phones), and one extracting suprasegmental material (e.g., a device building the prosodic structure of the current utterance). These two prelexical channels could operate independently, but could still both influence lexical processing. Another possibility is that there is a single processing channel which constructs an integrated multidimensional structure consisting of larger and smaller elements.

The current weight of evidence favors an autonomous account of on-line processing. If convincing data for on-line lexical–prelexical feedback were to be found, however, then it would be necessary to establish the cognitive function that such feedback serves in normal listening. Since on-line feedback appears to be of no benefit to word recognition, one possibility is that on-line effects are an epiphenomenon of the need for feedback in perceptual learning. If, however, the conclusion in favor of autonomy in on-line processing continues to stand the test

of time, then it will be necessary to develop a model of speech recognition in which lexical knowledge cannot modulate prelexical processing as it is happening, but can retune those processes over time.

An important constraint on the operation of the prelexical level, and thus possibly also on the nature of the representations constructed there, is that it must be flexible. The evidence we reviewed on perceptual learning in speech suggests that the way in which the speech signal is mapped onto the lexicon can be retuned after very brief exposure, and that that retuning can be specific to the speech of a single talker. It will be important to ascertain what the limits are on this kind of flexibility. For example, might there also be retuning of suprasegmental representations, and are other sources of knowledge (i.e., other than the lexicon) used to supervise perceptual learning?

Perhaps the greatest current challenge for cognitive modeling of speech perception, however, is how to include abstractionist and episodic components in the same model. Recent research suggests that episodic detail (e.g., how individual talkers produce specific speech sounds) is used to modulate the prelexical level. But this contribution of episodic knowledge to the flexibility of the prelexical processor is unlikely to be the whole story. Recent research also suggests that details of encounters with specific tokens of words are stored in long-term memory. The question, then, is how those long-term memories relate to abstract linguistic processing: do they exist only at the prelexical level, or also at the lexical level, or do they reside in a more general episodic memory store (i.e., not in the mental lexicon)?

It is important to note that this debate does not concern talker-specific segmental details alone. It also concerns suprasegmental details. For example, tokens of words differ in acoustic-phonetic detail because of their position in the prosodic hierarchy. Are all of these tokens stored, or is there abstraction of prosodic knowledge? This debate is also about the role of word frequency. One way in which the frequency of occurrence of a word can be coded is through storage of all encounters with that word, as in episodic models. But frequency can also be handled by models with abstract representations (either at the prelexical level or the lexical level, or both). Reconciling abstractionist and episodic accounts will thus entail specifying how multiple sources of information – about segments, suprasegmental structures, talker- and situation-specific details, and lexical frequency – are brought together as listeners hear spoken words. Experimentalists and computational modelers have plenty still to do.

REFERENCES

Allopenna, 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, 419–39.

Alphen, P. M. van & McQueen, J. M. (2006) The effect of voice onset time differences on lexical access in Dutch. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 178–96.

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994) The effect of subphonetic differences on lexical access. *Cognition*, 52, 163–87.

Aull, A. M. (1984) Lexical stress and its application in large vocabulary speech recognition. Masters dissertation, Massachusetts Institute of Technology.

Best, C. T. (1994) The emergence of language-specific phonemic influences in infant speech perception. In J. Goodman & H. C. Nusbaum (eds.), *Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* (pp. 167–224) Cambridge, MA: MIT Press.

Bond, Z. & Small, L. H. (1983) Voicing, vowel, and stress mispronunciations in continuous speech. *Perception and Psychophysics*, 34, 470–4.

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 Experimental Psychology: Human Perception and Performance*, 15, 567–75.

Cho, T. & McQueen, J. M. (2005) Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, 33, 121–57.

Cho, T., McQueen, J. M., & Cox, E. A. (2007) Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, 35, 210–43.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004) Phonological phrase boundaries constrain lexical access, I: Adult data. *Journal of Memory and Language*, 51, 523–47.

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, 521–33.

Cluff, M. S. & Luce, P. A. (1990) Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 551–63.

Cole, R. A. & Scott, B. (1974) Toward a theory of speech perception. *Psychological Review*, 81, 348–74.

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

Connine, C. M. & Clifton, C. (1987) Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 291–9.

Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997) Similarity mapping in spoken word recognition. *Journal of Memory and Language*, 37, 463–80.

Cooper, N., Cutler, A., & Wales, R. (2002) Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech*, 45, 207–28.

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

Cutler, A. (1987) Components of prosodic effects in speech recognition. *Proceedings of the 11th International Congress of Phonetic Sciences*, Tallinn, Estonia, 1, 84–7.

Cutler, A. (1994) Segmentation problems, rhythmic solutions. *Lingua*, 92, 81–104.

Cutler, A. (2005) Lexical stress. In D. B. Pisoni & R. E. Remez (eds.), *The Handbook of Speech Perception* (pp. 264–89). Oxford: Blackwell.

Cutler, A. & Butterfield, S. (1992) Rhythmic cues to speech segmentation:

Evidence from juncture misperception. *Journal of Memory and Language*, 31, 218–36.

Cutler, A. & Chen, H.-C. (1997) Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics*, 59, 165–79.

Cutler, A. & Clifton, C. (1984) The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhuis (eds.), *Attention and Performance, 10: Control of Language Processes* (pp. 183–96). Hillsdale, NJ: Lawrence Erlbaum.

Cutler, A., Dahan, D., & Donselaar, W. van (1997) Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40, 141–201.

Cutler, A. & Darwin, C. J. (1981) Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception and Psychophysics*, 29, 217–24.

Cutler, A. & Donselaar, W. van (2001) *Voornaam* is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech*, 44, 171–95.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986) The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25, 385–400.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987) Phoneme identification and the lexicon. *Cognitive Psychology*, 19, 141–77.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992) The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, 24, 381–410.

Cutler, A. & Norris, D. (1988) The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113–21.

Cutler, A. & Otake, T. (1994) Mora or phonemes? Further evidence for language-specific listening. *Journal of Memory and Language*, 33, 824–44.

Cutler, A. & Otake, T. (1999) Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustical Society of America*, 105, 1877–88.

Cutler, A. & Pasveer, D. (2006) Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition. *Proceedings of Speech Prosody 2006*, Dresden, Germany, 237–400.

Cutler, A., Weber, A., & Otake, T. (2006) Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, 34, 269–84.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001) Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16, 507– 34.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005) Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, 134, 222–41.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. (2002) Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 218–44.

Dell, G. S. & Newman, J. E. (1980) Detecting phonemes in fluent speech. *Journal of Verbal Learning and Verbal Behavior*, 19, 608–23.

Donselaar, W. van, Koster, M., & Cutler, A. (2005) Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology*, 58A, 251–73.

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

Eisner, F. & McQueen, J. M. (2005) The specificity of perceptual learning in speech processing. *Perception and Psychophysics*, 67, 224–38.

Eisner, F. & McQueen, J. (2006) Perceptual learning in speech: Stability over time. *Journal of the Acoustical Society of America*, 119, 1950–3.

Elman, J. L. & McClelland, J. L. (1986) Exploiting lawful variability in the speech wave. In J. S. Perkell & D. H. Klatt (eds.), *Invariance and Variability of Speech Processes* (pp. 360–80). Hillsdale, NJ: Lawrence Erlbaum.

Elman, J. L. & McClelland, J. L. (1988) Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, 27, 143–65.

Ernestus, M. & Mak, W. M. (2004) Distinctive phonological features differ in relevance for both spoken and written word recognition. *Brain and Language*, 90, 378–92.

Fear, B. D., Cutler, A., & Butterfield, S. (1995) The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America*, 97, 1893–904.

Foss, D. J. & Blank, M. A. (1980) Identifying the speech codes. *Cognitive Psychology*, 12, 1–31.

Foss, D. J. & Gernsbacher, M. A. (1983) Cracking the dual code: Toward a unitary model of phoneme identification. *Journal of Verbal Learning and Verbal Behavior*, 22, 609–32.

Foss, D. J., Harwood, D. A., & Blank, M. A. (1980) Deciphering decoding decisions: Data and devices. In R. A. Cole (ed.), *Perception and Production of Fluent Speech* (pp. 165–99). Hillsdale, NJ: Lawrence Erlbaum.

Fowler, C. A. (1984) Segmentation of coarticulated speech in perception. *Perception and Psychophysics*, 36, 359–68.

Fox, R. A. (1984) Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 526–40.

Fox, R. A. & Unkefer, J. (1985) The effect of lexical status on the perception of tone. *Journal of Chinese Linguistics*, 13, 69–90.

Frauenfelder, U. H., Segui, J., & Dijkstra, T. (1990) Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance*, 16, 77–91.

Ganong, W. F. (1980) Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 110–25.

Gaskell, M. & Marslen-Wilson, W. D. (1997) Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, 12, 613–56.

Gaskell, M. & Marslen-Wilson, W. D. (2002) Representation and competition in the perception of spoken words. *Cognitive Psychology*, 45, 220–66.

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, 1166–83.

Goldinger, S. D. (1998) Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–79.

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, 501–18.

Goldinger, S. D., Luce, P. A., Pisoni, D. B., & Marcario, J. K. (1992) Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1211–38.

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, 152–62.
Gordon, P. C. (1988) Induction of rate-dependent processing by coarse-grained aspects of speech. *Perception and Psychophysics*, 43, 137–46.
Gow, D. W., Jr. (2002) Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, 28, 163–179.
Gow, D. W., & Gordon, P. C. (1995) Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 344–59.
Hawkins, S. (2003) Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31, 373–405.
Healy, A. F. & Cutting, J. E. (1976) Units of speech perception: Phoneme and syllable. *Journal of Verbal Learning and Verbal Behavior*, 15, 73–83.
Johnson, K. (1997) Speech perception without speaker normalization: An exemplar model. In K. A. Johnson & J. W. Mullennix (eds.), *Talker Variability in Speech Processing* (pp. 145–66). San Diego, CA: Academic Press.
Johnson, K. (2005) Speaker normalization in speech perception. In D. B. Pisoni & R. Remez (eds.), *The Handbook of Speech Perception* (pp. 363–89). Oxford: Blackwell.
Johnson, K. A. & Mullennix, J. W. (eds.) (1997) *Talker Variability in Speech Processing*. San Diego, CA: Academic Press.
Keating, P., Cho, T., Fougeron, C., & Hsu, C. (2003) Domain-initial strengthening in four languages. In J. Local, R. Ogden, & R. Temple (eds.), *Laboratory Phonology 6* (pp. 143–61). Cambridge: Cambridge University Press.
Kim, J., Davis, C., & Cutler, A. (2008) Perceptual tests of rhythmic similarity, II: Syllable rhythm. *Language and Speech*, 51, 343–59.
Klatt, D. H. (1979) Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. A. Cole (ed.), *Perception and Production of Fluent Speech* (pp. 243–88). Hillsdale, NJ: Lawrence Erlbaum.
Klatt, D. H. (1989) Review of selected models of speech perception. In W. D. Marslen-Wilson (ed.), *Lexical Representation and Process* (pp. 169–226). Cambridge, MA: MIT Press.
Kolinsky, R., Morais, J., & Cluytens, M. (1995) Intermediate representations in spoken word recognition: Evidence from word illusions. *Journal of Memory and Language*, 34, 19–40.
Kouider, S. & Dupoux, E. (2005) Subliminal speech priming. *Psychological Science*, 16, 617–25.
Kraljic, T. & Samuel, A. G. (2005) Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51, 141–78.
Kraljic, T. & Samuel, A. G. (2006) Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review*, 13, 262–8.
Kraljic, T. & Samuel, A. G. (2007) Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56, 1–15.
Lee, C.-Y. (2007) Does horse activate mother? Processing lexical tone in form priming. *Language and Speech*, 50, 101–23.
Lehiste, I. (1970) *Suprasegmentals*. Cambridge, MA: MIT Press.
Lehiste, I. (1971) The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 51, 2018–24.
Leyden, K. van & Heuven, V. J. van (1996) Lexical stress and spoken word recognition: Dutch vs. English. In C. Cremers & M. den Dikken (eds.), *Linguistics in the Netherlands 1996* (pp. 59–170). Amsterdam: John Benjamins.
Liberman, A. M. & Mattingly, I. G. (1985) The motor theory of speech perception revised. *Cognition*, 21, 1–36.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991) Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, 89, 874–86.
Luce, P. A. & Large, N. R. (2001) Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes*, 16, 565–81.
Luce, P. A. & Lyons, E. A. (1998) Specificity of memory representations for spoken words. *Memory and Cognition*, 26, 708–15.
Luce, P. A. & Pisoni, D. B. (1998) Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1–36.
Maddieson, I. (1984) *Patterns of Sounds*. Cambridge: Cambridge University Press.
Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003) Lexical effects on compensation for coarticulation: A tale of two systems? *Cognitive Science*, 27, 801–5.
Mann, V. A. & Repp, B. H. (1981) Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, 69, 548–58.
Marslen-Wilson, W. D. (1987) Functional parallelism in spoken word-recognition. *Cognition*, 25, 71–102.
Marslen-Wilson, W. (1990) Activation, competition, and frequency in lexical access. In G. T. M. Altmann (ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* (pp. 148–72). Cambridge, MA: MIT Press.
Marslen-Wilson, W., Moss, H. E., & Halen, S. van (1996) Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1376–92.
Marslen-Wilson, W. & Warren, P. (1994) Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, 101, 653–75.
Martin, C., Mullennix, J., Pisoni, D., & Summers, W. (1989) Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 676–84.
Maye, J., Aslin, R. N., & Tanenhaus, M. T. (2008) The Weckud Wetch of the Wast: Lexical adaptation to a novel accent. *Cognitive Science*, 32, 543–62.
McClelland, J. L. (1987) The case for interactionism in language processing. In M. Coltheart (ed.), *Attention and Performance 12: The Psychology of Reading* (pp. 1–36). London: Lawrence Erlbaum.
McClelland, J. L. (1991) Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology*, 23, 1–44.
McClelland, J. L. & Elman, J. L. (1986) The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
McClelland, J. L., Mirman, D., & Holt, L. L. (2006) Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, 10, 363–9.
McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002) Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86, B33–B42.
McNeill, D. & Lindig, K. (1973) The perceptual reality of phonemes, syllables, words, and sentences. *Journal of Verbal Learning and Verbal Behavior*, 12, 431–61.
McQueen, J. M. (1991) The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 433–43.
McQueen, J. M. (1993) Rhyme decisions to spoken words and nonwords. *Memory and Cognition*, 21, 210–22.
McQueen, J. M. (2003) The ghost of Christmas future: Didn't Scrooge learn to be good? Commentary on Magnuson, McMurray, Tanenhaus, and Aslin (2003). *Cognitive Science*, 27, 795–9.

McQueen, J. M., Cutler, A., & Norris, D. (2003) Flow of information in the spoken word recognition system. *Speech Communication*, 41, 257–70.

McQueen, J. M., Cutler, A., & Norris, D. (2006) Phonological abstraction in the mental lexicon. *Cognitive Science*, 30, 1113–26.

McQueen, J. M., Jesse, A., & Norris, D. (2009) No lexical-prelexical feedback during speech perception or: Is it time to stop playing those Christmas tapes? *Journal of Memory and Language*, 61, 1–18.

McQueen, J. M., Norris, D., & Cutler, A. (1994) Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 621–38.

McQueen, J. M., Norris, D., & Cutler, A. (1999) Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1363–89.

McQueen, J. M., Norris, D., & Cutler, A. (2006) Are there really interactive processes in speech perception? *Trends in Cognitive Science*, 10, 533.

Mehler, J. (1981) The role of syllables in speech processing: Infant and adult data. *Philosophical Transactions of the Royal Society of London*, 295, 333–52.

Miller, J. L. (1981) Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (eds.), *Perspectives on the Study of Speech* (pp. 39–74). Hillsdale, NJ: Lawrence Erlbaum.

Miller, J. L. (1987) Mandatory processing in speech perception: A case study. In J. L. Garfield (ed.), *Modularity in Knowledge Representation and Natural-Language Understanding* (pp. 309–22). Cambridge, MA: MIT Press.

Miller, J. L. & Dexter, E. R. (1988) Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 369–78.

Miller, J. L., Green, K., & Schermer, T. M. (1984) A distinction between the effects of sentential speaking rate and semantic congruity on word identification. *Perception and Psychophysics*, 36, 329–37.

Minematsu, N. & Hirose, K. (1995) Role of prosodic features in the human process of perceiving spoken words and sentences in Japanese. *Journal of the Acoustical Society of Japan*, 16, 311–20.

Mirman, D., McClelland, J. M., & Holt, L. L. (2005) Computational and behavioral investigations of lexically induced delays in phoneme recognition. *Journal of Memory and Language*, 52, 424–43.

Mitterer, H. & Ernestus, M. (2006) Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, 34, 73–103.

Mullennix, J. W. & Pisoni, D. B. (1990) Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics*, 47, 379–90.

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, 365–78.

Murty, L., Otake, T., & Cutler, A. (2007) Perceptual tests of rhythmic similarity, I: Mora rhythm. *Language and Speech*, 50, 77–99.

Nearey, T. M. (1990) The segment as a unit of speech perception. *Journal of Phonetics*, 18, 347–73.

Newman, J. E. & Dell, G. S. (1978) The phonological nature of phoneme monitoring: A critique of some ambiguity studies. *Journal of Verbal Learning and Verbal Behavior*, 17, 359–74.

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

Norris, D. & McQueen, J. M. (2008) Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115, 357–95.

Norris, D., McQueen, J. M., & Cutler, A. (1995) Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1209–28.

Norris, D., McQueen, J. M., & Cutler, A. (2000) Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23, 299–370.

Norris, D., McQueen, J. M., & Cutler, A. (2003) Perceptual learning in speech. *Cognitive Psychology*, 47, 204–38.

Nygaard, L. C., & Pisoni, D. B. (1998) Talker-specific learning in speech perception. *Perception and Psychophysics*, 60, 355–76.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994) Speech perception as a talker-contingent process. *Psychological Science*, 5, 42–6.

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002) Learning phonotactic constraints from brief auditory exposure. *Cognition*, 83, B13–B23.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993) Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32, 258–78.

Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001) The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science*, 12, 445–9.

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, 309–28.

Pierrehumbert, J. (2002) Word-specific phonetics. In C. Gussenhoven & N. Warner (eds.), *Laboratory Phonology 7* (pp. 101–40). Berlin: Mouton de Gruyter.

Pitt, M. A. & McQueen, J. M. (1998) Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, 39, 347–70.

Pitt, M. A. & Samuel, A. G. (1990) The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 564–73.

Pitt, M. A. & Samuel, A. G. (1993) An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 699–725.

Port, R. F., Reilly, W. T., & Maki, D. P. (1988) Use of syllable-scale timing to discriminate words. *Journal of the Acoustical Society of America*, 83, 265–73.

Radeau, M., Morais, J., & Segui, J. (1995) Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 1297–311.

Raphael, L. J. (2005) Acoustic cues to the perception of segmental phonemes. In D. Pisoni & R. E. Remez (eds.), *The Handbook of Speech Perception* (pp. 182–206). Oxford: Blackwell.

Repp, B. H. & Liberman, A. M. (1987) Phonetic category boundaries are flexible. In S. R. Harnad (ed.), *Categorical Perception* (pp. 89–112). Cambridge: Cambridge University Press.

Rubin, P., Turvey, M. T., & Gelder, P. van (1976) Initial phonemes are detected faster in spoken words than in non-words. *Perception and Psychophysics*, 19, 394–8.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003) The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, 90, 51–89.

Samuel, A. G. (1977) The effect of discrimination training on speech perception: Noncategorical perception. *Perception and Psychophysics*, 22, 321–30.

Samuel, A. G. (1981a) Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, 110, 474–94.

Samuel, A. G. (1981b) The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 1124–31.
Samuel, A. G. (1987) Lexical uniqueness effects on phonemic restoration. *Journal of Memory and Language*, 26, 36–56.
Samuel, A. G. (1996) Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, 125, 28–51.
Samuel, A. G. (1997) Lexical activation produces potent phonemic percepts. *Cognitive Psychology*, 32, 97–127.
Samuel, A. G. (2001) Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science*, 12, 348–51.
Samuel, A. G. & Kat, D. (1996) Early levels of analysis of speech. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 676–94.
Samuel, A. G. & Pitt, M. A. (2003) Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language*, 48, 416–34.
Segui, J. & Frauenfelder, U. (1986) The effect of lexical constraints upon speech perception. In F. Klix & H. Hagendorf (eds.), *Human Memory and Cognitive Capabilities: Mechanisms and Performances* (pp. 795–808). Amsterdam: North-Holland.
Segui, J., Frauenfelder, U., & Mehler, J. (1981) Phoneme monitoring, syllable monitoring and lexical access. *British Journal of Psychology*, 72, 471–7.
Sekiguchi, T. & Nakajima, Y. (1999) The use of lexical prosody for lexical access of the Japanese language. *Journal of Psycholinguistic Research*, 28, 439–54.
Shattuck-Hufnagel, S. & Turk, A. E. (1996) A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25, 193–247.
Shatzman, K. B. & McQueen, J. M. (2006a) Segment duration as a cue to word boundaries in spoken-word recognition. *Perception and Psychophysics*, 68, 1–16.
Shatzman, K. B. & McQueen, J. M. (2006b) Prosodic knowledge affects the recognition of newly-acquired words. *Psychological Science*, 17, 372–7.
Shields, J. L., McHugh, A., & Martin, J. G. (1974) Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology*, 102, 250–5.
Shillcock, R. (1990) Lexical hypotheses in continuous speech. In G. T. M. Altmann (ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* (pp. 24–49). Cambridge, MA: MIT Press.
Slowiaczek, L. M. (1990) Effects of lexical stress in auditory word recognition. *Language and Speech*, 33, 47–68.
Slowiaczek, L. M. & Hamburger, M. (1992) Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1239–50.
Slowiaczek, L. M., McQueen, J. M., Soltano, E. G., & Lynch, M. (2000) Phonological representations in prelexical speech processing: Evidence from form-based priming. *Journal of Memory and Language*, 43, 530–60.
Small, L. H., Simon, S. D., & Goldberg, J. S. (1988) Lexical stress and lexical access: Homographs versus nonhomographs. *Perception and Psychophysics*, 44, 272–80.
Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001) Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, 45, 412–32.
Spinelli, E., McQueen, J. M., & Cutler, A. (2003) Processing resyllabified words in French. *Journal of Memory and Language*, 48, 233–54.

Stemberger, J. P., Elman, J. L., & Haden, P. (1985) Interference between phonemes during monitoring: Evidence for an interactive activation model of speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 475–89.
Stevens, K. N. (2002) Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, 111, 1872–91.
Stevens, K. & Blumstein, S. (1981) The search for invariant acoustic correlates of phonetic features. In P. Eimas & J. L. Miller (eds.), *Perspectives on the Study of Speech* (pp. 1–38). Hillsdale, NJ: Lawrence Erlbaum.
Strange, W. (1995) *Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research*. Timonium, MD: York Press.
Streeter, L. A. & Nigro, G. N. (1979) The role of medial consonant transitions in word perception. *Journal of the Acoustical Society of America*, 65, 1533–41.
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. 201–9). Amsterdam: North-Holland.
Tabossi, P., Burani, C., & Scott, D. (1995) Word identification in fluent speech. *Journal of Memory and Language*, 34, 440–67.
Tabossi, P., Collina, S., Mazzetti, M., & Zoppello, M. (2000) Syllables in the processing of spoken Italian. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 758–75.
Utman, J. A., Blumstein, S. E., & Burton, M. W. (2000) Effects of subphonetic and syllable structure variation on word recognition. *Perception and Psychophysics*, 62, 1297–311.
Vitevitch, M. S. & Luce, P. A. (1998) When words compete: Levels of processing in perception of spoken words. *Psychological Science*, 9, 325–9.
Vitevitch, M. S. & Luce, P. A. (1999) Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40, 374–408.
Vroomen, J. & Gelder, B. de (1995) Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 98–108.
Vroomen, J. & Gelder, B. de (1997) Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 710–20.
Vroomen, J., Linden, S. van, Gelder, B. de, & Bertelson, P. (2007) Visual recalibration and selective adaptation in auditory–visual speech perception: Contrasting build-up courses. *Neuropsychologia*, 45, 572–7.
Vroomen, J., Zon, M. van, & Gelder, B. de (1996) Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory and Cognition*, 24, 744–55.
Waibel, A. (1988) *Prosody and Speech Recognition*. London: Pitman.
Warren, R. M. (1970) Perceptual restoration of missing speech sounds. *Science*, 167, 392–3.
Warren, R. M. & Obusek, C. J. (1971) Speech perception and phonemic restorations. *Perception and Psychophysics*, 9, 358–62.
Weber, A. & Cutler, A. (2004) Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50, 1–25.
Whalen, D. (1984) Subcategorical phonetic mismatches slow phonetic judgments. *Perception and Psychophysics*, 35, 49–64.
Whalen, D. (1991) Subcategorical phonetic mismatches and lexical access. *Perception and Psychophysics*, 50, 351–60.

Wickelgren, W. A. (1969) Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, 76, 1–15.

Ye, Y. & Connine, C. M. (1999) Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, 14, 609–30.

Yip, M. C. (2001) Phonological priming in Cantonese spoken-word processing. *Psychologia: An International Journal of Psychology in the Orient*, 44, 223–9.

Zwitserlood, P. (1989) The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, 32, 25–64.