In *Phonetics and Phonology in Language Comprehension and Production. Differences and Similarities*. N. O. Schiller and A. S. Meyer (Eds.) (2003). Berlin: Mouton de Gruyter (pp. 39-78).

Continuity and gradedness in speech processing

James M. McQueen, Delphine Dahan and Anne Cutler

1. Introduction

Spoken language comprehension is a decoding process. The talker's message is encoded in the physical speech signal in complex patterns of acoustic energy, in the three dimensions of amplitude, frequency and time. The listener's task is to extract the underlying message from this code. The key to cracking the code is the listener's prior knowledge about the phonological form of words. This phonological information, however it may be stored in lexical memory, is the only means by which listeners can extract a message from the hissing, humming, chirping stream of sounds that impinges on their ears when someone speaks.

In this chapter, we review what is currently known about the way in which listeners map the speech signal onto stored lexical knowledge. We argue that the lexical access process involves the parallel evaluation of multiple lexical hypotheses. We also argue that lexical access is continuous: There are no discrete component stages in the process; instead, information flows in cascade through the recognition system. We then describe evidence which suggests that this evaluation process is graded: Not only are there no discrete processing stages, but also the information that is passed through the system is graded rather than categorical. For example, a word is not simply either in the lexical competitor set or out of it; each word has its own variable degree of support. Recent results suggest that the scale against which the support for different words is measured has a resolution that is more fine-grained than could be captured by a purely

phonemic analysis of the speech signal. That is, subphonemic differences in the signal appear to influence lexical access.

We then discuss speech production in the light of these findings about speech comprehension. While the assumptions of continuity and gradedness in lexical access are widely held in accounts of speech decoding, both of these assumptions are questioned in some accounts of speech encoding. In a leading theory of lexical access in speech production, for example, there are discrete processing stages, and word-form representations contain only phonemic information (Levelt, Roelofs, and Meyer 1999). We discuss why the processing of phonetic and phonological information may be so different in speech encoding and in speech decoding, and suggest that the evidence on the fine-grained detail in the speech signal challenges an aspect of the Levelt et al. model.

2. Continuous multiple evaluation in speech decoding

2.1. Activation

The recognition of a word involves the parallel evaluation of many other candidate words. As speech unfolds over time, the words that are consistent with the current input are considered in parallel. The "multiple activation" metaphor is often used to describe this process: Each candidate word is considered to have a continuously varying activation value associated with it. A candidate's activation represents the amount of support from the speech signal that that word has at that time. The activation metaphor captures the idea that multiple competitor words are evaluated at the same time, and that the evaluation is incremental.

This view of speech decoding is very plausible given the nature of the task with which the listener is faced. Speech is very complex and changes rapidly over time. Processing speech incrementally can therefore reduce the memory load of storing all the acoustic details of the current signal. It also reduces delay in the recognition process: Incremental processing allows a word to be recognized as soon as it

can be (when sufficient information has accumulated to distinguish it from its competitors), rather than after the delays which could arise as different serial processing stages reach completion in a non-incremental model.

Processing speech incrementally, however, implies processing it on the basis of partial and very often ambiguous information. There are an infinite number of possible utterances that a talker might say, but a very limited inventory of sounds with which a talker can encode any one utterance. One can estimate that there are likely to be more than 1000 times as many words in any given language as there are phonemes. Phoneme inventories generally lie nearer the lower end of the range 10-100 sounds (Maddieson 1984), while a lexicon is likely to be in the range 10,000-100,000 words (depending on how one defines what a word is). The lexical-phonological space is thus very dense, with many words sharing the same sound sequences (e.g., words which begin in the same way, words which rhyme, and words which have shorter words embedded entirely within them).

The ambiguity of speech is amplified by the variability of the speech signal (even the same talker will never pronounce the same word in exactly the same way twice), and by the fact that speech is often uttered in a noisy environment. Finally, the lack of fully reliable cues to word boundaries in continuous speech (as reliable as the white spaces between written words in an English text such as this) adds to the complexity of the word-recognition problem. Not only is a given stretch of speech likely to offer support for many different words; it is also unclear a priori how many words that stretch of speech might contain, and where they might begin and end.

The price that has to be paid for the benefits of incremental processing, therefore, is that it entails the analysis of a multiply-ambiguous signal. One way to deal with this ambiguity but still achieve optimal incremental recognition consists of considering all lexical candidates compatible with the current, yet incomplete, input, and settling on one interpretation when support for this interpretation safely outweighs support for the others. Later arriving information can then help to confirm or disconfirm earlier interpretations of the input. This processing is embodied in the assumptions of multiple

activation and competition, shared by all current models of spoken-word recognition.

It is important to point out, however, that the activation of a word can mean different things to different theorists. Some theories assume that a word corresponds to an abstract representation of the form of a word, itself associated with a representation or representations corresponding to that word's meaning. This form representation is a category that abstracts from all variations in the acoustic realization of a word. Other theories assume that no such abstract form representation exists. All instances or episodes of that word are stored with all their acoustic details (so called traces). On such accounts, a word is a category at the meaning level that is abstracted from all its form-based instances.

There is considerable empirical support for the assumptions of multiple activation and relative evaluation of lexical candidates. Evidence for the activation of multiple candidate words as the form of a spoken word unfolds over time comes from cross-modal semantic priming experiments. These studies show that partial information in the speech signal can trigger the activation of the meaning of multiple matching candidate words. Competitors beginning at the same time are activated (e.g., in Dutch, faster responses to associates of both *kapitein*, captain, and *kapitaal*, capital, were found when listeners heard [k▶p◀◀t] than when they heard the beginning of an unrelated word; Zwitserlood 1989; see also Moss, McCormick, and Tyler 1997; Zwitserlood and Schriefers 1995). Words embedded in longer words can also be activated (e.g., in English, listeners responded more rapidly to an associate of *bone* when they heard *trombone* than when they heard an unrelated word; Shillcock 1990, but see also Luce and Lyons 1999, Swinney 1981, and Vroomen and de Gelder 1997). Furthermore, words straddling word boundaries in the input are also activated. In English, faster responses to associates of both *lips* and *tulips*, for example, were found when listeners heard *two lips* than in a control condition (Gow and Gordon 1995). Likewise, in Italian, responses to an associate of *visite*, visits, for example, were faster when listeners heard *visi tediati*, bored faces, than in a control condition (Tabossi, Burani, and Scott 1995).

In recognition memory experiments, false positive errors have been found on words which had not been presented earlier in the experiment but which began in the same way as words which had been presented earlier (Wallace, Stewart, and Malone 1995; Wallace et al. 1995, 1998). These errors suggest that the non-studied words were indeed activated when the studied words were heard.

Eye-tracking experiments, where participants' fixations to pictures on a computer screen are collected while they are auditorily instructed to click on one of the pictures, have also provided evidence for multiple-candidate activation. As the name of the target picture unfolds over time, participants make more fixations to pictures with names compatible with the available spoken information (e.g., looks to picture of a beetle when the initial sounds of *beaker* are heard) than to unrelated pictures (Allopenna, Magnuson, and Tanenhaus 1998; see also Tanenhaus et al. 2000).

The meanings of word candidates are thus available before the word that was actually heard can be unambiguously identified. This fact has important consequences for theories of spoken-word recognition. It demonstrates that semantic representations of words can be activated when their corresponding form representations have been activated but before the support for one particular form has outweighed the support for other forms. The activation process is thus continuous, rather than staged, between form- and meaning-representation levels.

2.2. Competition

As multiple candidates are activated by partial spoken input, the degree of evidence for each of them is evaluated with respect to the other words, and this relative evaluation affects the recognition of the target word. This lexical competition process has considerable empirical support. Multiple lexical activation and evaluation can be inferred from the effects of manipulating the lexical neighborhood density of target words (the number and frequency of similar sounding words). It is harder to recognize a word in a dense neighborhood than in a sparse neighborhood because of stronger inter-word competition

in the denser neighborhood (Cluff and Luce 1990; Luce 1986; Luce and Large 2001; Vitevitch and Luce 1998, 1999).

The number of competitors beginning at a different point in the input than the target word also influences ease of target recognition. For example, recognizing a word embedded in a longer nonsense word tends to be harder when the nonsense word contains a sequence consistent with many other words than when that sequence is consistent with fewer words (Norris, McQueen, and Cutler 1995; Vroomen and de Gelder 1995).

Competition between specific candidate words has also been observed. Listeners find it harder to spot words embedded in the onsets of longer words (like *sack* in [s◀kr▶ f], the beginning of *sacrifice*) than in matched sequences which are not word onsets (like [s◀kr▶ k]; McQueen, Norris, and Cutler 1994). This kind of competition also occurs when the target and competitor begin at different points in the signal (e.g., spotting *mess* in [d▶ m◀s], the beginning of *domestic*, is harder than in the nonword onset [n▶ m◀s]; McQueen et al. 1994).

The effects of the competition process extend over time. In priming paradigms, responses to target words tend to be slower when they are preceded by phonologically related prime words than when they are preceded by unrelated words. This suggests not only that target words are activated when related primes are heard, and that they lose the competition process, but also that this has negative consequences for the subsequent processing of those targets. Inhibitory effects have been found in phonetic priming experiments (in which target words are preceded by primes which share phonetic features but no phonemes with the targets; Goldinger et al. 1992; Luce et al. 2000) and in phonological priming experiments (where primes and targets share onset phonemes; Monsell and Hirsh 1998; Slowiaczek and Hamburger 1992). Note, however, that inhibitory effects in phonological priming are sometimes weak or absent (see, e.g., Praamstra, Meyer, and Levelt 1994, and Radeau, Morais, and Segui 1995). This may be because the inhibitory effects are concealed by strategic factors (see, e.g., Monsell and Hirsh 1998, for discussion).

Models of spoken word recognition like the Cohort model (Marslen-Wilson 1987, 1993), TRACE (McClelland and Elman 1986),

1986), Shortlist (Norris 1994), the Distributed Cohort Model (DCM; Gaskell and Marslen-Wilson 1997), the Neighborhood Activation Model (NAM; Luce and Pisoni 1998) and PARSYN (Luce et al. 2000) differ in very many respects. They all have one thing in common, however. They all share the assumption that, as a listener hears a section of speech, the words that are consistent with that input are considered in parallel, with the respective evidence for each word evaluated relative to the other words.

This relative-evaluation algorithm is implemented in different ways in these models. One way to implement the relative evaluation algorithm is to allow lexical representations to compete directly and actively with one another (as in TRACE, Shortlist and PARSYN). Two other implementations have been proposed. First, as in the NAM and the Cohort model, relative evaluation can occur at a decision stage, where differential degrees of support for candidates are passively compared (i.e., unlike in active competition models, the evaluation has no influence on the activation of competitors). Second, relative evaluation can be achieved via the indirect form of competition or interference that occurs as a connectionist model with highly distributed lexical representations generates a particular activation pattern (as in the DCM).

Although each of these implementations can account for many effects, the available data impose some constraints on the choice between them. A recent eye-tracking study (Dahan et al. 2001b) found effects of a competitor's interference on the target's activation before the complete name of the target had been heard and processed. These data suggest that the evaluation of a candidate's activation proportional to its competitors' activation must take place in a continuous manner. These results thus challenge competition implementations in which relative evaluation only occurs at a discrete stage of processing.

Experiments showing that competition can take place between words beginning at different points in the speech stream (e.g., McQueen et al. 1994) support the implementation of competition via direct links between candidates, and call the plausibility of models with decision-stage competition into question. Direct competition provides a more efficient means than decision-stage competition by

which words that do not all start at the same point in the input can be evaluated relative to one another (see McQueen et al. 1995, for further discussion).

2.3. Summary

Speech decoding thus appears to involve the parallel activation of multiple lexical hypotheses, and the relative evaluation of those hypotheses. This process is incremental and continuous. Words are activated even when they match the signal only partially (e.g., when a given stretch of speech can be continued in different ways, a number of different lexical paths will be considered). Furthermore, activation does not stop at the level of word-form representations; it continues through to the semantic level, such that the meanings of competitors can be activated before the word that was actually present in the input can be fully identified. Information thus flows in cascade through the recognition system, with no serial sub-stages in the process.

3. Gradedness in speech decoding

How is lexical activation modulated during the comprehension process? There are two inter-related aspects to this question. The first concerns the parameters which determine whether a given word should enter or leave the competitor set. The second concerns the metric which is used to compute the goodness of fit of any given word to the input. We will argue that words are not activated in an all-or-none fashion. Instead, lexical representations are activated in a graded way. Activation levels reflect the degree of support the speech signal provides for particular words; they change continuously over time as the information in the signal changes. We will also argue that a phoneme-based evaluation metric in the computation of lexical goodness of fit is insufficient. Finer-grained information than can be captured by a phonemic transcription modulates lexical activation.

3.1. The determinants of lexical activation

Our review of the evidence for multiple activation of candidate words and for competition between those candidates suggests that all words that are consistent with the information in the speech signal are considered, and that partial information is sufficient for lexical activation. What are the constraints on this process, however? How much matching material does there have to be in the signal to cause activation? The evidence suggests that the position of the matching information in the word, the length of that word, and the number of lexical competitors it has are all determinants of its activation. The frequency of occurrence of words also plays a role in lexical activation (see, e.g., Dahan, Magnuson, and Tanenhaus 2001a; Luce and Pisoni 1998).

The recognition system appears to be quite intolerant of mismatching information in word-initial position. Marslen-Wilson and Zwitserlood (1989) found, in a Dutch cross-modal priming experiment, that responses for example to *bij*, bee, an associate of *honing*, honey, were faster after listeners had heard the prime *honing* than after they had heard an unrelated prime word. But there was no overall priming effect when the prime rhymed with the base word and indeed shared all segments with the base word except for its initial phoneme, neither when it was another word (*woning*, dwelling) nor when it was a nonword (*foning*). This result suggests that a very strict criterion may be used to determine whether a word is considered as a candidate: Mismatch of one phoneme in word-initial position may be sufficient to block lexical access.

The nature of the difference between the prime words and the base words seems to be critical, however (Connine, Blasko, and Titone 1993). Connine et al. observed cross-modal associative priming for base word primes (e.g., *service* as prime, *tennis* as target) and a weaker priming effect for nonword primes differing from the base words in only one or two features (*zervice-tennis*), but no reliable effect for nonword primes differing from the base words on more than two features (*gervice-tennis*). These featural distances were computed from the number of articulatory features that the two phonemes share (Jakobson, Fant, and Halle 1952). Marslen-Wilson,

Moss, and van Halen (1996) observed a similar pattern of results, using intra-modal (auditory-auditory) priming in Dutch: facilitation was strongest for target words preceded by associates (e.g., *tomaat-rood*, tomato-red, and *tabak-pijp*, tobacco-pipe), weaker when the prime was a nonword which differed by only one feature on its initial segment from the base word (*pomaat-rood*), and weaker still when the difference involved two or more features (*nabak-pijp*). In contrast to the Connine et al. study, however, the difference between the two mismatch conditions was not significant.

Featural distance manipulations have also been carried out using the phoneme monitoring task. The logic here is that phoneme monitoring response latencies reflect degree of lexical activation. Lexical influences on phonemic decision-making have been modeled either as the consequence of top-down feedback from the lexicon on prelexical phoneme representations (as in TRACE), or as a consequence of a feedforward process from the lexicon to a level of processing where explicit phoneme decisions are made (as in the Merge model, Norris, McQueen, and Cutler 2000). On either the feedback or feedforward account, if a word is more strongly activated, it will facilitate phonemic decision-making more strongly. Connine et al. (1997) asked listeners to detect the final /t/, for example, in the base word *cabinet*, a minimal mismatch nonword *gabinet* (one feature change on the initial phoneme), a maximal mismatch nonword *mabinet* (many features changed) and a control nonword *shuffinet*. Phoneme monitoring latencies were fastest for targets in base words, slower for targets in minimal mismatch nonwords, slower still for targets in maximal mismatch nonwords, and slowest of all for targets in control nonwords. These results are thus consistent with the claim that lexical activation does not depend on a perfect phonemic match in word-initial position.

Evidence of activation of rhyming words with initial mismatch has also been observed using the eye-tracking paradigm (e.g., listeners look at a picture of a speaker when they hear *beaker*; Allopenna et al. 1998). The tendency to look at pictures of rhyming competitors is, however, weaker than the tendency to look at pictures of competitors which begin in the same way as the spoken word (e.g., looks at a beetle given *beaker*; Allopenna et al. 1998). This finding reflects a

general tendency that competitors which begin in the same way as target words are more strongly activated (in spite of perhaps greater mismatch) than rhyme competitors (compare, for example, the results of Zwitserlood, 1989, which gave evidence of activation of *kapitaal* when the initial sounds of *kapitein* were heard, with those of Zwitserlood and Marslen-Wilson, 1989, where there was apparently no activation of *honing* by *woning*). This tendency is likely to be due to the relative position of the mismatching information, to the temporal properties of speech, and to lexical competition. In the Allopenna et al. example, *beetle* may be just as plausible a candidate as *beaker* early in the *beaker* sequence, so for at least some time they are likely to be equally strong competitors. But *speaker* will always be at a disadvantage because of its initial mismatch; it can therefore never become as strong a competitor as the target *beaker*.

Recent support for this view of the dynamics of lexical activation comes from a phoneme monitoring study. Frauenfelder, Scholten, and Content (2001) found evidence of lexical activation of long French words when the words were distorted by a single feature change on their initial phoneme (e.g., *vocabulaire*, vocabulary, produced as *focabulaire*). Responses to target phonemes were faster in these distorted words than in control nonwords, but only when the target phoneme was word-final (i.e., according to Frauenfelder et al., only when enough time had elapsed for the positive evidence later in the word to override the negative effects of the early mismatch).

Frauenfelder et al. (2001) also examined the impact of mismatch occurring later in the input. There was no evidence of activation of *vocabulaire* given *vocabunaire*, for example (i.e., responses to target phonemes in these distorted words, e.g., the final /r/ of *vocabunaire*, were no faster than in control nonwords). This result suggests that the activation of words which have already been activated (given their initial perfect match) is strongly reduced when mismatching material is heard. Soto-Faraco, Sebastián-Gallés, and Cutler (2001) reached a similar conclusion on the basis of a series of cross-modal fragment-priming experiments. Spanish listeners' responses to *abandano*, abandonment, for example, were faster, relative to a control condition, if they had just heard the matching fragment *aban*, and slower if they had just heard the mismatching fragment *abun*, the

onset of *abundancia*, abundance. Soto-Faraco et al. argue that this inhibitory effect reflects the joint influence of the mismatching information and lexical competition (e.g., inhibition of *abandano* by *abundancia*).

It appears, therefore, that polysyllabic words which begin in a different way from what was actually heard can be activated in spite of the initial mismatch, and that long words, once activated, are penalized when a later-occurring mismatch occurs. Shorter (i.e., monosyllabic) words, however, appear to be less strongly activated when they mismatch with the input. Research on the effects of initial mismatch with monosyllabic words has suggested that robust activation of any particular monosyllabic candidate depends on how many words are close matches to the signal. Milberg, Blumstein, and Dworetzky (1988) observed intra-modal priming on lexical decisions to targets preceded by nonwords differing from associates of those targets by one or more features on the initial phoneme (e.g., responses to *dog* were faster after the prime *gat* than after an unrelated prime, presumably due to the activation of *cat*). But this effect may depend on the fact that *gat* is itself not a word, leaving *cat* as the best match to the signal. When there is a strong alternative candidate word, however, there may be no activation of mismatching words. Gow (2001), for example, found no evidence of activation (in a cross-modal form-priming experiment) of monosyllabic words like *guns* when listeners heard close lexical competitors like *buns*.

Connine, Blasko, and Wang (1994), also using a cross-modal priming task, presented listeners with auditory stimuli in which the initial sound was ambiguous between two different phonemes, such as a sound half way between /b/ and /p/, and in which both interpretations of the sequence was a word (e.g., [?◀◀g], consistent with both *big* and *pig*). Facilitative priming was observed on responses to visually presented associates of both these words (e.g., *little* and *hog*). This suggests that the lexical access process is more tolerant of mismatch when the input differs from a word by less than one phoneme. But this effect was not replicated by Marslen-Wilson et al. (1996): There was no facilitation of responses to *wood*, for example, after hearing [?l◀🔇k], which is consistent with both *plank* and *blank*. Marslen-Wilson et al. found a priming effect, however, when only

one of the endpoints was a word: Responses to *job*, for example, an associate of *task*, were facilitated when [?►sk] was heard, where [?] was ambiguous between /t/ and /d/ and *dask* is a nonword. It therefore again appears to be the case that degree of lexical activation of mismatching words depends on the lexical competitor environment.

Finally, it is important to note that tolerance to mismatching information is modulated by the phonological context. A body of research has examined how the recognition system deals with the variation in the signal caused by phonological processes such as assimilation (see, e.g., Coenen, Zwitserlood, and Boelte 2001; Gaskell and Marslen-Wilson 1996, 1998, 2001; Gow 2001; Marslen-Wilson, Nix, and Gaskell 1995). These studies have shown that words can be recognized in spite of the phonemic changes caused by assimilation, but only when those changes are contextually appropriate. Thus, for example, the word *night* is activated given the input [na◄◄p], but only if it appears in a context which licenses the assimilation of place of articulation of the final coronal consonant /t/ to bilabial [p], as in *night bus*.

The evidence on the effect of mismatch on lexical activation thus suggests that the lexical access process is not highly tolerant of mismatching information. Words that mismatch with the signal by more than a phoneme are unlikely to be considered as serious candidates if the mismatching information is at or near the beginning of the word, or rapidly rejected as plausible candidates if the mismatch occurs later in the word. The position of the mismatch, the length of the word, the number of lexical competitors, and the phonological context all appear to influence the tolerance of the system. The pattern of results on this issue is complex, however, and further work will be required to establish how these different factors interact in determining lexical activation. Nevertheless, it seems clear that, as speech unfolds over time, candidate words become, remain or cease to be active depending both on the amount of bottom-up support they have and on the amount of support other words have. When the available evidence does not clearly favor one word, all plausible candidates remain activated, but as soon as disambiguating information is available, the system appears to settle rapidly on the winning candidate

and to reject the losers (McQueen, Norris, and Cutler 1999; Norris et al. 2000).

3.2. Graded goodness of fit

This view of the dynamics of the lexical access process suggests that each word's activation reflects its moment-by-moment goodness of fit with the available input. What metric is used in this computation? One possibility is that the degree of activation of a word reflects the activation of its components. The simplest metric that could be used to compute a word's activation would be to count the number of components of that word which are consistent with the signal. Word activation could then vary as a function of the number of matching components. This metric would of course depend on a level of processing, prior to lexical access, at which those components would be recognized, and on specification of what those components are.

Theories of speech decoding which assume abstract lexical form representations often also assume prelexical abstract representations. The minimal difference between one word and any other word in the listener's language must be a phonemic difference (a word's nearest lexical neighbor cannot differ from that word by less than one phoneme). One obvious candidate for the abstract representations that exist at the prelexical level is therefore the phoneme, as indeed is instantiated in Shortlist and TRACE. Other theories have questioned the benefits of an intermediate analysis of the signal, since this may discard useful acoustic information. For these models, the degree of activation of a word reflects the similarity between the signal and its non-decomposable form representation (Klatt 1979, 1989), or all stored traces (Goldinger 1998). Nevertheless, the assumption of an abstract prelexical level in many models has led to a focus on the effects of abstract differences (such as phonemic differences) on lexical activation.

Could lexical activation thus depend simply on the number of matching phonemes each word has with a given input? The results of the studies on mismatch in lexical access described above suggest that lexical activation levels cannot be based on this simple metric.

Several of those studies have shown that subphonemic differences influence lexical activation. Connine et al. (1993, 1997) showed that the number of features with which a phoneme mismatches a lexically specified phoneme influences the degree of activation of that word.

Further evidence that lexical activation varies as a function of subcategorical differences comes from an auditory-auditory associative priming study by Andruski, Blumstein and Burton (1994). Lexical decision responses to *fruit*, for example, were faster when *fruit* was preceded by *pear* than when it was preceded by an unrelated word (*jet*). This priming effect was modulated, however, by the Voice Onset Time (VOT) of the initial unvoiced stop consonants of the related primes (e.g., of the [p] of *pear*). The [p] was presented in its normal form, with the VOT reduced by one-third, and the VOT reduced by two-thirds. The reductions made the VOT less like that of a prototypical [p] and more like that of the voiced counterpart [b], but both types of reduction produced tokens which were still heard as [p]. Although all three forms of the word *pear* primed *fruit*, responses were significantly slower after the more extremely edited prime had been heard than after the less extremely edited prime or the natural prime. These results suggest again that lexical activation is graded: words beginning with unvoiced stops appear to have been more weakly activated when their stops were shorter than normal than when their stops were of normal duration. Similar effects have also been observed using the identity priming task, where target words were preceded either by the same natural tokens of those words, or by tokens with shortened VOTs (Utman, Blumstein, and Burton 2000).

Yet another demonstration that lexical activation is modulated by fine-grained information in the speech signal has arisen from research on assimilation. As mentioned above, this research has shown that listeners can recognize the word *night* given the input [na◂◂p] but only if it appears in an appropriate context, such as *night bus*. Recent data suggests that the recognition system is sensitive to subphonemic cues to assimilation (Gow 2002): The [ra◂◂p] in *right berry* is not the same as the [ra◂◂p] in *ripe berry*, and this allows listeners to resolve potential lexical ambiguities caused by assimilation.

The influence of subphonemic variation on lexical activation has also been observed in studies examining the perception of words and nonwords containing mismatching acoustic-phonetic information (Dahan et al. 2001b; Marslen-Wilson and Warren 1994; McQueen et al. 1999; Streeter and Nigro 1979; Whalen 1984, 1991). Such items are created by cross-splicing sequences that originate from different words and nonwords. For example, a cross-spliced version of a nonword like *smob* can be constructed by concatenating the initial portion (up to the vowel) of the word *smog* or the nonword *smod* with the final consonant of a token of the nonword *smob* (i.e., *smo*[*g/d*] + [*smo*]*b*). Although these cross-spliced versions would both consist of the phonemic sequence /sm► b/, the vocalic portion would contain formant-transition information consistent with a velar [►] or a dental [d], which would mismatch with the final bilabial stop release burst [b]. A variety of lexical and phonetic tasks have shown that the lexical status of the cross-spliced portions of such stimuli (e.g., /sm► / from the word *smog* or the nonword *smod*) influences how much effect the mismatching coarticulatory information has (see Dahan et al. 2001b; Marslen-Wilson and Warren 1994; and McQueen et al. 1999 for further details). The interaction of the effects of subphonemic information and lexical information in tasks which probe lexical activation shows that subcategorical information influences processes at the lexical level.

All of these subphonemic effects contradict the suggestion that word activation is computed on the basis of the number of matching phonemes. More generally, they challenge the view that the prelexical stage is phonemic and discrete. If a categorical phonemic representation of the speech signal were computed at the prelexical level, and this were to occur in a serial fashion, such that a phonemic parse of the input was completed prior to lexical access, the lexical level would not be sensitive to featural differences among phonemes. One phoneme would be like any other, and lexical goodness of fit would have to be based on some measure of the number of matching phonemes. Such models can therefore be rejected.

These results, however, are consistent with models in which prelexical representations are activated in proportion to their acoustic match with the input and in which a word's activation in turn reflects

the prelexical activation pattern. Although the manipulations in the above studies have all been subcategorical, the effects can still be described phonemically. Number of mismatching features, for example, can be represented in terms of degree of support for particular phonemes. Likewise, subcategorical variation in VOT can be represented by the relative activation of voiced versus unvoiced stops, and subcategorical mismatch in cross-spliced words can modulate the amount of support for each of the phonemes involved in the splice.

These results are thus consistent with models like TRACE and Shortlist in which the prelexical representations are phonemic. In these models, information spreads continuously up to the lexical level. There is no serial stage at which an absolute phonemic categorization of the input is made prior to lexical access. TRACE is an interactive-activation model in which activation cascades continuously between representations (McClelland and Elman 1986). Although in the implemented version of Shortlist there is categorical phonemic input to the lexicon, this implementation is considered to be a mere approximation of a more continuous process (Norris 1994; Norris et al. 2000). If the degree of activation of prelexical phoneme representations can vary continuously, and this activation can spread to lexical representations, then subphonemic effects on lexical activation can be explained. The present results would of course also be consistent with models in which the prelexical representations are larger or smaller than the phoneme, so long as those representations have graded activation values and pass activation continuously up to the lexicon.

3.3. Phonemic decoding is not sufficient

Results from several recent experiments, however, impose stronger constraints on the granularity of the lexical activation process. In these new experiments, the relative activation of different words sharing the same phonemic sequences was measured. In contrast to the studies described above, therefore, the information that was varied in these new studies did not offer differential support for alternative phonemes. Instead, it provided support for one or another lexical

interpretation of the same phonemic sequence. As we describe in more detail below, there is no straightforward way to represent this kind of information in terms only of the relative degree of activation of different phonemes.

Tabossi et al. (2000) have shown in Italian that the phonetic consequences of syllabic structure on the realization of phoneme sequences affect the activation of words that match those sequences. A word that mismatched the syllabic structure of the input (e.g., *si.lenzio*, silence, when the input consisted of the syllable fragment [sil]) received less support from the input than a word that matched this structure (e.g., *sil.vestre*, silvan). The reverse was true when the input was the fragment [si.l], taken from *si.lenzio*. On a purely phonemic analysis, the fragments were identical. Nevertheless, the subphonemic difference between the two types of fragment (cued at least in part by a small but robust durational difference in the vowels) seems to have been fed forward to the lexicon, influencing word activation. It might appear that the results could be modeled in terms of the degree of activation of prelexical phonemic representations (the amount of activation of /s/, /i/ and /l/, for example). But, because the evidence does not at the same time favor alternative phonemes and thus alternative words with different phonemic transcriptions, there is no way in such an account for the lexical level to distinguish between the different types of input. Additional, non-phonemic information must therefore influence lexical activation.

Spinelli, McQueen, and Cutler (2003), in a study of liaison in French, examined the activation of vowel- and consonant-initial words (e.g., *oignon*, onion, and *rognon*, kidney) in phrases like *C'est le dernier oignon* (It's the last onion). In this context, the final [►] of *dernier* is produced and resyllabified with the following syllable, making the phrase phonemically identical to *C'est le dernier rognon*. Acoustic analyses revealed however that there were reliable durational differences in the pivotal consonants depending on the speaker's intentions (e.g., the medial [►] was longer in *dernier rognon* than in *dernier oignon*). In cross-modal identity priming experiments, only responses to the words that the speaker intended to utter were facilitated reliably. Although in both cases the information was consistent with an [►], the durational distinction appears to have

influenced the lexical level, helping listeners to retrieve the speaker's intended message.

One way of accommodating these results on syllabification and liaison is to assume that prelexical representations are allophonic variations of phonemes, rather than context-independent phonemes (as in the PARSYN model, Luce et al. 2000). Allophones are variants of phonemes that are conditioned by the context in which they occur. This context can be the position of the phoneme within a syllable (such as syllable onset or coda), or whether the syllable in which the phoneme occurs is stressed or unstressed.

Allophonic analysis of the speech signal could account for Tabossi et al.'s (2000) results (e.g., the [l] in [si.l] could be a different allophone from that in [sil], leading to differential activation of *silenzio* and *silvestre*). Likewise, the results of Spinelli et al. (2003) could be explained if liaison consonants (like the [►] *in dernier oignon*) provided more support for a syllable-final allophone while word-initial consonants (like the [►] in *dernier rognon*) provided more support for a syllable-initial allophone (note that on this account, resyllabification in liaison contexts is incomplete).

An allophonic model could also account for the effects on word activation of lexical stress or pitch-accent patterns in languages that use these prosodic factors. Lexical stress information appears to influence the degree of activation of words in languages like Spanish (Soto-Faraco et al. 2001) and Dutch (Cutler and Donselaar 2001), that is, in languages where this information is important for lexical disambiguation (see Cutler, Dahan, and Donselaar 1997, for a review). Soto-Faraco et al., for example, found an inhibitory stress mismatch effect in cross-modal fragment priming (e.g., the fragment *prinCI-*, the beginning of *prinCIpio*, which is stressed on the second syllable, produced slower responses to the visual target *principe*, which is stressed on the first syllable, *PRINcipe*, than did an unrelated fragment).

It has been suggested that lexical stress information is not used in the initial lexical access process in English because it is not useful for lexical disambiguation (Cutler 1986). More recent research, however, has shown that lexical activation is modulated by stress information in English, but less so for native speakers than for Dutch-

English bilinguals (Cooper, Cutler, and Wales submitted). Stress information may modify word activation more strongly in the bilinguals because they have had more opportunity to learn the value of this information (i.e., in processing the native language, Dutch). These results therefore support the suggestion that suprasegmental information is used to the extent that it is useful. In fixed-stress languages like French, therefore, where lexical stress information is not contrastive, this information does not appear to modulate lexical activation (Dupoux et al. 1997; Peperkamp and Dupoux 2002). A different kind of suprasegmental information, that for pitch-accent patterns in Japanese words, also appears to be used in lexical access (Cutler and Otake 1999). Again, pitch-accent information can be used for lexical disambiguation in Japanese.

Suprasegmental influences on lexical activation could be captured by models with prelexical allophonic representations. Allophonic as well as phonemic models, however, are challenged by experiments which have examined the recognition of sequences which, on either a phonemic or allophonic transcription, would be lexically ambiguous. Gow and Gordon (1995) compared the lexical activation generated by ambiguous sequences that consist of one or two words (such as *two lips* or *tulips*). Their results suggest that word activation can be modulated by the presence of acoustic cues marking word onsets in the signal. Evidence for the activation of a word embedded in the sequence (e.g., *lips*) was found in two-word sequences (e.g., *two lips*), that is, when word-onset cues may be available, but not in matched one-word sequences (e.g., *tulips*).

Recent research on the activation of words embedded in the onsets of longer words also challenges models which only encode purely segmental information (even allophonic models with context-sensitive segments). Davis, Marslen-Wilson, and Gaskell (2002) and Salverda, Dahan, and McQueen (submitted) have shown that subtle durational differences between productions of an ambiguous sequence (e.g., /p►n/ in Dutch), as either a monosyllabic word (*pan*, id.) or as the onset of a longer word (*panda*, id.), bias listeners' interpretation of the sequence in favor of the speaker's intentions. For example, Salverda et al. demonstrated that the temporary activation of the embedded word *pan*, upon hearing the carrier word *panda*,

was larger when the syllable *pan* was of a longer duration. This bias in word activation may arise from the tendency (in the sample Salverda et al. recorded, and presumably in the Dutch language in general) for monosyllabic words to be longer than equivalent sequences which form the initial portion of polysyllabic words. Salverda et al. suggest that this may be the result of segmental lengthening at the edge of prosodic domains.

3.4. Summary

There is a growing body of evidence demonstrating fine-grained modulation of the amount of support for particular words during lexical access. A model in which the number of matching phonemes between each candidate word and the input are counted is therefore not realistic. Nor are models in which there is a discrete and categorical stage of processing prior to lexical access: Just as word-form activation appears to spread continuously to word meanings, so too does the activation of prelexical representations spread to word forms.

Some results on the spread of fine-grained information to the lexicon are consistent with a variety of prelexical representational options: These are experiments in which the information could be used to evaluate the relative support in the input for different phonemic sequences. But other results do impose constraints on the nature of prelexical processing: These are experiments which have shown that there is variation in lexical activation even when only one phonemic sequence is strongly supported by the signal (i.e., where two signals with the same phonemic transcription have differential effects on the activation of words). A purely phonemic analysis would not capture allophonic variation in the speech signal (e.g., that due to syllable structure or lexical stress patterns); nevertheless, such variation does appear to influence lexical activation. Allophonic representations (i.e., one for each contextually-constrained variant of each phoneme) may therefore be preferred. But there is now evidence that lexical activation is also sensitive to differences that cannot be captured by allophonic representations.

It is not yet clear how best to model the latest data on lexical activation. One can consider two possible approaches. One is to maintain prelexical segmental representations (e.g., in terms of phonemes), but to add a parallel level of suprasegmental representations (i.e., representation of syllabic structures, lexical stress patterns, prosodic-domain boundaries, etc.). It is interesting to note here that the fine-grained information which appears to modulate lexical activation, while it can be described as subphonemic, or even suballophonic, can also be viewed as suprasegmental, in that it involves prosodic structures which are larger than the segment. On this account, word activation would be modulated by the match with both segmental and suprasegmental representations. An attractive feature of this approach is that it provides a unified account of, on the one hand, the data that could perhaps be explained by a model with prelexical allophonic representations (e.g., Spinelli et al. 2003; Soto-Faraco et al. 2001; Tabossi et al. 2000) and, on the other hand, the data which challenge allophonic models (Davis et al. 2002; Gow and Gordon 1995; Salverda et al. submitted).

The other possibility is to reject a prelexical level of processing and to assume instead that the signal is directly mapped onto lexical representations. These representations could consist of prototypes of the form of each word (as in the model proposed by Klatt 1979, 1989) or of the combination of all the traces associated with each word (as in the episodic view of Goldinger 1998). In both of these types of direct-mapping model, considerable detail about the acoustic-phonetic form of words can be stored at the lexical level. Either class of direct-mapping model could thus account for the sensitivity of the lexical access process to all the fine-grained aspects of the input, as long as those cues are word specific.

Speech decoding therefore involves the parallel graded activation of multiple candidate words. This process is continuous: There are no discrete sub-stages of processing – information flows in cascade from the prelexical to the lexical level, and from representations of word form to representations of word meaning. This process is also graded: The activation of representations at each of these levels changes continuously over time, as information from the speech signal accrues, and as different candidate words compete with each

other. Differences in degree of lexical activation appear to reflect aspects of the speech signal which cannot be captured by a purely segmental description of that signal.

4. Speech production

The view that the processing of phonological information in spoken word comprehension is continuous and graded stands in stark contrast to the view that lexical access in speech production is staged and categorical (Levelt et al. 1999). Why might the flow of information through the speech encoding process, and the nature of that information, be different from that in speech decoding? In this section, we will examine the arguments concerning these two issues in speech production, in the light of the comprehension evidence.

4.1. Flow of information in production and perception

We have argued that, in perception, activation spreads continuously from the prelexical level to the word-form level, and on up to the meaning level. But in WEAVER++ (Levelt et al. 1999; see also Roelofs, this volume), word-form production consists of two discrete stages of processing (Levelt et al. refer to a rift between the conceptual/syntactic domain and the phonological/articulatory domain). There is spread of activation involving multiple words between the conceptual and lemma levels (lemmas are syntactic representations of words which code grammatical properties like gender). There is also spread of activation among multiple representations at the word-form and phonological encoding levels. But there is a discrete step between the lemma and word-form representations: Only the form of the selected lemma is activated.

Levelt et al. motivate this assumption of seriality in two ways: first, on the theoretical grounds that it would be counterproductive to activate unnecessary phonology; and second, on empirical grounds (see, e.g., Levelt et al. 1991). More recent experiments, however, have shown that the strongest version of this seriality hypothesis is not tenable (e.g., Peterson and Savoy [1998] presented evidence for

parallel activation of the phonological forms of both members of synonym pairs like *couch-sofa*). Levelt et al. (1999) therefore suggest that multiple activation of word forms may be limited to cases where more than one lemma is selected, as when a near-synonym has to be produced under time pressure. The assumption of seriality in WEAVER++ can thus be preserved: Only word-forms for selected lemmas are activated, but there are some circumstances where more than one lemma can be selected.

In addition to the findings of Peterson and Savoy (1998), Jescheniak and Schriefers (1998) and Cutting and Ferreira (1999) have provided evidence suggesting that, at least under some circumstances, activation does flow continuously from semantics to phonology during speech production. Such results, while they can be explained by the WEAVER++ model (see Levelt et al. 1999 for discussion), also support the assumptions of continuous spreading activation in the DSMSG model (Dell 1986; Dell et al. 1997; Dell and Gordon, this volume). The DSMSG model is an interactive two-step account of lexical access in production. The first step is lemma access, the second is phonological access. During lemma access, activation spreads from semantic units to lemma units but also cascades down to phonological units. In addition to this feedforward activation, there is positive feedback from lemmas to semantic representations and from phonological representations to lemmas. The most activated lemma nodes are therefore the target and its semantically and formally related neighbors: The most highly activated lemma node is selected. The second step begins when the selected lemma node is given a large jolt of activation. Activation then spreads to the phonological units associated with the selected word, and, via the feedback connections, back to lemma and semantic representations. In contrast to WEAVER++, the DSMSG model therefore embodies an interactive rather than modular theory. But, because activation from the serially ordered jolts dominates the activation pattern, the model is only locally interactive. Activation at the semantic level has only mild effects at the phonological level and vice-versa. Nevertheless, the model correctly predicts that there are situations where there is (weak) activation of phonological representations that are not required for the utterance that is actually produced.

In the production literature, therefore, there is no consensus on whether information flow is staged or cascaded. This contrasts with the agreement that has been reached that processing operates in cascade in speech comprehension. We suggest that there are two reasons for this difference. The first is the evidence. Our review of the comprehension literature makes clear that there is overwhelming empirical support for continuous flow of information up to the meaning level. The data on cascaded processing in production are scarcer, and what results there are can be explained by a staged model (Levelt et al. 1999).

The second reason is based on arguments about the nature of speech encoding and decoding. Levelt et al. (1999) have argued that activating the phonology of an unintended word during speech production is unlikely to assist phonological encoding, and thus that it is inefficient to activate unnecessary phonology. This is a key motivation for the assumption of staged processing in WEAVER++. This is also a motivation for the activation jolts in the DSMSG model, which act to bias phonological encoding strongly in favor of the intended word. Limited cascade (i.e., only enough to activate the phonology of the intended word before lemma selection is completed) could be of some benefit, however. As Dell et al. (1997) point out, it might be helpful to have access to the phonological form of a candidate lemma to ensure that its form is available before that lemma is selected. It is to the speaker's advantage if (s)he chooses a lemma whose form will later be easy to find. Dell et al. claim that this would reduce the incidence of tip-of-the-tongue (TOT) states, where the speaker makes a commitment to a word for which the phonological form is not accessible (or only partially available).

Note, however, that this motivation for limited cascade of information in production is dependent on the assumption of feedback: for a benefit to accrue, the phonological level must be able to impact on processing at the lemma level. It is therefore not clear whether even limited cascade would be beneficial to speech production. Only models with feedback could use cascaded processing to reduce the number of TOT states. In a model without feedback, continuous flow from the lemma level to the wordform level would not help to reduce TOT states. This potential benefit of cascaded processing thus de-

pends on the additional assumption of feedback in the production system. There is no evidence which makes it necessary to make this assumption (Levelt et al. 1999; see Dell and Gordon, this volume, and Roelofs, this volume, for further discussion of the evidence for and against feedback in the production system). In the absence of good evidence in support of feedback, no strong argument can be made for the benefits of even limited cascaded processing in speech encoding. It may therefore be better to interpret the results which can be taken as evidence for cascade in production (Cutting and Ferreira 1999; Jescheniak and Schriefers 1998; Peterson and Savoy 1998) in ways which are consistent with a feedforward staged model (i.e., as Levelt et al. 1999 do). It is clear, however, that, irrespective of whether or not there is feedback in phonological encoding, widespread cascade of information right through the production system would be counterproductive because it would make speaking harder.

One might want to argue that cascaded processing in perception is also counterproductive. It might be unproductive to activate unnecessary meanings during comprehension, that is, the meanings of the candidate words which lose the lexical competition process. Would it not be efficient to restrict meaning activation to that of the winning word form? One could imagine a two-stage process: The first stage would be to select one word form on the basis of its fit with the signal; the second stage would be to access its meaning and integrate it into the context.

The listener's task, however, is to derive the message the talker intended from an infinite range of possible utterances. Furthermore, the input to phonological processing in perception is more likely to be impoverished than the input to phonological processing in production. As we have argued earlier, cascaded processing from the acoustic signal to the lexical-form level assists in the decoding process when information is missing from or not yet available in the input. Likewise, it is also useful for information to cascade from the word-form level to the meaning level in perception. Some ambiguities may be impossible to resolve on the basis of form-based information alone (e.g., those due to polysemous words). Since meaning constraints are thus sometimes essential for comprehension, it makes sense to use them as soon as possible. Higher-level information may

then also help to resolve temporary ambiguities in the signal (i.e., before disambiguating form information has been heard). Activating the candidates' meanings incrementally allows some candidates to be disfavored on the basis of the integration of their meaning within the context.

A number of studies have indeed demonstrated very early effects of context on spoken word recognition. In these studies, listeners heard spoken sentences while event-related brain potentials were recorded. As the initial sounds of a word that matched or mismatched the context were heard, but before the acoustic information allowed listeners to distinguish the word from its competitors, brain responses were shown to vary as a function of the semantic congruency of the word (Van Berkum et al. submitted; Van den Brink, Brown, and Hagoort 2001; Van Petten et al. 1999). Contextual influences occurring before sufficient acoustic information has accrued for listeners to be able to identify a word uniquely show not only that listeners have rapid access to word meanings, but also that they use this information in their evaluation of the incoming speech signal as soon as that information is available. Since the meaning level can therefore assist in the comprehension process, it is highly beneficial to pass information continuously up to that level.

This comparison of speech production with speech comprehension thus suggests that the two systems may differ with respect to how information flows during talking versus listening. There is more evidence in favor of continuous flow of information in comprehension than in production, and that which is available on production can be explained by a staged model. Even in production models with cascaded processing there are limits on the extent to which information flows between the different stages of lexical encoding. Furthermore, there are good design reasons why there may be cascaded processing in perception and staged processing in production. The nature of the task faced by the listener makes fully cascaded processing valuable in comprehension, while the nature of the task faced by the talker makes fully cascaded flow of information in production detrimental.

4.2. Granularity in production

We argued above that the lexical level of the comprehension system is sensitive to fine-grained (i.e., subphonemic) differences in the speech signal. This means that those differences must be a systematic part of the signal (i.e., they are not just noise). It therefore follows that the speech production system must produce those differences in a systematic fashion. In WEAVER++ (see Levelt et al. 1999 for details), however, a word form in production is a "bare-bones" representation, consisting of a sequence of phonemes that is unsyllabified and has no stress pattern (unless the word has an irregular stress pattern). Syllables and regular stress patterns are built on the fly during "prosodification" – a post-lexical stage of phonological encoding which computes, among other things, the syllabification of phoneme strings within phonological phrases.

One of the reasons which Levelt et al. use to motivate this assumption is that syllabification depends on surrounding context (e.g., the final /v/ of *save* is syllable final, but, at least on some accounts, will be syllable initial in the cliticized phrase *save it*). That is, the syllabification of a word is not fixed and immutable (see Levelt et al. 1999 for further discussion). In WEAVER++, therefore, there is no lexical representation of, for example, the duration of the first syllable of *panda* or of the first and only syllable of *pan*: Both syllables are simply the string of phonemes /p/, /►/, /n/. But listeners are sensitive to the durational differences between tokens of syllables like /p►n/ coming from these different contexts, and talkers tend to produce such syllables in a systematic way (Davis et al. 2002; Salverda et al. submitted). Similarly, listeners are sensitive to other fine-grained details in the speech signal (e.g., that due to syllabification, liaison or assimilation; see above), and talkers produce those details. How then might a model like WEAVER++ account for this production behavior?

One possibility is that, in the context of WEAVER++, the post-lexical prosodification procedure could be enriched with more prosodic knowledge (e.g., in the embedded word case, knowledge that results in segmental lengthening at the edge of prosodic domains). Specification of this prosodic hierarchy would run in parallel with the lexical-segmental encoding process, and these prosodic specifications could then be added to the phonological words generated dur-

ing prosodification (i.e., the same process as in the existing model, but with a richer prosodic component). During production, therefore, there would be no lexical specification of segmental duration (or any other subphonemic detail) in particular words: Durational differences would emerge as a result of the specifications provided by the prosodic hierarchy.

In perception, however, as the evidence we have summarized shows, the recognition system uses subphonemic details to modulate the activation of lexical representations. Perceptual word-form representations must therefore be sensitive in some way to these subphonemic differences. Note that this does not mean that each individual lexical representation in the perceptual system must include detailed acoustic information (e.g., durational specifications). As we suggested earlier, subphonemic acoustic information could influence the activation of prelexical suprasegmental representations, which in turn would modulate word-form representations. A word's activation would thus change as a function of a match to an abstract specification, rather than as the result of a direct match with subphonemic information. For example, the activation of *pan* could be boosted if the duration of the syllable [p►n] indicated that the word *pan* was aligned with the edge of a prosodic domain. Irrespective of how exactly subphonemic information exerts its effect on lexical activation in perception, though, it is clear that this information can arise from a production system in which that information is not coded lexically.

The proposal that fine-grained information modulates lexical activation in perception but is specified post-lexically in production is consistent with our claim that the two processing systems are fine-tuned to the different task demands of speech decoding and encoding. The listener needs to be able to recognize that a word in an utterance is a token of a particular word, and knowledge that goes beyond that word's segmental make-up can assist in that process (and indeed appears to do so). Phonetic detail assists comprehension because the more information there is for listeners to use, the easier it will be for them to distinguish one word from another.

The talker, on the other hand, needs to build an utterance given a conceptual message. While the segmental material for a given word must be stored lexically and retrieved when that word is to be spo-

ken, it could be much more efficient to complete the phonological encoding of that word in its utterance context using post-lexical rules. There is certainly no need for phonetic detail at the stage of lemma selection, where the choice between words is semantic. There might also be no need for phonetic or phonological details, beyond the bare segmental information, at the word-form stage, since words are selected on the basis of semantic specifications (i.e., spread of activation from lemmas). There may therefore be an interesting asymmetry between the lexical selection process in perception, where phonetic/phonological information is primary and semantic information is secondary, and the lexical selection process in production, where semantic information is primary and phonetic/phonological information is secondary.

The evidence on subphonemic detail in the speech signal is thus consistent with the assumptions of "bare-bones" word-form representations and post-lexical prosodification in WEAVER++. But this evidence challenges another assumption of this model: the mental syllabary (Levelt et al. 1999; Levelt and Wheeldon 1994). The output of the phonological encoding (prosodification) stage in WEAVER++ forms the input to the phonetic encoding stage, where gestural programs for articulation are generated. According to the theory, gestural programs for high-frequency syllables are stored, in precompiled form, in a syllabary (there is also a second mechanism for the phonetic encoding of infrequent or novel syllables). Talkers tend to use only a relatively small inventory of common syllables for most of their speech output (one can estimate that 500 syllables are enough to cover 80% of all English speech, Levelt et al. 1999, and 85% of all Dutch speech, Schiller et al. 1996). The syllabary is thus motivated by the idea that it would be efficient to store precompiled motor programs for a set of frequently recurring syllabic patterns.

The production data associated with the study of subphonemic effects in perception suggest, however, that each token of a syllable that a talker produces is not always the same. The [►] in the third syllable of *dernier rognon* will tend to be longer than the [►] in the third syllable of *dernier oignon* (Spinelli et al. 2003), the [l] in the second syllable of *two lips* will tend to be longer than the [l] in *tulips* (Gow and Gordon 1995), the syllable [p►n] will tend to be longer

when the talker intends the word *pan* than when the talker wants to say *panda* (Salverda et al. submitted), and so on. These findings call into question the motivation for the syllabary that speech consists by and large of a relatively small number of recurring patterns, and, more generally, cast doubt on the notion of the syllabary.

Levelt et al. (1999) point out that the fine detail of syllables can change as a consequence of coarticulation (where motor instructions for successive syllables overlap in time). But this suggestion concerns a process which occurs after the syllabary has been accessed and is therefore consistent with the syllabary hypothesis. In the cases of subphonemic differences which disambiguate words or sequences of words which would otherwise be identical, however, these differences need to be specified before phonetic encoding. That is, the phonetic encoder needs, as part of its input, a specification of the difference between the two readings of a phonemically ambiguous sequence. If the fine detail were to arise at the prosodification stage (as we have suggested it might in order to generate the segmental duration differences between *dernier rognon* and *dernier rognon*, between *two lips* and *tulips*, between *pan* and *panda*, etc.), then it would be specified before the syllabary was accessed. The same would be true if the details were coded at the lexical level in the model. But if there is only one gestural program for each syllable, syllabary access would obliterate these prespecified distinctions.

It seems clear that, in any account of speech production, there must be a means by which speakers can generate very fine-grained, but nonetheless systematic phonetic details. In WEAVER++, it appears that lexical or prosodic specifications would have to be able to modify motor programs after they have been accessed from the syllabary. This, however, seems to undermine any benefit that could be had from the storage of only a limited number of precompiled syllabic motor programs.

5. Conclusions

We have argued that speech decoding is continuous and graded. Information flows through the recognition system in cascade all the

way up to the meaning level, with no discrete processing stages. In this system, multiple words are evaluated in parallel; these candidate words compete with each other, and their activation is modulated by subphonemic detail in the speech signal. We have suggested that such a system is well suited to the demands of listening to speech.

The way that phonetic and phonological information is processed in speech encoding appears to be very different. Lexical access is a two-stage process, with, on some accounts, strict seriality, and, on other accounts, limited cascade between levels. In no current production model is there massive parallel activation of word forms. Furthermore, it appears that subphonemic detail need not be specified at the lexical level in production. Instead, this type of detail could be filled in by post-lexical rules. Again, this view appears well suited to the task demands of speaking. The evidence on subphonemic detail in the speech signal, however, calls into question the hypothesis that the phonetic encoding of speech involves a mental syllabary. This evidence therefore demands not only the development of speech decoding models which can accommodate subphonemic effects but also an account of the genesis of these effects within models of how talkers encode speech.

References

Allopenna, Paul D., James S. Magnuson and Michael K. Tanenhaus
1998 Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language* 38: 419-439.

Andruski, Jean E., Sheila E. Blumstein and Martha W. Burton
1994 The effect of subphonetic differences on lexical access. *Cognition* 52: 163-187.

Cluff, Michael S. and Paul A. Luce
1990 Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance* 16: 551-563.

Coenen, Else, Pienie Zwitserlood and Jens Bölte
2001 Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes* 16: 535-564.

Connine, Cynthia M., Dawn G. Blasko and Debra Titone

1993 Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language* 32: 193-210.

Connine, Cynthia M., Dawn G. Blasko and Jian Wang
1994 Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context. *Perception & Psychophysics* 56: 624-636.

Connine, Cynthia M., Debra Titone, Thomas Deelman and Dawn G. Blasko
1997 Similarity mapping in spoken word recognition. *Journal of Memory and Language* 37: 463-480.

Cooper, Nicole, Anne Cutler and Roger Wales
submitted Constraints of lexical stress on lexical access in English: Evidence from native and nonnative listeners.

Cutler, Anne
1986 *Forbear* is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech* 29: 201-220.

Cutler, Anne, Delphine Dahan and Wilma van Donselaar
1997 Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40: 141-201.

Cutler, Anne and Wilma van Donselaar
2001 *Voornaam* is not a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech* 44: 171-195.

Cutler, Anne and Takashi Otake
1999 Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustical Society of America* 105: 1877-1888.

Cutting, J. Cooper and Victor S. Ferreira
1999 Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25: 318-344.

Dahan, Delphine, James S. Magnuson and Michael K. Tanenhaus
2001a Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology* 42: 317-367.

Dahan, Delphine, James S. Magnuson, Michael K. Tanenhaus and Ellen M. Hogan
2001b Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes* 16: 507-534.

Davis, Matt H., William D. Marslen-Wilson and M. Gareth Gaskell
2002 Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 28: 218-244.

Dell, Gary S.
1986 A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93: 283-321.

Dell, Gary S., Myrna F. Schwartz, Nadine Martin, Eleanor M. Saffran and Deborah A. Gagnon

1997 Lexical access in aphasic and nonaphasic speakers. *Psychological Review* 104: 801-838.

Dupoux, Emmanuel, Christophe Pallier, Núria Sebastián-Gallés and Jacques Mehler
1997 A destressing deafness in French. *Journal of Memory and Language* 36: 399-421.

Frauenfelder, Uli H., Mark Scholten and Alain Content
2001 Bottom-up inhibition in lexical selection: Phonological mismatch effects in spoken word recognition. *Language and Cognitive Processes* 16: 583-607.

Gaskell, M. Gareth and William D. Marslen-Wilson
1996 Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 22: 144-158.

Gaskell, M. Gareth and William D. Marslen-Wilson
1997 Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes* 12: 613-656.

Gaskell, M. Gareth and William D. Marslen-Wilson
1998 Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 24: 380-396.

Gaskell, M. Gareth and William D. Marslen-Wilson
2001 Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language* 44: 325-349.

Goldinger, Stephen D.
1998 Echoes of echoes?: An episodic theory of lexical access. *Psychological Review* 105: 251-279.

Goldinger, Stephen D., Paul A. Luce, David B. Pisoni and Joanne K. Marcario
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, David W.
2001 Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language* 45: 133-159.

Gow, David W.
2002 Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance* 28: 163-179.

Gow, David W. and Peter C. Gordon
1995 Lexical and prelexical influences on word segmentation: evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance* 21: 344-359.

Jakobson, Roman, C. Gunnar M. Fant and Morris Halle

1952 *Preliminaries to Speech Analysis: The Distinctive Features and their Correlates*. Cambridge, MA: Massachusetts Institute of Technology Press.

Jescheniak, Jörg D. and Herbert Schriefers
1998 Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the co-activation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24: 1256-1274.

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

Klatt, Dennis H.
1989 Review of selected models of speech perception. In: William D. Marslen-Wilson (ed.), *Lexical Representation and Process*, 169-226. Cambridge, MA: Massachusetts Institute of Technology Press.

Levelt, Willem J. M., Ardi Roelofs and Antje S. Meyer
1999 A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22: 1-75.

Levelt, Willem J. M., Herbert Schriefers, Dirk Vorberg, Antje S. Meyer, Thomas Pechmann and Jaap Havinga
1991 The time course of lexical access in speech production: A study of picture naming. *Psychological Review* 98: 122-142.

Levelt, Willem J. M. and Linda R. Wheeldon
1994 Do speakers have access to a mental syllabary? *Cognition* 50: 239-269.

Luce, Paul A.
1986 Neighborhoods of Words in the Mental Lexicon (Ph.D. dissertation, Indiana University). In: *Research on Speech Perception*, Technical Report No. 6, Speech Research Laboratory, Department of Psychology, Indiana University.

Luce, Paul A., Stephen D. Goldinger, Edward T. Auer and Michael S. Vitevitch
2000 Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics* 62: 615-625.

Luce, Paul A. and Nathan R. Large
2001 Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes* 16: 565-581.

Luce, Paul A. and Emily A. Lyons
1999 Processing lexically embedded spoken words. *Journal of Experimental Psychology: Human Perception and Performance* 25: 174-183.

Luce, Paul A. and David B. Pisoni
1998 Recognizing spoken words: The Neighborhood Activation Model. *Ear and Hearing* 19: 1-36.

Maddieson, Ian
1984 *Patterns of Sounds*. Cambridge: Cambridge University Press.

Marslen-Wilson, William D.
1987 Functional parallelism in spoken word-recognition. *Cognition* 25: 71-102.
Marslen-Wilson, William D.
1993 Issues of process and representation in lexical access. In: Gerry T. M. Altmann and Richard Shillcock (eds.), *Cognitive Models of Speech Processing: The Second Sperlonga Meeting*, 187-210. Hillsdale, NJ: Erlbaum.
Marslen-Wilson, William D., Helen E. Moss and Stef van Halen
1996 Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 22: 1376-1392.
Marslen-Wilson, William D., Andy Nix and M. Gareth Gaskell
1995 Phonological variation in lexical access: Abstractness, inference and English place assimilation. *Language and Cognitive Processes* 10: 285-308.
Marslen-Wilson, William D. and Paul Warren
1994 Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review* 101: 653-675.
Marslen-Wilson, William D. and Pienie Zwitserlood
1989 Accessing spoken words: the importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance* 15: 576-585.
McClelland, James L. and Jeffrey L. Elman
1986 The TRACE model of speech perception. *Cognitive Psychology* 10: 1-86.
McQueen, James M., Anne Cutler, Ted Briscoe and Dennis Norris
1995 Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes* 10: 309-331.
McQueen, James M., Dennis Norris and Anne Cutler
1994 Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20: 621-638.
McQueen, James M., Dennis Norris and Anne Cutler
1999 Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance* 25: 1363-1389.
Milberg, William, Sheila E. Blumstein and Barbara Dworetzky
1988 Phonological factors in lexical access: Evidence from an auditory lexical decision task. *Bulletin of the Psychonomic Society* 26: 305-308.
Monsell, Stephen and Katherine W. Hirsh

1998 Competitor priming in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24: 1495-1520.

Moss, Helen E., Samantha F. McCormick and Lorraine K. Tyler
1997 The time course of activation of semantic information during spoken word recognition. *Language and Cognitive Processes* 10: 121-136.

Norris, Dennis
1994 Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52: 189-234.

Norris, Dennis, James M. McQueen and Anne Cutler
1995 Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21: 1209-1228.

Norris, Dennis, James M. McQueen and Anne Cutler
2000 Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences* 23: 299-325.

Peperkamp, Sharon and Emmanuel Dupoux
2002 A typological study of stress 'deafness'. In: Carlos Gussenhoven and Natasha Warner (eds.), *Papers in Laboratory Phonology 7*, 203-240. Berlin: Mouton de Gruyter.

Peterson, Robert R. and Pamela Savoy
1998 Lexical selection and phonological coding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24: 539-557.

Praamstra, Peter, Antje S. Meyer and Willem J. M. Levelt
1994 Neurophysiological manifestations of phonological processing: Latency variation of a negative ERP component time-locked to phonological mismatch. *Journal of Cognitive Neuroscience* 6: 204-219.

Radeau, Monique, José Morais and Juan Segui
1995 Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology: Human Perception and Performance* 21: 1297-1311.

Salverda, Anne Pier, Delphine Dahan and James M. McQueen
submitted The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension.

Schiller, Niels O., Antje S. Meyer, R. Harald Baayen and Willem J. M. Levelt
1996 A comparison of lexeme and speech syllables in Dutch. *Journal of Quantitative Linguistics* 3: 8-28.

Shillcock, Richard C.
1990 Lexical hypotheses in continuous speech. In: Gerry T. M. Altmann (ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, 24-49. Cambridge, MA: Massachusetts Institute of Technology Press.

Slowiaczek, Louisa M. and Mary B. Hamburger

1992 Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18: 1239-1250.

Soto-Faraco, Salvador, Núria Sebastián-Gallés and Anne Cutler
2001 Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language* 45: 412-432.

Spinelli, Elsa, James M. McQueen and Anne Cutler
2003 Processing resyllabified words in French. *Journal of Memory and Language* 48: 233-254.

Streeter, Lynn A. and Georgia N. Nigro
1979 The role of medial consonant transitions in word perception. *Journal of the Acoustical Society of America* 65: 1533-1541.

Swinney, David A.
1981 Lexical processing during sentence comprehension: Effects of higher order constraints and implications for representation. In: Terry Myers, John Laver and John Anderson (eds.), *The Cognitive Representation of Speech*, 201-209. Amsterdam: North-Holland.

Tabossi, Patrizia, Cristina Burani and Donia Scott
1995 Word identification in fluent speech. *Journal of Memory and Language* 34: 440-467.

Tabossi, Patrizia, Simona Collina, Michela Mazzetti and Marina Zoppello
2000 Syllables in the processing of spoken Italian. *Journal of Experimental Psychology: Human Perception and Performance* 26: 758-775.

Tanenhaus, Michael K., James S. Magnuson, Delphine Dahan and Craig Chambers
2000 Eye movements and lexical access in spoken language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research* 29: 557-580.

Utman, Jennifer A., Sheila E. Blumstein and Martha W. Burton
2000 Effects of subphonetic and syllable structure variation on word recognition. *Perception & Psychophysics* 62: 1297-1311.

Van Berkum, Jos J. A., Pienie Zwitserlood, Peter Hagoort and Colin Brown
submitted Discourse-dependent N400 effects in spoken language comprehension.

Van den Brink, Dannie, Colin Brown and Peter Hagoort
2001 Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience* 13: 967-985.

Van Petten, Cyma, Seana Coulson, Susan Rubin, Elena Plante and Marjorie Parks
1999 Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25: 394-417.

Vitevitch, Michael S. and Paul A. Luce
1998 When words compete: Levels of processing in spoken word recognition. *Psychological Science* 9: 325-329.

Vitevitch, Michael S. and Paul A. Luce
1999 Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40: 374-408.
Vroomen, Jean and Beatrice de Gelder
1995 Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 21: 98-108.
Vroomen, Jean and Beatrice de Gelder
1997 Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 23: 710-720.
Wallace, William P., Mark T. Stewart and Christine P. Malone
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, William P., Mark T. Stewart, Thomas R. Shaffer and John A. Wilson
1998 Are false recognitions influenced by prerecognition processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24: 299-315.
Wallace, William P., Mark T. Stewart, Heather L. Sherman and Michael Mellor
1995 False positives in recognition memory produced by cohort activation. *Cognition* 55: 85-113.
Whalen, Doug H.
1984 Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics* 35: 49-64.
Whalen, Doug H.
1991 Subcategorical phonetic mismatches and lexical access. *Perception & Psychophysics* 50: 351-360.
Zwitserlood, Pienie
1989 The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition* 32: 25-64.
Zwitserlood, Pienie and Herbert Schriefers
1995 Effects of sensory information and processing time in spoken-word recognition. *Language and Cognitive Processes* 10: 121-136.