LANGUAGE AND COGNITIVE PROCESSES, 2001, *16* (5/6), 469–490

Spoken word access processes: An introduction

James M. McQueen and Anne Cutler
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

We introduce the papers in this special issue by summarising the current major issues in spoken word recognition. We argue that a full understanding of the process of lexical access during speech comprehension will depend on resolving several key representational issues: what is the form of the representations used for lexical access; how is phonological information coded in the mental lexicon; and how is the morphological and semantic information about each word stored? We then discuss a number of distinct access processes: competition between lexical hypotheses; the computation of goodness-of-fit between the signal and stored lexical knowledge; segmentation of continuous speech; whether the lexicon influences prelexical processing through feedback; and the relationship of form-based processing to the processes responsible for deriving an interpretation of a complete utterance. We conclude that further progress may well be made by swapping ideas among the different sub-domains of the discipline.

The proportion of the world's population that is kept awake at night worrying about spoken word access processes is, undoubtedly, vanishingly small. After all, hardly anyone has even heard the phrase *Spoken word access processes*. Such things are hardly the stuff mobile-phone conversations are made on. And yet in another way, as we psycholinguists know, spoken word access processes are what those conversations depend on. An English listener chosen at random might not understand this phrase as a whole, but this would not be because of a failure to recognise the words themselves.

Spoken word recognition is remarkably robust and seemingly effortless. Chances are high that our native speaker of English would be able to recognise our four-word phrase with no difficulty, even if it were spoken over a mobile phone in a noisy station by someone whose voice he or she had never heard before, irrespective of how fast that talker spoke, and

Requests for reprints should be sent to James M. McQueen, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands.
Email: james.mcqueen@mpi.nl

http://www.tandf.co.uk/journals/pp/01690965.html DOI: 10.1080/01690960143000209

irrespective of the sex or age of the talker. This is in spite of the fact that all of these factors can radically alter the acoustic characteristics of the phrase. The robustness and effortlessness of spoken word recognition make it something that listeners simply take for granted. They do not appreciate that there is a complex problem to be solved in speech recognition; it just happens. How often have spoken-word recognition researchers, in answer to polite questions about what it is they do, found themselves explaining not their research itself, but rather that there is in fact a domain of enquiry there to be studied? One of the problems here is the lack of a simple word in English to describe the act of recognising words in the spoken domain, like "to read" in the written domain. Curiously, other languages do have speech-specific listening words (like "verstaan" in Dutch and "kikitoru" in Japanese); anglophones appear particularly uninterested in speech recognition.

Spoken word access processes, therefore, are mental processes which listeners take for granted. They are not the processes by which a listener interprets the sequence of words in an utterance, but rather the processes by which that sequence of words is derived from the acoustic speech signal. They are the perceptual processes which take the sequence of buzzes, bursts, and chirps that make up the raw acoustic signal and convert them into a sequence of words. These word-sequences form our primary perceptual representation of spoken language and form the input to the interpretative processes by which we derive the meaning of utterances.

This volume contains articles and short reports based on presentations at the workshop Spoken Word Access Processes (SWAP), held in Nijmegen in May 2000. The papers cover the major issues that the field is currently concerned with, and thus, like the workshop, provide a snapshot of the state of the SWAP art. We summarise those issues here, as they relate to the papers in this volume, and to the other papers presented at the SWAP workshop (as listed in the Appendix to the Preface). We will refer to the authors of papers and short reports in the present issue using bold type, and to the others who presented at SWAP using italics.

Although the field is only just over a quarter of a century old, much has already been learned about spoken word recognition. As the present papers attest, however, much remains uncertain. Fortunately, as the papers also attest, the field is still very active. We hope that in another 25 years we will know yet more about spoken word access processes, even if the mobile-phone users of the 2020s continue to be blissfully ignorant of them.

THE SPOKEN WORD

What, then, are spoken words? What are the mental entities which we recognise when we listen to spoken language? Many answers to this question have been proposed, ranging from the claim that words in the mental lexicon are highly detailed episodic representations (i.e., each individual utterance of a word is stored in the form in which it is heard, coding, for example, information about the speaker that spoke it and the nature of any background noise) to the claim that the lexicon consists of highly abstract phonological representations (such as those in under-specification theory which code only the information necessary to specify the phonological form of a word). Three themes related to this issue were discussed at SWAP and are represented in this volume: the unit of perception, phonological representation, and semantic and morphological representation.

The unit of perception

The spoken word does not consist of easily identifiable and extractable subunits. While it is possible to describe a spoken word as a sequence of phonemes, or as a sequence of syllables, or as a complex bundle of acoustic-phonetic features, it is extremely hard to identify those units (of whatever grain size) in the acoustic speech signal. In particular, it is impossible to identify the exact temporal locations of the beginnings or ends of phonemes (or features, syllables or words). Since talkers coarticulate (the vocal-tract gestures for neighbouring sounds can be made simultaneously), the speech signal does not consist of a sequence of discrete units. One of the most venerable but still recurring issues in spoken word recognition, therefore, is whether there is an intermediate level of representation between the input and the mental lexicon, and, if so, what the ‘‘units of perception’’ at that level of processing are.

This issue is of course intimately related to the issue of the form that words take in the mental lexicon. Since these intermediate units form the lexical access code (the means by which the speech signal makes contact with the lexicon) then, on the simplest account, access and lexical representations should have the same units (e.g., phonemic access representations should map onto a phonemically structured lexicon). At the very least, there should be a straightforward mapping of prelexical representations onto lexical representations.

Nearey argues that phonemes play a central role in speech recognition. He presents simulations showing how the recognition of nonsense syllables can be very well predicted from the recognition of their component phonemes. He also argues that a model in which syllables are factored into their phonemes can account for the results of multidimensional phonetic categorisation experiments. **Miller** takes Nearey’s line of argument further. She bases her argument on research examining variation in speaking rate. She has shown in earlier work that the processes which make adjustments for changes in speaking rate are mandatory (e.g., Miller & Dexter, 1988); listeners automatically adjust their criteria for interpreting durational cues to speech sounds depending on the speaking rate. One such cue is Voice Onset Time (VOT), in English stop consonants. A VOT which counts as a good /p/, for example, at one speaking rate, will be judged to be a poorer /p/ at a different speaking rate (Miller & Volaitis, 1989). In the present paper, Miller describes how changes in speaking rate and changes in lexical context have qualitatively different effects on category goodness judgements. A key underlying assumption, however, consistent with Nearey's arguments, is that there are prelexical representations which are essentially phonemic in nature. Category goodness judgements are therefore considered to be based on phonemic categories which are extracted prelexically, and which have internal structure.

On the basis of these two papers, it might therefore appear that there is now agreement not only that there are ''units of perception'', but also that they are phonemic. This, however, is definitely not the case. *Goldinger*, following up on his earlier research (e.g., Goldinger, 1996, 1998), argued at SWAP for the episodic view: that the lexicon consists of detailed episodic traces of individual words. Listeners certainly appear to have detailed memories of specific instances of words. Palmeri, Goldinger, and Pisoni (1993), for example, showed that listeners could recognise more rapidly that a word in a spoken list had already occurred in that list if the word was said by the same speaker earlier in the list than if it had been said by someone else. Goldinger showed, furthermore, that subjects appear to imitate words that they have previously heard, that is, their utterances of a particular word tend to be more like another token of that word (spoken by another speaker) after they have heard that token than before they have heard it. These results appear inconsistent with the view that the lexicon consists only of abstract phonological representations, which can only be accessed by abstract phonemic representations. *Hawkins and Nguyen*, in their SWAP poster, also argued against this strong abstractionist position. They showed that listeners' decisions about word-final stops are influenced by the acoustic-phonetic nature of the [l] at the onset of those words. These kinds of data suggest that coarticulation can create dependencies even between non-adjacent sounds in production. It would also be more consistent with a model in which lexical entries contain subtle acoustic-phonetic details rather than abstract strings of phonemes.

Dupoux, Pallier, Mehler, and Kakehi, on the other hand, argue for abstract intermediate representations. Results from an experiment on Japanese vowel epenthesis suggest that Japanese listeners use their knowledge about the phonological structure of Japanese to create an abstract lexical access code. They appear to create, for example, epenthetic vowels in locations where such vowels should appear in the speech signal (i.e., between consonants) even if there is no vocalic information actually present in the input. While Dupoux et al. are agnostic about the nature of prelexical representations (they could be phonemic, syllabic, or moraic), their data certainly challenge the view that there are no such representations.

It therefore seems that the most extreme positions in the debate about "units of perception" are no longer tenable. Those arguing for an episodic account of the lexicon, with no intermediate abstract phonological representations, need to address the data which demonstrate abstraction and normalisation. The reverse is of course equally true: those arguing for prelexical representations, be they phonemic or otherwise, need to confront the data showing that listeners are able to remember the fine detail associated with particular utterances of particular words. There is therefore considerable scope for the development of what one might call "hybrid models", that is, accounts of spoken word recognition with both episodic and abstractionist components.

One of the major problems with the old "units of perception" debate (are syllables, phonemes, features or something else the basic units?) was that the data simply did not distinguish clearly between these alternatives (see, e.g., McQueen & Cutler, 1997; Pisoni & Luce, 1987). As *Whalen* pointed out at SWAP, this remains a problem. Marslen-Wilson and Warren (1994), for example, made the case that acoustic-phonetic features are the basic lexical access units, but it has since been shown that their data can equally well be explained by a model with phonemic access representations (**Dahan, Magnuson, Tanenhaus, & Hogan**; McQueen, Norris, & Cutler, 1999; Norris, McQueen & Cutler, 2000). We hope that more progress may be made on this issue if more attention is devoted to the distinction between abstractionist and episodic accounts than to distinctions among different types of access units.

Phonological representation

Spoken words are highly variable. In normal speech, words rarely appear in their citation forms, that is, as they would do if spoken slowly and carefully in isolation. Some of this variation is due to factors which we have already mentioned; changes in the rate of speech of the talker, the talker's age, sex, dialect, and speaking style, as well as changes in the amount and nature of background noise, all influence the acoustic form of a spoken word. Though some of this variation is very unpredictable (the exact form of a novel talker's vocal tract; how carefully a talker will utter a particular word), some of it is more predictable (the spectral structure of a talker's vowels, indicating for example that she is female, can be used to predict the form of her vowels later in the utterance; faster speaking rate will tend to shorten all segments in an utterance). This kind of variation has in fact provided one of the primary motivations for a prelexical level of processing. Normalisation processes can act on the signal at this level of processing, to adjust for this variability and to generate an abstract code which can then be used for lexical access. Further research on the effects of fine-grained acoustic-phonetic detail due to this kind of variation on the lexical access process should lead to advances in our understanding of prelexical representations.

Another kind of variation, however, is very predictable. Changes in the surface form of spoken words can be predicted by phonological rules of assimilation, epenthesis, deletion, and resyllabification. Several presentations at SWAP addressed how this kind of variation is dealt with during spoken word recognition. How is the word *sweet*, for example, recognised when it is pronounced as [swik] in the phrase *sweet girl*? Answers to this question again have an intimate relationship with the issue of the phonological form of words in the mental lexicon. One proposal, for example, is that the lexicon consists of underspecified phonological representations (e.g., Lahiri & Marslen-Wilson, 1991). Since English words ending in coronal stops, like *sweet*, can undergo place assimilation, and take on the place of articulation of a following bilabial or velar consonant, their lexical representations can be considered to lack specification of the place of articulation of their final phonemes. Thus, when a listener hears [swik], for example, the final [k] will not mismatch with the lexical representation of *sweet*.

Coenen, Zwitserlood, and Bölte present evidence from German which challenges this view. They show that recognition of an assimilated form depends on the availability of an appropriate phonological context: the pattern of results for words in isolation was not the same as for words in sentence contexts. These results are consistent with similar findings in English (Gaskell & Marslen-Wilson, 1996, 1998; Marslen-Wilson, Nix, & Gaskell, 1995). Just as in phonology, where many theorists have moved away from underspecification theory (towards, e.g., Optimality Theory), so too are psycholinguists moving away from the view that the lexicon consists of underspecified representations. Results such as those of Coenen et al. suggest that, even if lexical representations are abstract, there need also to be what Gaskell and Marslen-Wilson (1996, 1998) have referred to as phonological inference processes: mechanisms which evaluate the surface form of words in the phonological context of their neighbouring words.

Pierrehumbert also discusses the phonological form of words in the mental lexicon, but from a rather different perspective. The question which she addresses concerns learning: how we come to master the phonological regularities of our native language. She shows that as vocabulary size increases, even though different listeners will have different words in their vocabularies, more complex phonological regularities can be inferred. If the recognition of words that have undergone assimilation (or had some other phonological process applied to them in production) requires inference processes, then these processes must themselves be acquired. Accounts of the acquisition of spoken language therefore need to explain not only how spoken forms are learned, and how those forms are linked to meanings, but also how phonological inferencing processes can be acquired through exposure to the very words which will later be recognised through the operation of those processes.

Semantic and morphological representation

The third issue concerning representation that was discussed at SWAP was the relationship between the phonological form of a word and the other information stored in the mental lexicon. How is the knowledge about the form of a word linked to its syntactic and semantic properties? *Shillcock* showed that there are small but striking relationships between form and meaning: similar sounding words tend to be similar in meaning. Phonetic symbolism of this type has implications for the structure of the lexicon and, Shillcock argued, may be the result of the brain's tendency to store information topographically. *Rodd, Gaskell and Marslen-Wilson's* poster focused more exclusively on the semantic level. They discussed how semantically ambiguous words are recognised. It appears that words with multiple meanings (e.g., *bark*) are recognised more slowly than unambiguous words, while those with multiple senses (e.g., *twist*) are recognised more rapidly. Rodd et al. argued that these results are consistent with models of the lexicon with distributed semantic representations: competition between multiple meanings delays recognition, while the additional information provided by words with multiple senses benefits recognition.

How might these semantic representations be linked to representations of word forms? One common assumption is that, intervening between sound and meaning representations, there are morphemic representations. Morphemes are, after all, the link between form and content. While most psycholinguists agree that there is a morphological level of representation (e.g., Marslen-Wilson, Tyler, Waksler & Older, 1994; Schreuder & Baayen, 1995; see McQueen & Cutler, 1998, for a review), there is considerable disagreement about exactly what kind of information is coded there. Are morphologically complex words stored as wholes, or in their component parts, and, if so, is this equally true for derived, inflected, and compounded words? **Marslen-Wilson** reviews recent cross-linguistic work on this issue. This kind of research was reported in several posters at SWAP: work on Arabic by *Boudelaa and Marslen-Wilson*, on French by *Meunier, Marslen-*

Wilson and Ford and on Polish by *Reid and Marslen-Wilson*. It appears that there are no language-universals here: the kind of morphological information that is stored in a given listener's mental lexicon seems to depend on the morphological structure of that listener's native language; some languages encourage a more strongly decompositional form of representation than others.

Cross-linguistic research has proved to be very valuable in developing our understanding of spoken word recognition, for example in the domain of speech segmentation (see, e.g., Cutler, Mehler, Norris & Segui, 1986; Otake, Hatano, Cutler & Mehler, 1993). Such comparisons can show which aspects of speech processing are shared by speakers of all languages, and which reflect adaptation to particular structures in the listener's native language. They can also show the limitations in theorising which may be imposed when most of the available data come from a very small number of languages. We therefore welcome and encourage more cross-linguistic work, not only in the domain of morphology, where it has again already proved to be very valuable, but also in other domains of SWAP.

ACCESS PROCESSES

Competition

How then are spoken words accessed and recognised? The field began to make serious progress on this aspect of the topic with the advent of computational models in the mid-1980s. There is now virtually unanimous agreement on the broad outline of the lexical access process. From an enormous amount of research we know that lexical access involves continuous activation of multiple candidate words, and that there is a process of competition between the activated candidates out of which the eventual winning words emerge (see, for example, Allopenna, Magnuson, & Tanenhaus, 1998; Cluff & Luce, 1990; Gow & Gordon, 1995; Marslen-Wilson, 1987, 1990; Marslen-Wilson, Moss, & Van Halen, 1996; McQueen, Norris, & Cutler, 1994; Norris, McQueen, & Cutler, 1995; Shillcock, 1990; Swinney, 1981; Tabossi, Burani, & Scott, 1995; Vroomen & de Gelder, 1995, 1997; Wallace, Stewart, & Malone, 1995; Wallace, Stewart, Sherman, & Mellor, 1995; Zwitserlood, 1989; Zwitserlood & Schriefers, 1995).

Evidence from phonological priming studies (Goldinger, Luce, & Pisoni, 1989; Goldinger, Luce, Pisoni, & Marcario, 1992; Monsell & Hirsh, 1998; Radeau, Morais, & Segui, 1995; Slowiaczek & Hamburger, 1992; and *Sereno & Quené*, at SWAP) also suggests that candidate words compete with one another during word recognition. *Amano and Kondo's* statistical analyses of the structure of the Japanese lexicon were inspired by the idea that multiple candidate words are activated when a spoken word is heard. *Goswami and De Cara* showed that the emergence of phonological awareness in children, specifically children's judgements about rhyme, is influenced by the number of words sharing the same rime. These results suggest yet again that spoken word recognition involves the activation of multiple candidate words.

Vitevitch and Luce (1998, 1999) have argued that the spoken word recognition system is sensitive to sound similarities between words at two distinct levels of processing. If many words share the same sequence of phonemes, that sequence will tend to occur often in speech. Vitevitch and Luce observed facilitatory effects due to the frequency of phoneme sequences, which they attributed to a sublexical level of processing, where common sequences of sounds are easier to process than rare sequences. This level of processing might be analogous to the prelexical level of processing discussed above – an intermediate level that acts as an interface between the acoustic-phonetic speech signal and the lexicon. Vitevitch and Luce also observed inhibitory effects of sound similarity and argued that they were due to competition between words at the lexical level. If a word has many lexical neighbours (i.e., similar sounding words), which are all activated when that word is heard, that word will be harder to recognise than a word in a sparse neighbourhood. **Luce and Large** provide further evidence for the simultaneous operation of sublexical facilitation and lexical competition during spoken word recognition.

Mismatch

Not all the details of the activation and competition process are agreed on, however. Many issues are still subject to research and debate, including: which information affects activation; whether some types of information are more strongly weighted than others; and exactly how the competition process is structured (e.g., does it involve lateral inhibition?). These issues received considerable attention at SWAP. Among the posters was work on the use of tone information in lexical access in Mandarin (*Tao*), and work on how both native and non-native listeners process lexical stress information while listening to English (*Cooper*). The question here is whether prosodic information is used to constrain lexical access. Answers to this question have clear implications not only for the form of lexical representations, but also for the form of prelexical representations. Purely phonemic access representations, for example, cannot code lexical stress information. The study of non-native speech processing is of course interesting in its own right. How do listeners recognise the words of a second language? Second language listeners certainly draw on access processes used for first language comprehension in recognising their second language (see Cutler, in press, for a review); but are they able to learn new procedures or to suppress inappropriate old procedures? Second-language research is, however, also relevant to models of native word recognition; the way listeners process their second languages is likely to be informative about how they process their first languages. As with other cross-linguistic work, second language research provides a means of examining to what extent access processes are language-universal, and to what extent they are tuned to the specific phonological properties of particular languages.

The poster by *Bölte and Coenen* examined segmental rather than suprasegmental constraints on lexical access. To what extent is a word activated when the input mismatches with that word by one phoneme? *Van der Lugt* examined whether mismatching information in a carrier word fragment would influence the lexical competition process, and hence the detection of a shorter word embedded in the carrier. The work by Bölte and Coenen and by van der Lugt follows in the tradition of work on segmental mismatch carried out by Connine and colleagues (e.g., Connine, Blasko, & Wang, 1994; Connine, Titone, Deelman, & Blasko, 1997). This kind of work has clear parallels with that done on assimilation. In the assimilation case, a phoneme is substituted which, though appropriate to the phonological environment, is inconsistent with the isolated, citation form of the word. In the other work on mismatch, however, the substituted phoneme is not licensed by the phonological environment. Although the system appears fairly intolerant of this kind of mismatching information, small amounts of mismatch do not necessarily block lexical access (Connine et al., 1994, 1997; Marslen-Wilson et al., 1996). The access system appears to be more tolerant, however, of the ‘‘mismatch’’ caused by assimilation or other phonological processes. In other words, the system seems to tolerate natural variation.

Frauenfelder, Scholten, and Content test the predictions that the TRACE model (McClelland & Elman, 1986) and the Shortlist model (Norris, 1994) make about the effects of mismatching segmental information on lexical activation. They present results from two phoneme monitoring experiments in which target phonemes appeared in real French words and in nonwords which had been constructed by changing one or more phonemes of French words. They manipulated the position of the target phoneme and, in the nonwords, the position of the phoneme(s) which mismatched with the original real word (e.g., the [f] in *focabulaire* and the [n] in *vocabulaine*, both based on *vocabulaire*, vocabulary). Frauenfelder et al. argue that their results challenge TRACE (where mismatching information does not actively count against the activation of candidate words), but are consistent with Shortlist (where bottom-up inhibition from mismatching phonemes can strongly deactivate candidate words). As this paper attests, the evaluation of computational implementations of models of spoken word recognition has become an important part of the field.

Subsegmental constraints on lexical access were also examined at SWAP. The poster by *van Alphen* showed that variation in the amount of prevoicing in Dutch voiced stops does not appear to influence the degree of activation of words containing those stops. In this particular case, then, variation in the signal does not seem to affect the outcome of the lexical access process. In their poster, however, *Davis, Marslen-Wilson, and Gaskell* argued that words embedded in other words, like *cap* in *captain*, are acoustically different from unembedded tokens of those words (i.e., when the talker actually intends *cap*), and that the lexical access system is sensitive to these differences.

Dahan et al. examine other subsegmental effects: the influence on lexical access of mismatch between place of articulation cues in vowels and place cues in following stops (e.g., in a token of *net*, made by splicing the [nɛ] from *neck*, with cues in the vowel signalling the upcoming velar [k], to a final alveolar [t] release burst). Using an eye-tracking paradigm (subjects were asked to look at a visual display and to follow instructions like "click on the net"), Dahan et al. found evidence that word activation is influenced by subsegmental mismatch and evidence of competition between activated candidate words. These results support those obtained with similar materials using gating, lexical decision, and phoneme decisions tasks (Marslen-Wilson & Warren, 1994; McQueen et al., 1999). Dahan et al. also simulate their data with the TRACE model. Their paper again demonstrates how important it is to evaluate how well computational models can account for a particular set of data. The interplay between data and modelling will, we hope, continue to drive research in spoken word recognition.

The eye-tracking paradigm holds considerable promise; it appears that subtler differences in lexical activation can be tracked (and with greater temporal resolution) with this paradigm than with more traditional reaction time paradigms. While the lexical access system may be sensitive to the subsegmental differences, however, it is also very robust. Thus the asymptotic state of the recognition system should be the same for different surface variants of the same underlying utterance (i.e., the same words will be recognised in spite of small variations in the surface forms). It is thus perhaps not surprising that subsegmental mismatch effects appear to be rather subtle. We feel that considerable work remains to be done to map out the effects of fine-grained phonetic information on the lexical access process, using eye-tracking and other paradigms. As we pointed out earlier, the speech signal contains finer quality information than can be captured by a phonemic transcription, both with respect to the information available at any one moment in time and with respect to how that information can change over a very fine time-scale. Attention to this kind of detail will be important for the development not only of accounts of prelexical representation but also of the lexical competition process.

Segmentation

Spoken words usually occur not by themselves but in the middle of a running stream of speech sounds. Coarticulation of speech sounds occurs not only within but between words. This means that, at least within phonological phrases, speech tends to be a continuous stream of sounds, rather than a discontinuous sequence of words. Speech certainly lacks the reliable and unambiguous marking of word boundaries which is provided by the white spaces between the words in this text. One of the essential components of the lexical access process is therefore segmentation: how are discrete words recovered from the speech stream? Several presentations at SWAP were devoted to this theme.

Content, Meunier, Kearns, and Frauenfelder examine segmentation in French. They challenge the old idea that word recognition in French is based on a syllabic classification of the speech stream (i.e., that the prelexical "unit of perception" is the syllable; Mehler, 1981; Mehler, Dommergues, Frauenfelder, & Segui, 1981). They argue instead that syllables provide cues for lexical segmentation, specifically, that syllable onsets indicate the likely locations of word boundaries. While the theory of segmentation based on syllabic classification necessarily makes a strong claim about the nature of prelexical representations, the theory proposed by Content et al. is neutral on this topic. No claim about the size of the units in the lexical access code needs to be made; all that is required is that the locations of syllable onsets can be extracted prelexically and then used to constrain lexical access.

The view proposed by Content et al. has close links with that discussed by **Norris, McQueen, Cutler, Butterfield, and Kearns**. On Norris et al.'s view, segmentation is achieved through lexical competition (i.e., as in both TRACE and Shortlist). The competition process is, however, enriched by a segmentation procedure, the Possible Word Constraint (PWC; Norris, McQueen, Cutler, & Butterfield, 1997). The PWC uses cues in the speech signal to the location of likely word boundaries (including, for example, syllable onset locations in French). The PWC also uses information about what speech material constitutes a plausible part of the ongoing lexical parse. A section of speech between a candidate word and the location of a likely word boundary must contain a vowel. If there is only consonantal material, then that candidate word is not likely to be part of the utterance and its activation is therefore reduced (thus, for example, *arm* in "sheep farm" is penalised because the [f] between the beginning of *arm* and the boundary between the [p] and the [f], in this case signalled by the phonotactics of English, is not a possible word). Note that the boundary cues listeners use are necessarily language-specific. The sequence [pf], for example, though impossible within an English syllable, is possible in German syllables. Norris et al. review evidence which suggests that the PWC itself is, however, a simple, language-universal constraint. Irrespective of the phonological constraints as to what constitutes a well-formed word in any particular language, the information that is used in on-line segmentation is simply whether the residue of speech between a candidate word and a likely word boundary contains a vowel.

Four posters at SWAP also addressed issues in segmentation. *Dumay, Frauenfelder and Content* presented further evidence on French segmentation, arguing that in addition to the cues provided by syllable onsets, lexical competition has an important role to play. Dumay et al. also argued, along with *Smith and Hawkins*, that fine differences in the available acoustic-phonetic information have consequences for segmentation: some boundaries may be marked more clearly than others. In Smith and Hawkins' case, these acoustic differences were due to allophonic variation (differences between the syllable-onset and syllable-coda allophones of English consonants). *Kirk* had a similar point to make about English listeners' use of allophonic variation in segmentation. She also argued that constraints from phonological grammar (e.g., the tendency to maximise the number of consonants in a syllable onset, or the tendency for stressed syllables to attract consonants) can also influence how continuous speech is segmented. *Weber* showed that German listeners use their knowledge of both German phonotactics and English phonotactics in the segmentation of English speech. These findings are consistent with other demonstrations of the use of phonotactics in segmentation (Van der Lugt, in press; McQueen, 1998; Yip, 2000). Weber's results also show how the procedures that are adopted when learning to segment and recognise one's native language are recruited in segmenting a second language.

There is therefore now a relatively long list of cues which listeners appear to use for segmentation and which vary among languages, including phonotactics, allophonics, other acoustic-phonetic cues, silence, and metrical cues based on the input language's rhythmic structure. An important issue which remains to be addressed is the relative ranking of these cues: do some cues carry more weight in segmentation than others? Another issue for future research links work on segmentation to work on mismatching information in lexical access. To what extent do cues in the speech signal have their effect by directly influencing the degree of activation of lexical representations, and to what extent do they provide boundary markers, by which a segmentation procedure like the PWC can then, indirectly, influence lexical activation?

Feedback?

Despite the agreement that has been achieved on the sort of architecture needed to model access processes, there is one giant issue of disagreement still remaining, namely whether the flow of information within the spoken-word recognition system is unidirectional or bidirectional. Is there feedback from the lexicon to the prelexical level, such that lexical knowledge can influence the earlier stages in the lexical access process? There was extensive discussion of this issue at the SWAP workshop. Indeed, one of the motivations for the workshop was to continue the discussion of this issue as centered round the Merge model (Norris et al., 2000), in which there is no feedback. Many of those present at SWAP had contributed commentaries on the *Behavioral and Brain Sciences* target article, and many had attended a stimulating discussion on the topic at the fall 1999 ASA meeting in Columbus, Ohio. SWAP was thus round three in this debate.

One important result to come out of the meeting was that it appeared that a consensus had been reached on what counts as a necessary test of feedback. Experimental demonstrations of lexical involvement in phonemic decision-making can be explained either as a result of feedback from the lexicon to prelexical phonemic representations, the activation of which then determine the decisions (as in TRACE), or as a result of feedforward connections from the lexicon to a level of processing where explicit phonemic decisions are made (as in Merge). What is needed to distinguish between these accounts is therefore an unambiguous measure of prelexical processing. We can then ask whether the lexicon influences prelexical processing (as a feedback model would predict) or not (as a model without feedback would predict). One such measure of prelexical processing is the demonstration that the perceptual system appears to adjust for coarticulation between neighbouring consonants. Specifically, identification of stop consonants varies depending on the nature of preceding fricatives (Mann & Repp, 1981) or liquids (Mann, 1980). Most psycholinguists are agreed that this compensation for coarticulation process has a prelexical locus. Although some have argued that the effect is due to general, low-level acoustic contrast effects (Lotto & Kluender, 1998; Lotto, Kluender & Holt, 1997), recent research suggests that it reflects speech-specific processes (Fowler, Brown, & Mann, 2000).

Elman and McClelland (1988), in a well-known paper, appeared to demonstrate lexical involvement in fricative-stop compensation for coarticulation, as predicted by TRACE. Pitt and McQueen (1998), however, have shown that this effect was likely to be due to transitional probability differences (between the vowels and fricatives in the different lexical contexts). They also demonstrated a dissociation between lexical involvement in decisions about the fricative sounds and, simultaneously, no lexical involvement in compensation for coarticulation. This dissociation is predicted by Merge, but not by TRACE.

Given the importance of compensation for coarticulation to the feedback debate, it perhaps comes as no surprise that five talks at SWAP were on this topic. Three of these presentations described the beginnings of ongoing research projects. *Fowler and Brancazio* proposed an ingenious test of the Merge model, where lexical information is provided by visual (lip-read) speech. Does this visual information also influence compensation for coarticulation? *Tanenhaus, Magnuson, McMurray and Aslin* proposed another ingenious test, this time using an artificial language-learning paradigm to manipulate ‘‘lexical status’’ and transitional probabilities in nonsense contexts, and then to measure, using the eye-tracking paradigm, whether these contexts influence compensation for coarticulation. *Kingston* sought to examine the effect of compensation for coarticulation on both stop identification and stop discrimination. The important issue here is whether the contexts (in Kingston’s case, the liquids [l] and [r]) influence perceptual sensitivity.

Vroomen and de Gelder, like Fowler and Brancazio, examine the influence of visual speech on fricative-stop compensation for coarticulation. They found, in a parallel to Pitt and McQueen (1998), a dissociation: listeners used lipread information in identifying fricatives, but did not appear to use this information to identify the following stops. In contrast to the recent results of Fowler et al. (2000), lipread information did not appear to modulate the compensation for coarticulation mechanism. While this particular contradiction remains to be resolved, it seems clear that the study of the integration of auditory and visual speech cues can be valuable not only in the feedback debate, but perhaps also in other domains of spoken word recognition. For example, do visual cues to speech sounds carry as much weight as auditory cues in the activation of lexical hypotheses?

Samuel also examines lexical involvement in compensation for coarticulation. He describes recent results, which, like those of Elman and McClelland (1988), appear to show an influence of lexical context (in fricative-final words) on the identification of following stops. Unlike Elman and McClelland’s materials, however, the new materials are controlled for the transitional probabilities between the word-final fricatives and the preceding vowels. These results, always assuming that they do not, like those of Elman and McClelland, prove to be open to an explanation which does not require lexical feedback, appear to challenge Merge’s assumption that there is no feedback from the lexicon to prelexical levels.

Another potential measure of prelexical processing, and thus a possible tool for testing claims about feedback, is the selective adaptation effect. When listeners hear multiple tokens of a speech sound, and are then required to label a continuum of sounds between the adaptor phoneme and another phoneme, they tend to label ambiguous sounds less often as the adaptor than as the other phoneme (Eimas & Corbit, 1973). Samuel (1997) has demonstrated lexical influences on selective adaptation, using words with noise-replaced phonemes as adaptors, and **Samuel** describes similar effects using words with ambiguous phonemes as adaptors. He argues that this is further evidence for feedback. As we have discussed elsewhere (Norris et al., 2000), however, we are not yet convinced that the locus of the adaptation effect with noise-replaced phonemes is purely prelexical. The same argument applies to adaptation with ambiguous phonemes. Selective adaptation may thus not provide a critical test of feedback.

One other paper also discussed lexical effects and the feedback issue. **Pitt and Shoaf** describe the Verbal Transformation Effect (VTE): when listeners hear the same word repeated very many times at a rapid rate, the word tends to be perceived as other words. They report lexical effects in the VTE, and examine their cause. These lexical effects could perhaps be used to evaluate whether there is feedback in spoken word recognition. This issue is certainly not settled, so new approaches to the problem are certainly to be welcomed. As we argued in Norris et al. (2000), feedback is unnecessary since it cannot benefit word recognition. Since feedback is useless for word recognition, convincing evidence is required to show that it does in fact exist in the human speech recognition system. We were delighted to see so many people at SWAP taking up this challenge.

In addition to these developments, further progress may be made by considering the feedback issue in the context of the other issues discussed here. For example, consider the fact that the argument for feedback depends on the assumption that there are phonemic prelexical representations, from which explicit phonemic decisions are made. In a radically episodic model, for example, with no prelexical level of representation, or in a model where the prelexical representations are featural, there are no phoneme units that lexical activation could be fed back to. If phoneme representations have to be constructed in the context of such models to explain lexical effects on phonemic decision-making, then those representations are, by definition, not part of the lexical access system, and any flow of information to those representations therefore cannot entail feedback. More generally, the feedback question hinges on assumptions about a hierarchy of representations including, minimally, prelexical and lexical levels. In the context of a highly distributed model, with no discrete representations of word forms or of sublexical units, for example, it becomes hard to define what would or would not constitute lexical involvement in prelexical processing, and feedback might not be a relevant concept. Claims about information flow and claims about representations are thus interdependent.

Higher-level processes

A major challenge facing the field is the unification of work on lexical semantics with work on lexical form. How, if at all, does the retrieval of a word's meaning impinge upon recognition of its form? Word meanings are usually considered to be involved at the level of processing where an interpretation of an utterance is achieved, that is, at a later stage of processing than that at which a parse of the lexical forms of an utterance is constructed. This is therefore another question about feedback, but now at the interface between interpretative and lexical-form processing rather than the interface between lexical-form and prelexical processing. Several presentations addressed this issue at SWAP, in a number of different ways.

Gaskell asks whether sentential context can influence the identification of potentially assimilated forms of words. For example, the [rʌm] in [... rʌmpɪks ...] could either be the word *rum* or an assimilated token of the word *run*. Gaskell shows that *run* is activated by this input, but only in the context of a preceding sentence about running. *Nooteboom, Janse, Quené and te Riele*, in a rhyme detection experiment, showed that sentence context can have a very rapid effect on the activation of words. Words rhyming with a prespecified cue word were detected faster when they were predictable from the sentence context. *Van den Brink, Brown and Hagoort*, using an Event-Related Potential (ERP) measure, also demonstrated rapid use of sentential context during spoken word recognition. ERP waveforms diverged 150 ms after word onset for words in congruent versus incongruent contexts. *Mauth* showed that Dutch listeners in a phonetic categorisation task tend to interpret an ambiguous word-final sound on a [xat]-[xak] continuum as [t], thus forming the inflected Dutch verb *gaat* (goes), rather than the nonword *gaak*. But again this effect was only observed in a sentence context.

Bard, Sotillo, Kelly, and Aylett, finally, review evidence suggesting that word recognition requires the use, not only of acoustic-phonetic and lexical information, but also discourse information. They argue that there is so much variability in casual continuous speech, caused by phenomena such as deaccenting, vowel reduction, consonant deletion, and assimilation, that there is no simple way to predict or constrain these phonological changes. They therefore suggest that one way listeners deal with this variability is that they use their knowledge about the ongoing discourse to resolve the ambiguities in the signal.

One could interpret all of these results as evidence of feedback from interpretative levels of processing to the level of word-form processing. Sentential context could thus exert a direct effect on the activation of lexical candidates. But there is an alternative explanation, akin to that offered by the Merge model to account for lexical context effects on phonemic processing. This is that sentence context has its influence by feeding information forward to a level of perceptual decision-making, rather than back to the representations of word-forms initially activated by the speech signal. This is not a new idea. Indeed, several authors have argued that sentence context does not influence the process of speech encoding, but instead influences perceptual decision-making (Van Alphen & McQueen, in press; Connine, 1987; Samuel, 1981). A common view is that sentential context can influence lexical selection (the choice from a set of activated candidate words) but not which words are activated (which is determined by the signal alone; see, e.g., Marslen-Wilson, 1987; Zwitserlood, 1989). But it remains unclear whether feedback is required to explain effects of context on lexical selection. Feedback is involved only if the context modulates the activation of word-form representations. If instead context influences some other representation of that word, such as its morphemic representation, or has its influence only on perceptual decisions, or if one considers that lexical selection occurs at the same level of processing as the construction of the interpretation of the utterance, then there is no need to postulate feedback to explain sentence context effects. What might be very useful in this debate would be an experiment analogous to those on compensation for coarticulation—that is, an experiment which examined whether sentence context influences a process which is an integral part of the process of word-form activation.

CONCLUSIONS

One of the benefits of a workshop conference is that presenters are encouraged (if not forced by persistent questioners) to re-evaluate their work in the context of other presentations. While written research papers do of course discuss new empirical results in the context of the relevant literature, what counts as relevant enough to be included in a particular article will in part be determined by space limitations. A workshop like SWAP thus provides a potentially much broader challenge than that thrown up by the requirement to write a well-integrated research article. Not many articles on mismatching information in lexical access, for example, will discuss the representation of morphologically complex words. And yet this kind of comparison of seemingly rather disparate issues raises important questions. What, for example, is accessed when a mispronounced word is heard? If both a representation of the word's phonological form and a representation of its morphological structure are activated, is the activation of both representations modulated by the mismatching information? Does the activation of the morphemic representation depend upon some criterial level of activation of the phonological representation?

We have therefore attempted in this review to draw attention to the connections between different issues in spoken word recognition. Assumptions that are taken for granted in one domain may well be those that are directly tested in another. This of course makes good sense, since meaningful experimental questions can only be asked by carving up the topic into smaller parts. But a look at the connections between sub-topics may well reveal new research questions, or new ways of looking at old questions. There are many more links to be drawn than we could discuss here; we encourage those reading the papers in this special issue to continue to draw them.

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*, 419–439.

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*, 551–563.

Connine, C.M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language*, *26*, 527–538.

Connine, C.M., Blasko, D.G., & Wang, J. (1994). Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context. *Perception and Psychophysics*, *56*, 624–636.

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

Cutler, A. (in press). Listening to a second language through the ears of a first. *Interpreting.*

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.

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

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*, 143–165.

Fowler, C.A., Brown, J.M., & Mann, V.A. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 877–888.

Gaskell, G., & Marslen-Wilson, W.D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 144–158.

Gaskell, M.G., & Marslen-Wilson, W.D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 380–396.

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

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

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

Goldinger, S.D., Luce, P.A., Pisoni, D.B., & Marcario, J.K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1211–1238.

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*, 344–359.

Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, *38*, 245–294.

Lotto, A., & Kluender, K. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception and Psychophysics*, *60*, 602–619.

Lotto, A., Kluender, K., & Holt, L. (1997). Perceptual compensation for coarticulation by Japanese quail (*coturnix coturnix japonica*). *Journal of the Acoustical Society of America*, *102*, 1134–1140.

Mann, V.A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception and Psychophysics*, *28*, 407–412.

Mann, V.A., & Repp, B.H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, *69*, 548–558.

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

Marslen-Wilson, W.D. (1990). Activation, competition, and frequency in lexical access. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 148–172). Cambridge, MA: MIT Press.

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

Marslen-Wilson, W., Nix, A., & Gaskell, G. (1995). Phonological variation in lexical access: Abstractness, inference and English place assimilation. *Language and Cognitive Processes*, *10*, 285–308.

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

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychological Review*, *101*, 653–675.

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

McQueen, J.M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, *39*, 21–46.

McQueen, J.M., & Cutler, A. (1997). Cognitive processes in speech perception. In W.J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 566–585). Oxford: Blackwell.

McQueen, J.M., & Cutler, A. (1998). Morphology in word recognition. In A. Spencer & A.M. Zwicky (Eds.), *The handbook of morphology* (pp. 406–427). Oxford: Blackwell.

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*, 621–638.

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*, 1363–1389.

Mehler, J. (1981). The role of syllables in speech processing: Infant and adult data. *Philosophical Transactions of the Royal Society*, *Series B*, *295*, 333–352.

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

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*, 369–378.

Miller, J.L., & Volaitis, L.E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception and Psychophysics*, *46*, 505–512.

Monsell, S., & Hirsh, K.W. (1998). Competitor priming in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1495–1520.

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

Norris, D., McQueen, J.M., & Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1209–1228.

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

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

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, *32*, 258–278.

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

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

Pitt, M.A., & McQueen, J.M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, *39*, 347–370.

Radeau, M., Morais, J., & Segui, J. (1995). Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1297–1311.

Samuel, A.G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110*, 474–494.

Samuel, A.G. (1997) Lexical activation produces potent phonemic percepts. *Cognitive Psychology*, *32*, 97–127.

Schreuder, R., & Baayen, R.H. (1995). Modeling morphological processing. In L.B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–154). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Shillcock, R. C. (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., & Hamburger, M.B. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1239–1250.

Swinney, D.A. (1981). Lexical processing during sentence comprehension: Effects of higher order constraints and implications for representation. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech* (pp. 201–209). Amsterdam: North-Holland.

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

Van Alphen, P., & McQueen, J.M. (in press). The time-limited influence of sentential context on function word identification. *Journal of Experimental Psychology: Human Perception and Performance*, *27*.

Van der Lugt, A. (in press). The use of sequential probabilities in the segmentation of speech. *Perception and Psychophysics*.

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

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

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 98–108.

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

Wallace, W.P., Stewart, M.T., & Malone, C.P. (1995). Recognition memory errors produced by implicit activation of word candidates during the processing of spoken words. *Journal of Memory and Language*, *34*, 417–439.

Wallace, W.P., Stewart, M. T., Sherman, H.L., & Mellor, M. (1995). False positives in recognition memory produced by cohort activation. *Cognition*, *55*, 85–113.

Yip, M.C.W. (2000). Recognition of spoken words in continuous speech: Effects of transitional probability. In B. Yuan, T. Huang, & X. Tang (Eds.), *Proceedings of ICSLP 2000*, Vol. 3 (pp. 758–761). Beijing: China Military Friendship Publish.

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

Zwitserlood, P., & Schriefers, H. (1995). Effects of sensory information and processing time in spoken-word recognition. *Language and Cognitive Processes*, *10*, 121–136.