Cognition, 38 (1991) 245–294

The mental representation of lexical form: A phonological approach to the recognition lexicon*

ADITI LAHIRI
Max-Planck-Institute for Psycholinguistics, Nijmegen

WILLIAM MARSLEN-WILSON
Max-Planck-Institute for Psycholinguistics, Nijmegen and MRC Applied Psychology Unit, Cambridge, U.K.

Received September 9, 1988, final revision accepted September 30, 1990

Abstract

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

We propose a psycholinguistic model of lexical processing which incorporates both process and representation. The view of lexical access and selection that we advocate claims that these processes are conducted with respect to abstract underspecified phonological representations of lexical form. The abstract form of a given item in the recognition lexicon is an integrated segmental–featural representation, where all predictable and non-distinctive information is withheld. This means that listeners do not have available to them, as they process the speech input, a representation of the surface phonetic realisation of a given word-form. What determines performance is the abstract, underspecified representation with respect to which this surface string is being interpreted.

These claims were tested by studying the interpretation of the same phonological feature, vowel nasality, in two languages, English and Bengali. The underlying status of this feature differs in the two languages; nasality is distinctive only in consonants in English, while both vowels and consonants

*The order of authors is alphabetical. Each contributed equally to this paper. Comments from Gary Dell, Lyn Frazier, Bruce Hayes, Pat Keating, Pim Levett and Joan Sereno were greatly appreciated. We would also like to thank Henning Reetz for the speech facilities which made this research possible. Requests for reprints should be addressed either to A. Lahiri, Max-Planck-Institute for Psycholinguistics, Wundtlaan 1, NL – 6525 XD Nijmegen, The Netherlands, or to W.D. Marslen-Wilson, Birkbeck College, Department of Psychology, Malet St., London WC1E 7HX, U.K.

contrast in nasality in Bengali. Both languages have an assimilation process which spreads nasality from a nasal consonant to the preceding vowel. A cross-linguistic gating study was conducted to investigate whether listeners would interpret nasal and oral vowels differently in the two languages. The results show that surface phonetic nasality in the vowel in VN sequences is used by English listeners to anticipate the upcoming nasal consonant. In Bengali, however, nasality is initially interpreted as an underlying nasal vowel. Bengali listeners respond to CVN stimuli with words containing a nasal vowel, until they get information about the nasal consonant. In contrast, oral vowels in both languages are underspecified for nasality and are interpreted accordingly. Listeners in both languages respond with CVN words (which have phonetic nasality on the surface) as well as with CVC words while hearing an oral vowel.

The results of this cross-linguistic study support, in detail, the hypothesis that the listener's interpretation of the speech input is in terms of an abstract underspecified representation of lexical form.

1. Introduction

A psycholinguistic theory of spoken language comprehension will have to combine both psychological and linguistic approaches – as conventionally defined – to the study of language. Experimental psychological research specifies the processing constraints on the functional properties of mental representations, while linguistic research specifies the potential structure and content of these representations. In this paper we are concerned with the mental representation of *lexical form*, and with the way the speech signal is mapped onto these representations during the recognition of spoken words. We propose here a novel psycholinguistic theory of the properties of these representations, drawing chiefly on developments in phonological theory over the past decade, and making the claim that spoken word recognition is conducted with respect to abstract underspecified phonological representations, from which all predictable and non-distinctive information is withheld.

1.1. Goals and problems

In the context of a theory of language comprehension, one of the major goals of a psycholinguistic theory of lexical form is to specify the properties of what we can call the *recognition lexicon*: namely, those aspects of the representation of lexical form that participate directly in the process of recognising spoken words, allowing the listener to identify the sequence of lexical items being produced by a given speaker.

The fundamental problem faced by a theory of the recognition lexicon is the problem of *variation* – the fact that no two tokens of a given lexical form are identical. There are two major sources for this variability – variations in the physical properties of speakers and variations in the phonological conditions under which a given form is realised.

Speakers differ, first, in the properties of their vocal tracts. This leads to wide variations in the acoustic–phonetic realisation of the same forms by different speakers. The recognition lexicon must either abstract away from this variability or else depend on some form of normalisation prior to entry into the lexicon.

More importantly, speakers also differ in the properties of their phonological systems. Different dialects of English vary in the kinds of phonological rules that they apply. American English, for example, flaps intervocalic alveolar stops under certain conditions, while British English does not. Phonologically based variation is not, however, just a matter of dialect variation between speakers. In the output of any individual speaker, the phonetic realisation of a given form can vary quite radically, according to the phonological conditions under which it is produced. These conditions include not only the properties of the immediate phonological environment (what precedes and follows the form in the speech stream), but also factors such as register and speech rate.

The problem for the recognition lexicon, again, is to find a way of representing the form of words so that it can cope with this degree of variation; with the fact that a word-form may be produced with its segmental – even syllabic – properties changed in different ways. Segments may be deleted, they may appear as allophonic variants, they may change their form due to assimilation with neighbouring items, and so on.

These problems of variability introduce a fundamental conflict into the demands placed on the recognition lexicon. On the one hand, the system must make choices. It must choose to represent a given word-form in a specific way, and it must do so sufficiently restrictively to keep it distinct from other lexical items in the language. On the other hand, this representation cannot directly correspond to any particular token of the form in question. It must be sufficiently abstract – where the details of phonetic realisation are concerned – to allow for recognition under the conditions of variability sketched above. And the more abstract the representation is assumed to be, then the greater the divergence between the properties of the word-form as presented in the speech input and the properties of this form as captured in the recognition lexicon.

What we need, therefore, is a theory of lexical representation which can resolve this conflict between specificity and abstractness. Our goal in this paper is to propose a particular resolution to this conflict. The question we ask is how much detail must there be in the representation of a word-form in the mental lexicon for it to be recognised? In other words, how abstract is abstract?

1.2. Background

Research in psycholinguistics has had very little to say about the representation of lexical form – and even less about its potential role in solving the problem of variation. There have been two reasons for this. The first was simply the assumption, implicit or explicit, that the properties of lexical form representations did not raise significant theoretical issues. The important psycholinguistic questions were assumed to be questions about process – asking, for example, whether words are accessed in parallel or in series, whether the access process is modular or interactive, and so on.

The second reason for the neglect of lexical representations was the assumption – almost universal in speech research as well as in psycholinguistics – that the important action in speech recognition took place prior to the lexicon, in the mapping from the speech signal onto a pre-lexical level of representation. It was here, if anywhere, that the problems of variation were solved, and where the noisy clamour of the speech signal was converted into strings of abstract labels for input to higher levels. There have been a variety of claims about exactly which units constituted this intervening level – features, phonemes, diphones, syllables – but the dominant view, in all areas of language research, has been that the speech processor generates as input to the lexicon a string of discrete phoneme-like units (cf. Pisoni & Luce, 1987), and that the primary business of acoustic–phonetic analysis was to map from the speech signal onto this pre-lexical segmental level. Consistent with this, the representations of words in the recognition lexicon were viewed as listings of linear strings of phonemic labels, and treated, in effect, as the auditory analogue of the representations of written words as strings of letters.

There are many problems with this view of the access of word-forms from the speech signal. One basic set of difficulties concern the linguistic adequacy of its assumptions about mental representation. At the sublexical level, the approach depends on the assumption that there is a meaningful definition of the notion "phoneme" (or equivalent unit), such that independent units of this type can be the perceptual targets for the acoustic–phonetic analysis process, as well as the means by which information about the outcome of these analyses is transmitted to the lexical level. At the lexical level, this approach assumes that the properties of lexical form can be adequately characterised as a linear string of phonemes.

The theoretical basis for these assumptions about representation can be traced back to the systematic phonemic level in taxonomic linguistics, and to the SPE (The Sound Pattern of English, Chomsky & Halle, 1968) view of form representations as linear strings of segments. But these historical connections are no longer valid. Current linguistic analyses of lexical structure no longer provide the theoretical underpinnings for a processing system that assumes either an autonomous level of segmental representation outside the lexicon, or a linear, single-level analysis of lexical representation. Although there may be a level of description in phonological accounts that roughly corresponds to the phonemic unit of analysis, this is just one level in a system of analytic categories ranging from the feature up through the segment to higher levels of hierarchically organized units, and where this entire apparatus is necessary to capture the abstract phonological form of a word.

The second set of difficulties concern the adequacy of the conventional model in dealing with the problems of phonological variation. On this account, the problem of variation is assumed not to involve the lexical level. Instead it is considered to be an issue in speech research, involving the processes that extract and interpret the phonetic cues in the speech input and map them onto segmental labels. Although this may be an appropriate strategy for coping with vocal tract variation (which could be dealt with by normalising processes operating early in speech analysis), it is not a strategy that can solve the problems of phonological variability. The problem here is not to relate variable phonetic cues to segmental labels but to relate variable segmental cues to lexical form representations. Let us take an example from English to clarify this point.

The final consonant cluster in the word *hand*, which could be classified as [nd] in isolated careful speech, can undergo phonological variation in different contexts. The [d] can become a [dʒ] when followed by *you*; it can be lost and the [n] can become a [m] when preceding a word like *me*; or the loss of the [d] can be accompanied by the change of [n] to a velar [ŋ] when followed by *care*. This kind of neutralising variation, where processes of assimilation change the identity of segments (e.g., from [n] to [m]), is extremely common in normal speech, and it poses problems which cannot be solved pre-lexically. Even if one assumes that there is an intervening level of the conventional sort, the output of the acoustic phonetic processor would be [nd], [dʒ], [m] and [ŋ] in the four cases, and there would be no grounds, pre-lexically, for determining that they were all realisations of the same underlying form. This leaves to the lexical level the problem of determining that each of these strings is indeed a realisation of the word *hand*. To do so, it must be able to match the various segmental labels with a lexical representation which uniquely describes this word.

The conventional model provides no principled way of dealing with this problem. The only solution it can offer is the assumption that all phonological variants of a word are listed in the recognition lexicon. This is not only inelegant, but also, we assume, not feasible as a general solution, given the productivity of the phonological processes involved. The alternative, which we will argue for here, is to assume that there is a *single* underlying representation of each lexical item, which abstracts away from all surface detail, and which is compatible with all phonologically permissible variants.

This is an approach which leads to a much deeper involvement of the lexicon in the on-line solution to the variability problem, and which is rooted much more firmly in linguistic conceptions of the representation of lexical form. It requires us to abandon the notion of an intervening segmental level, and to argue instead for a system where the input to the lexical level is featural (Klatt, 1989; Stevens, 1986), and where phonological variation is resolved directly at the lexical level, as a consequence of the abstract properties of the representations of lexical items in the mental recognition lexicon.

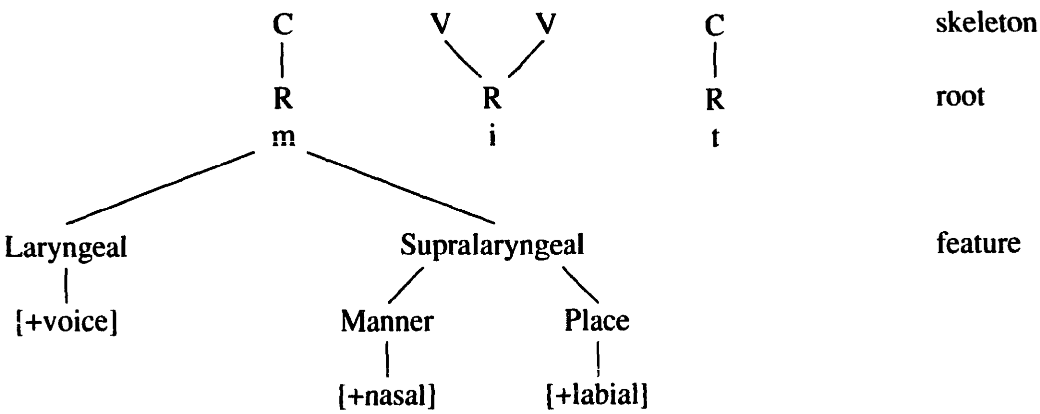
1.3. A phonological approach to the recognition lexicon

The psycholinguistically relevant representations of lexical form must be *abstract* in nature – that is, they must in some way abstract away from the variabilities in the surface realisation of lexical forms. This means that our account of the properties of the recognition lexicon must be an account in terms of some set of *underlying representations*. The only systematic hypotheses about the properties of these underlying representations are those that derive, directly or indirectly, from past or present phonological theories. It is the central business of phonological theory – and, indeed, *only* of phonological theory – to try to deduce the underlying, abstract properties of linguistic form, and it does so on the basis of a very wide range of empirical data about the phonological patterning of the world's languages.

For these reasons, and given the need to find an alternative to the representational assumptions that we discussed in the previous sections, we look to phonological theory to suggest the broad outlines of our hypotheses about the nature of the recognition lexicon (see Frauenfelder & Lahiri, 1989; Frazier, 1987, for a related discussion). From the consensus of current opinion about the general properties of phonological representations we can extract claims about the content and structure of lexical form representations. First, we will look briefly at the claims about structure.

As we mentioned earlier, psycholinguistic theories have assumed, explicitly or implicitly, that the lexical representation consists of a string of unstructured phoneme-like units. This was consistent with the account given by Chomsky and Halle (1968), where underlying segments were represented as unordered columns of features, concatenated in linear strings. Over the last decade, however, cross-linguistic evidence from various phonological processes has demonstrated that a more hierarchical organization is necessary to capture the properties of sound patterns in natural language. This hierarchical representation reflects the fact that phonological processes consistently affect certain subsets of features and not others. Individual features or subsets of features are functionally independent units and are capable of acting independently.

To illustrate the kind of representation which has emerged from these analyses, we give below a partial sketch of the representation for the word *meet*:[1]



This particular representation is primarily based on Clements (1985, 1989). Our interest is not so much in its specific details, but rather in the basic concepts which it exemplifies. This is a hierarchically organized representation, where features are organized into functionally related groups dominated by abstract class nodes (such as place). The phonological features are the terminal nodes, and the entire feature structure is dominated by the root node (made up of the major class features like [consonantal] and [sonorant]) which corresponds to the traditional notion of a single segment. The root,

[1]In the last decade, research in phonology has demonstrated the need for several levels of hierarchically organized prosodic structures, like syllables, feet, prosodic word, phonological phrase, etc. (Nespor & Vogel, 1986). The syllable is assumed to be built on the "skeleton", which consists of abstract timing units, and mediates between the root and the syllable. Everything above the skeleton (syllables, feet, etc.) is assumed to be derived and therefore not part of the underlying representation of a lexical item. We will, therefore, only describe in detail the featural organization of words.

however, does not contain any information about quantity. It is linked to a slot or slots in the skeletal tier which represent abstract phonological timing positions (C = [−syllabic], V = [+syllabic]).[2] Note that a representation of this sort does not choose between one or the other levels. There is no question of a choice between the skeletal tier, the root tier, and the feature tier. The underlying phonological structure is an integrated representation of all the levels together.

These claims about the structure of representations are closely tied to claims about their actual *content*. What is to be treated as part of the abstract properties of linguistic representation, and what is to be treated as part of the surface realisation, and therefore not represented underlyingly?

The basic principle of phonological analysis is the hypothesis that every linguistic item has a single unique underlying representation which is minimally specified in its phonetic description. This abstract representation consists only of idiosyncratic and unpredictable information from which all the surface forms are derived. All predictable features, because they can be derived by rule, are not specified in the underlying representation. The notion of predictability is itself defined in terms of two formal principles. These provide the criteria for deciding whether or not to assign properties to the underlying representation. The general principle of *redundancy* determines which individual features are to be specified, and, the principle of *underspecification* determines which value of the feature is to be represented.

The phonological inventory of a language consists of the *minimum* number of distinct segments which are not predictable by any phonological rule. Each underlying segmental unit represents a complex set of distinctive features selected from a subset of possible features that can be used distinctively in any language. Other relevant but non-distinctive features may be filled in by *redundancy rules*. Such rules can be language specific, depending on what features may or may not be distinctive in a given language. Aspiration in Bengali is distinctive – both voiceless and voiced stops have aspirated counterparts in Bengali, and minimal pairs like [kɑl] "tomorrow" and [k^{h}al] "drain" occur frequently. This feature, therefore, must be specified in the underlying representation. Not so in English, since the aspirated voiceless stops occur in

[2]It is generally accepted that there is a skeletal level mediating between the featural level and the syllable level. However, opinions differ regarding the actual nature of the units in the skeleton. Some researchers have argued for units which do not have any syllabic information in the skeleton (Levin, 1985), while others have argued for incorporating time and weight information in one moraic unit as part of the skeleton (Hayes, 1989). We have used the C and V as the units of the skeleton for purely expository reasons. Nothing crucially hinges on this for the present research. What matters, of course, is that there is a separate skeleton independent of the features.

a predictable environment, namely syllable initially, and this feature can be appropriately assigned in the surface form.

Some redundancy rules, however, are not language specific. These are universal default rules which will be present in all languages. The fact, for example, that all nasals are by default sonorant will be expressed in all languages by an appropriate rule.

These considerations of redundancy provide one principle upon which the content of underlying representations are determined. Current phonological theory goes a step further, with a second principle, that of *underspecification* (cf. Archangeli, 1984; Kiparsky, 1985; Pulleyblank, 1983). In the SPE framework, all features are fully specified for each segment before any phonological rule can apply. Although the redundancy rules indicate which features are distinctive, they do not provide any information about the *markedness* of these features – where the unmarked value of a feature can be regarded, for present purposes, as corresponding to its default value.

In SPE-type underlying representations, features were assigned marked or unmarked values which were translated into binary "+" or "−" values by marking conventions. Current theories of underspecification differ from this by adopting the principle that the unmarked feature is not specified at all in the representation. For example, in Bengali the feature [nasal] is distinctive for vowels, and both oral and nasal vowels occur as underlying segments. However, only the [+nasal] value is specified underlyingly, with the non-nasal (oral) vowels remaining unspecified for nasality. This is because the feature [+nasal] is assumed to be marked; [−nasal] is unmarked, and can be filled in by the appropriate redundancy rules. The feature array for a given segment will not contain a specification for any feature, distinctive or not, that has the unmarked value. Consequently, the only specifications in the underlying representation, on this account, are those for features which are (a) distinctive, and (b) have the marked (or non-default) value.

The principle of underspecification[3] is supported by a wide range of phonological evidence. One important piece of evidence comes from the patterning of assimilation processes, since if a class of sounds is not specified for a given feature, a feature-filling rule can spread a feature from a neighbouring segment. But this is assumed to be impossible if the feature is

[3]Theories of underspecification differ with respect to degrees of underspecification. One view is that both values of a feature are specified if it contrasts two segments; non-contrastive redundant values are left unspecified (Clements, 1987; Mester & Ito, 1989). The other, more radical view, does not allow any predictable information to be specified (universal or language specific) and permits only one value of a feature to be specified in the underlying representation (Archangeli, 1984).

already specified. In many languages coronal nasals (like [n] can assimilate to the following stop, while velar or palatal nasals (like [ŋ] or [ɲ]) remain unchanged. In Catalan, for instance, one finds /n/ assimilation in words like *so[m]pocs* "they are few", *so[n,]žermans* "they are brothers" (cf. the basic form *so[n]amics* "they are friends"), while there is no assimilation with palatal /ɲ/ and velar /ŋ/ – *a[ɲ]feliç* "happy year", *ti[ŋ]pa* "I have bread" (Mascaró, 1976). This asymmetry in the assimilation is readily explicable if we assume that coronal consonants are not specified for place and can "borrow" the place feature from the following consonant, whereas velars and palatals, which are underlyingly specified for place, cannot do so.

In summary, while there is still controversy over certain aspects of underspecification, it is now widely accepted that underlying representations exploit principles not only of redundancy but also of underspecification in determining the minimal set of features necessary to differentiate the underlying representation of the segments of a given language. We assume that this consensus has consequences for how we should view the recognition lexicon. Phonological theory cannot be interpreted literally, as a direct description of mental representations, but it can be interpreted as specifying the functional properties of these representations. Our hypothesis here, therefore, is that the lexical representations deployed in speech recognition also contain only distinctive and marked information. Predictable information, whether language independent or language dependent, will not be specified.

The crucial implication of this for the lexical access process is that the system will assign a different status to information in the signal as a function of its relationship to what is or is not directly specified in the recognition lexicon. In particular, if a given feature is not specified, then variations in the phonetic realisation of that feature should not affect the goodness of fit between the speech input and the form specification for the item in question.

1.4. The processing environment for lexical access

In the preceding sections we outlined some claims about the mental representation of lexical forms. To be able to evaluate these claims in a psycholinguistic framework, they need to be interpreted in the context of a theory of the general processing environment for lexical access. The theory that we will assume here is the *cohort model* of spoken word recognition (Marslen-Wilson, 1984, 1987; Marslen-Wilson & Welsh, 1978).

The cohort model distinguishes an initial, autonomous process of *lexical access and selection*, responsible for the mapping of the speech signal onto the representations of word-forms in the mental lexicon. The general properties of these processes can be laid out as follows. For each lexical unit, there is a discrete, computationally independent recognition element, where each such unit represents the functional co-ordination of the bundle of phonological, morphological, syntactic, and semantic properties defining a given lexical entry (here, of course, we are concerned only with the phonological aspects of the representation). The recognition lexicon is constituted, then, by the entire array of such elements, for a given listener.

The second major property of the system is that it allows for the simultaneous, parallel activation of each recognition element by the appropriate input from the pre-lexical processes of acoustic–phonetic analysis. This is coupled with the further assumption that the level of activation of each element reflects the goodness of fit of the input to the form specifications for each element. As more matching input accumulates, the level of activation will increase. When the input pattern fails to match, the level of activation starts to decay.

These assumptions lead to the characteristic cohort view of the form-based access and selection process. The process begins with the multiple access of word-candidates as the beginning of the word is heard. All of the words in the listener's mental lexicon that share this onset sequence are assumed to be activated. This initial pool of active word-candidates forms the *word-initial cohort*, from among which the correct candidate will subsequently be selected. The selection decision itself is based on a process of successive reduction of the active membership of the cohort of competitors. As more of the word is heard, the accumulating input pattern will diverge from the form specifications of an increasingly higher proportion of the cohort membership.

This process of reduction continues until only one candidate remains still matching the speech input – in activation terms, until the level of activation of one recognition element becomes sufficiently distinct from the level of activation of its competitors. At this point the form-based selection process is complete, and the word-form that best matches the speech input can be identified.

For our current concerns, the most important feature of this processing model is that it is based on the concept of *competition* among alternative word-candidates. Perceptual choice, in the cohort approach, is a contingent choice. The identification of any given word does not depend simply on the information that this word is present. It also depends on the information that other words are not present, since it is only at the point in the word where no other words fit the sensory input – known as the "recognition-point" – that the unique candidate emerges from among its competitors.

The recognition of a word does not depend on the perceptual availability of a complete specification of that word in the sensory input, either where individual segments are concerned, or where the word as a whole is concerned. The information has to be sufficient to discriminate the word from its competitors, but this is a relative concept. The significance of this is that it makes the basic mode of operation of the recognition process compatible, at least in principle, with the basic analytic procedures of phonological theory – namely, the attempt to strip away from underlying representations any information which is not necessary to maintain the distinctiveness of the item in question. In a contingent, competitor-based recognition system, this will maintain the distinctiveness of the item in the recognition process as well as in the linguistic analysis.

The second important aspect of the cohort approach is its emphasis on the continuous and sequential nature of the access and selection process. The speech signal is based on a continuous sequence of articulatory gestures, which result in a continuous modulation of the signal. In recent research (Warren & Marslen-Wilson, 1987, 1988), we have shown that this continuous modulation of the speech signal is tracked in detail by the processes responsible for lexical access and selection. As featural information becomes available in the signal, its consequences are immediately felt at the lexical level. Durational cues, for example, to the voicing of a post-vocalic consonant in English, start to affect the listener's behaviour well before the end of the vowel is reached (Warren & Marslen-Wilson, 1988). Similarly, the presence of vowel nasalisation, which indicates (for languages like English) that the following consonant is nasal, is also picked up and used to guide lexical choice early in the preceding vowel (Warren & Marslen-Wilson, 1987). In each case, the listener uses this featural information to select words that are compatible with these cues, even though the final segment cannot yet be uniquely identified.

On a number of counts, then, the cohort view of lexical access is compatible with the phonological view of lexical representation that we outlined earlier. It allows for an on-line process of competition between minimally specified elements, where this minimal specification is still sufficient to maintain distinctiveness, and, second, it allows for this competition to be conducted, with maximal on-line efficiency, in terms of a continuous stream of information about the cues that the speech signal provides to lexical identity, where these cues are defined in featural terms.

Given this preliminary sketch of our claims about lexical representation in the context of a model of lexical processing, we now turn to an experimental investigation of the psycholinguistic model that has emerged.

1.5. An empirical test

The view of lexical access and selection that we have developed here makes the claim that these processes are conducted with respect to an abstract *underspecified* representation, corresponding approximately to the underlying representations of current phonological theory. We will contrast this with the view that the perceptually relevant representation in some way directly encodes the surface phonetic form of the word. Naturally, this "surface representation" will be abstract as well, in the restricted sense that it will not include the idiosyncratic details of a word's realisation by a given speaker in a given phonetic environment. Where it differs from our proposal is because it is a *fully specified* phonetic representation, corresponding, in phonological terms, to the representation of a word's phonological form after all the phonological rules have applied. This type of surface representation hypothesis is the view of lexical representation held by essentially all current theories of spoken word recognition, where the properties of the recognition lexicon reflects some sort of averaging over individual listeners' experience with the surface phonetic properties of word-forms in their language.

We hypothesize, in contrast, that listeners do not have available to them, as they process the speech input, a representation of the surface phonetic realisation of a given word-form. What determines their performance, instead, is an abstract, underspecified representation with respect to which they interpret the speech input.

To test this claim we need to show that it is indeed an underspecified representation that determines the interpretation of a given phonetic cue and not a representation of fully specified surface form. This requires us to test listeners' lexical processing under conditions where the same surface feature varies in its underlying phonological status – and, therefore, in its representation in an abstract underspecified recognition lexicon. We can do this in two ways. First, within a given language, we can compare the interpretation of cues corresponding to the marked and unmarked values of a given feature, where the marked value is underlyingly specified and the unmarked value is not. Secondly, by testing cross-linguistically, we can look at the interpretation of the same surface feature as its underlying phonological status varies across different languages. In each case, if it is the surface form per se which controls the recognition process, then its interpretation should remain constant. If it is an abstract underspecified representation that controls performance, then the interpretation of the surface feature should change as its phonological status changes.

1.5.1. The oral/nasal contrast in English and Bengali

The feature that we chose to concentrate on was the oral/nasal contrast for vowels. We made this choice for several reasons; first, because of the uncontroversial status of the feature [nasal] in phonological theory. It is universally accepted that nasal vowels are marked and that they exist only in languages which have the unmarked oral counterparts. Therefore, distinctive nasal vowels are assumed to be marked underlyingly as [+nasal], whereas the oral vowels are left unspecified (e.g., Archangeli, 1984). This provides the first basic contrast we need, where, within a given language, there is an asymmetry in the representation of the marked and unmarked values of a given feature.

The second reason for choosing the feature [nasal] was because vowel nasalisation in the surface (we will refer to this as phonetic nasalisation) can come from one of two different sources. It can come from an underlying nasal vowel or it can be derived, by a process of feature assimilation, from a neighbouring nasal segment. Such processes of regressive assimilation, where nasality spreads from a nasal consonant to a preceding oral vowel, are widespread cross-linguistically. In English, for example, the word *ban*, where an oral vowel is followed by a nasal consonant, is standardly pronounced as [bæ̃n], because the nasality of the consonant spreads to the preceding vowel. This gives us the second basic contrast that we need, where the same surface feature can have a varying phonological status, contrasting both within a given language and between different languages.

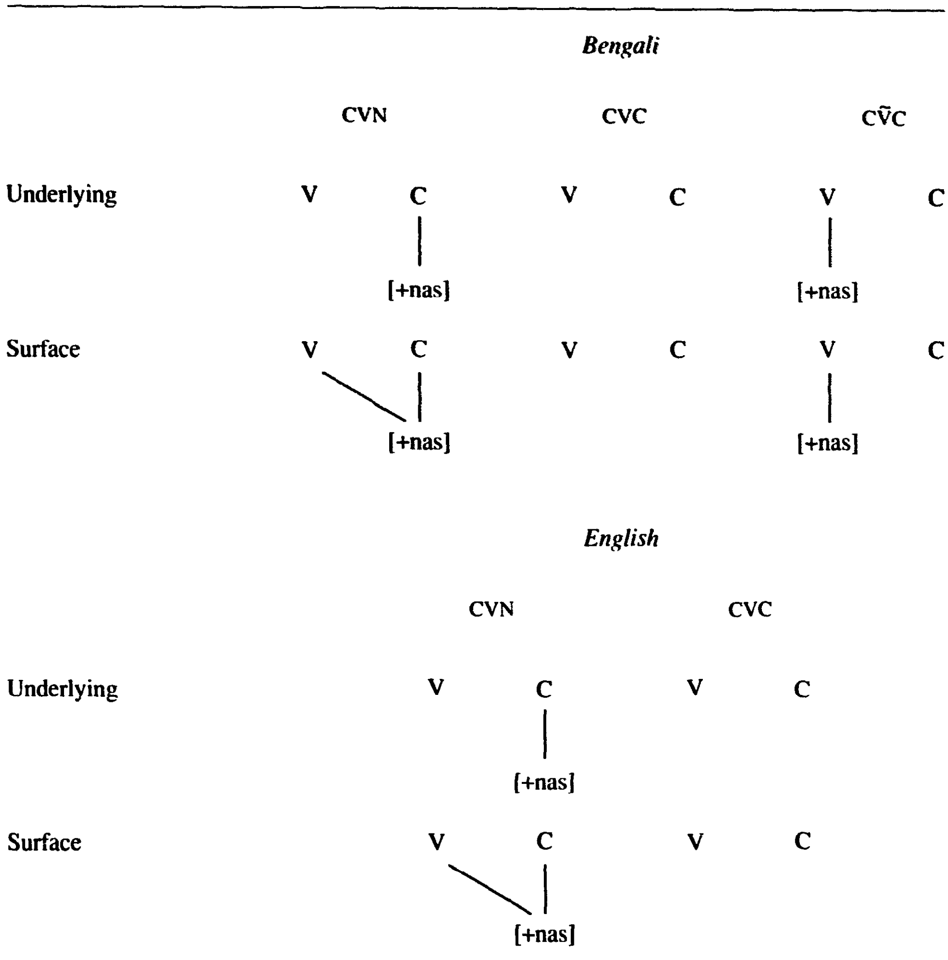
The final reason for choosing the feature [nasal] was the availability of two languages (English and Bengali) which allowed us to realise these contrasts in the appropriate stimulus sets. The relevant linguistic facts are summarised in Table 1 (remember that these two levels of underlying and surface phonological representation correspond, respectively, to the two competing hypotheses about the contents of the mental recognition lexicon).

English, as shown in the lower half of the table, has only oral vowel segments in the underlying representation. For both CVC and CVN words the vowel is underlyingly oral.[4] However, an allophonic rule nasalises all oral vowels when followed by a nasal consonant. This is indicated in Table 1 as the feature nasal spreading from the consonant to the preceding vowel in the CVNs, which gives surface contrasts like *ban* [bæ̃n] and *bad* [bæd].[5]

[4]The notations used throughout should be interpreted as follows: C = any consonant, V = any oral vowel, Ṽ = any nasal vowel, and N = any nasal consonant. Everything between square brackets is a surface phonetic form, and between slashes an underlying form.

[5]The assumption that the vowel in the CVN is underlyingly oral follows from the fact that surface nasalisation is always predictable, and therefore need not be specified underlyingly.

Table 1. *Underlying and surface phonological representations in Bengali and English*



Bengali, on the other hand, does have both underlyingly oral and nasal vowel segments. Each of the seven oral vowels in the language has a corresponding nasal vowel, as in the minimal pairs [pɑ̃k] "slime" and [pɑk] "cooking" (Ferguson & Chowdhury, 1960). This is illustrated in Table 1, where the vowel in the CṼC is specified as underlyingly nasal. Following the principle of underspecification, the oral vowels in the CVCs and the CVNs are not specified underlyingly for this feature. Apart from underlyingly nasal vowels, Bengali has an additional source of surface nasalisation. A rule of regressive nasal assimilation, similar to that in English, spreads the specified nasal feature of the consonant to the previous vowel in CVNs.[6] This gives surface contrasts like [bɑ̃n] "flood", [bɑ̃d] "dam", and [bɑd] "difference".

Thus, assuming that the vowels in CVNs are underlyingly oral, and given the nasal assimilation rule, this leads to the situation shown in the upper half of Table 1. Surface nasalisation in Bengali is ambiguous, since the vowels in both CVNs and in CṼCs are realised as nasal. Unlike English, therefore, the nasal assimilation rule in Bengali is neutralising – it neutralises the oral/nasal contrast in a given environment – and creates potential ambiguity.[7]

1.5.2. Experimental predictions

The phonological pattern of surface and underlying representations laid out in Table 1 allows us to specify the content of form representations in the mental recognition lexicon according to the two competing hypotheses: one where lexical representations are abstract and underspecified and the other where the recognition process is conducted in terms of a fully specified representation of the surface form of the word in question. This in turn leads to a differential set of predictions for the two hypotheses.

The hypothesis that representations are abstract and underspecified generates an integrated set of predictions, covering the interpretation of both nasal and oral vowels, for languages which have underlying nasal vowels and languages which do not. These predictions combine claims about representation with claims about the process of competition between candidates during lexical access, where the candidate best fitting the speech input eventually emerges as the lexical choice of the listener. The processing system computes a goodness-of-fit measure for each candidate, matching the specifications of

[6]This rule applies across morpheme boundaries as well as within morphemes:

/pɑn/ → [pɑ̃n] "betel leaf"
/pɑ+n/ → [pɑ̃n] "you (honorific) get"
/pɑ+ʃ/ → [pɑʃ] "you (familiar) get"

Since nasal and oral vowels do not contrast before nasal consonants, we need additional arguments to establish that the nasalised vowels preceding nasal consonants in tautomorphemic words (such as /pɑn/ above) are indeed underlyingly oral. One argument is that the nasalised vowels followed by nasal consonants are entirely predictable. A rule of nasal assimilation is independently needed for heteromorphemic words (as illustrated in the derivation of [pɑ̃n] from /pɑ+n/). Thus, our assumptions about representation of redundant features and underspecified underlying forms (especially with unmarked values (Kiparsky, 1985, p. 92)) suggest that for VN sequences the underlying representation of the vowel segment should be an unmarked oral vowel.

[7]The rule of nasal assimilation that applies in Bengali and English is postlexical – that is, it applies on the output of syntax, and is not a rule which is constrained to apply only after certain specific morphological operations. The surface variation as a result of the application of the rule – allophonic in English (introduces a feature on a vowel not present in the underlying representation) and neutralising in Bengali – is therefore phonological and cannot be attributed to any morphological domains.

words in the recognition lexicon against the information in the speech stream. Note that we are assuming here, as throughout, that the initial stage of speech processing is the extraction from the speech input of the phonetic cues corresponding to its featural properties. This featural information is projected directly onto the lexical level, and our predictions concern the way this information is interpreted over time. Because the underspecification hypothesis makes explicit claims about exactly which features are specified in the lexical representation, this enables us to make precise predictions about which word-candidates will provide the best match to a given speech input.

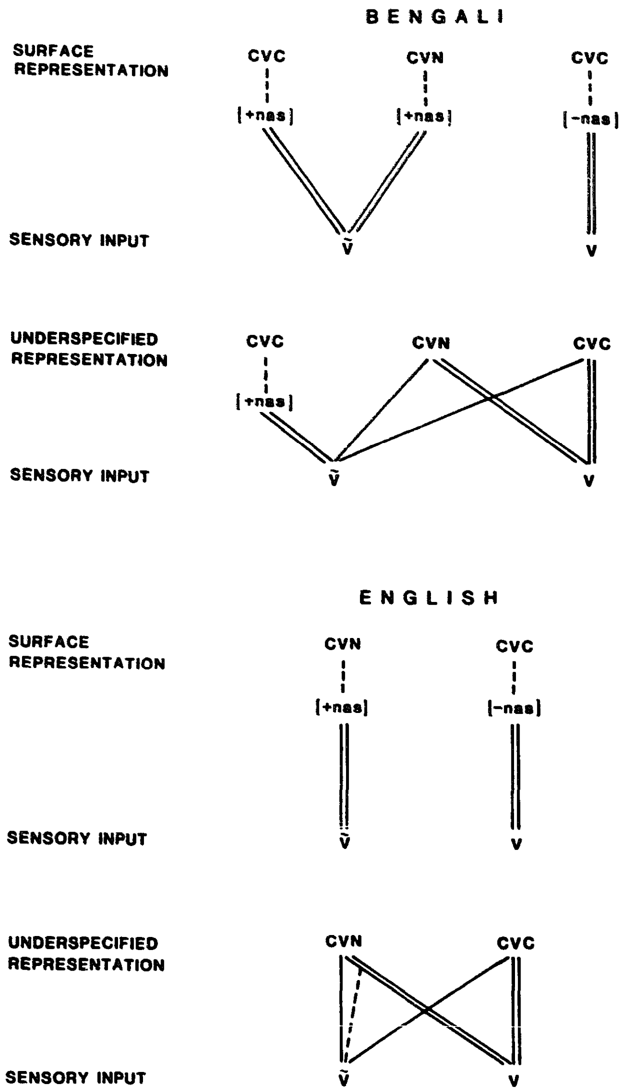
These predictions need to be tested during the period in which the listener is hearing the oral or nasal vowel (in monosyllables of the type illustrated in Table 1), and before the consonant is heard. Once the listener hears the following consonant, then the interpretation of the vowel becomes unambiguous, and the predictions of the competing theories no longer differ. The *gating* task allows us to establish how the listener is responding to the vowel before the consonant is heard. This is a task (Grosjean, 1980; Tyler & Wessels, 1983), in which listeners are presented with gradually incrementing information about the word being heard. At each increment they are asked to say what word they think they are hearing, and this enables the experimenter to determine how the listener is interpreting the sensory information presented up to the point at which the current gate terminates.

Previous research (Warren & Marslen-Wilson, 1987) shows that responses in this task are sensitive to the presence of phonetic cues such as vowel nasalisation, as they become available in the speech input. Other research (e.g., Marslen-Wilson, 1984, 1989; Tyler & Wessels, 1983; Zwitserlood, 1989) shows that gating responses give an accurate picture of the types of word-candidates generated by listeners during on-line processing of speech inputs, and of the timing with which candidates emerge from among their competitors.

What are the predictions for listeners' gating responses to phonetically oral and nasal vowels in English and Bengali? Figure 1 illustrates how each type of sensory input will match lexical representations in each language, under the contrasting representational assumptions of the two hypotheses.

For the abstract underspecification hypothesis, the double solid lines link nasal and oral vowels in the sensory input to words containing, respectively, nasal or oral vowels. Here there is a complete match between input and representation. The single solid lines, linking nasal vowels in the sensory input to words with underlying oral vowels, indicate cases where there is no mismatch between input and lexical representation. When a vowel is underlyingly oral, and therefore unspecified for the feature [nasal] in the recognition lexicon, the presence of vowel nasalisation does not create a mismatch. This means, within the cohort framework, that CVCs and CVNs remain as

Figure 1. *Contrasting predictions of underspecification and surface hypotheses for Bengali and English (see text). Double lines indicate a predicted complete match between input and lexical representation; single lines indicate an incomplete match but without mismatch, and dotted lines indicate linking to nearest available appropriate feature.*



possible candidates until the following consonant is heard. Note that, for Bengali, there is no linkage between oral vowels in the sensory input and words with nasal vowels. This is because there is an explicit mismatch between the input and a specified feature in the lexical representation. The consequence of this is that CṼC words should not remain as candidates in the cohort. The dotted line, finally, for the English nasalised vowel, indicates the anticipatory linking to a following segment containing a similar feature.

Turning to the surface representation hypothesis, here we see only double solid lines linking inputs to representations. The representation in the recognition lexicon reflects the surface phonetic form of words, so that phonetically oral vowels in the sensory input will only match CVC words in the lexicon, while phonetically nasal vowels in the input will only match CVN and CṼC words. There is no possibility for the form of linkage indicated by the single solid lines for the underspecification hypothesis.

We can now straightforwardly summarise the predictions of the two hypotheses, beginning with listeners' responses to the presence of vowel nasalisation. For English, the abstract underspecification hypothesis allows vowel nasalisation to be interpreted as a cue to the presence of a nasal consonant since no other nasal feature is available in the lexicon (see Fowler, 1984, for evidence that listeners are indeed able to factor out overlapping cues in this manner). Listeners can therefore start to give CVN words as responses as soon as they detect nasalisation in the signal. Since there is, however, no mismatch with the CVC words, they can also give CVCs as responses to a nasal input.

Under the surface hypothesis, the nasal vowel in English will directly map on to CVNs where the vowel is represented as nasal. But it will not match CVC words, predicting that these will be ruled out as responses to nasal vowel inputs.

The predictions of the two hypotheses diverge more strongly when we turn to Bengali. If the Bengali listeners are interpreting vowel nasalisations relative to the underspecified representations sketched in Figure 1, then they will treat nasalisation as an unambiguous cue to the status of the vowel being heard. Since it is only CṼCs that are specified as having nasal vowels, CṼC words will provide the best match with the speech input, and will therefore be preferred as responses. Note, however, that CVCs and CVNs are not ruled out as responses, since there is no mismatch here which would exclude them from the cohort of active candidates.[8]

[8]One additional point needs to be made regarding the representation of the oral vowels. Under the underspecification hypothesis, no oral vowel is specified for nasality. Hence all vowels, whether nasal or oral, are good matches for any V in the representation if the input vowel shares the appropriate quality features (like [high] or [round]). Thus, CVNs and CVCs which are unspecified for [nasal] are not ruled out as possible candidates on hearing a nasal vowel. Nonetheless, because the CṼC words do provide a more complete match with the signal, these should be preferred as candidates. This preference should hold whether the listener is hearing a CVN or a CṼC, since it is only when the final consonant is heard that the two types of word should diverge. Not being specified for [nasal], really means that a word can potentially match to a larger set of vowels, both nasal as well as non-nasal.

On a surface hypothesis, in contrast, vowel nasalisation is thoroughly ambiguous between CVN and CṼC words. The listener's experience of each type of word is with a phonetically nasal vowel, and this should be reflected in its representation in the recognition lexicon. Listeners should give both CVN and CṼC words as responses, and with relatively equal frequency.

The second important set of predictions concern listeners' responses to phonetically oral vowels – that is, to the CVC stimuli in English and Bengali. On the surface hypothesis, the absence of nasalisation (the presence of an oral vowel) should be just as informative and just as discriminative as the presence of nasalisation (as indicated in Figure 1). On the underspecification hypothesis, there is a basic asymmetry between the presence and the absence of nasalisation. Despite differences between English and Bengali in the phonological status of vowel nasalisation, the status of the lack of nasalisation is the same for both languages. In each case, words with underlyingly oral vowels are unspecified for the feature [nasal] in the lexical representation. This leads to crucial differences in the predictions of the two hypotheses.

For both English and Bengali, the underspecification hypothesis predicts that oral vowels will be ambiguous between CVN and CVC words, as indicated by the double lines in Figure 1. In each language CVNs and CVCs are equally good matches to phonetically oral vowels, since neither type of word contains vowels specified for nasality. Thus, for both Bengali and English listeners, CVNs should be possible responses to CVC inputs. In strong contrast, CṼC words should not be produced as responses to oral vowels, because here there is a mismatch with the underlying specification of the vowel.

The predictions of the surface hypothesis are quite different. If the lexical representations used in recognition directly capture the fact that CVNs are produced with phonetically nasal vowels, then CVNs should not be produced as responses to CVCs in either language. Moreover, if the oral vowel is marked as [−nasal] in the lexicon, the presence of an oral vowel should be just as informative, relative to the choice between CVCs and CVNs, as the presence of a nasalised vowel. And if any CVNs are produced in response to CVCs, then listeners should produce CṼCs as responses as well.

We will examine these predictions, for listeners' responses to oral and nasal vowels, in three different stimulus sets. Two sets reflect the structure laid out in Table 1: a set of CVC, CVN, and CṼC triplets in Bengali, and a set of CVC and CVN doublets in English. To allow a more direct comparison with English – and an even stronger test of the underspecified representation hypothesis – we will also include a set of Bengali doublets, consisting of CVN and CVC pairs where the lexicon of the language does not contain a CṼC word beginning with the same consonant and vowel as the CVN/CVC pair. This will place the Bengali listeners, as they hear the CVN stimuli, in a superficially similar position to English listeners exposed to an English CVN. In each case, the item is lexically unambiguous, since there are no corresponding CṼC words lexically available.

The surface representation hypothesis predicts that Bengali listeners should produce a very high proportion of CVN responses when hearing a CVN from one of these doublets. CVN words are represented with nasalised vowels, and there are no appropriate CṼC words available as responses. CVC responses, in contrast, should be very infrequent – and certainly no more frequent than to the CVNs in the triplet set.

The underspecification hypothesis makes a different prediction. On this account, the listener will not be able to find any complete lexical match with the nasal vowel, since the CVN is not specified as [nasal], and there is no lexically available CṼC. This suggests that if there is an increase in CVN responses (reflecting the absence of the CṼC competitor), this will be accompanied by an increase in CVC responses as well, since the CVCs are equally good fits to the available input, and will also benefit from the absence of CṼC competitors. It is only when the listeners hear the nasal consonant following the nasalised vowel that the correct analysis should become available to them.

In summary, the view of lexical access and selection that we advocate makes the claim that these processes are conducted with respect to abstract underspecified representations of lexical form. These are integrated segmental–featural representations where all predictable information is withheld, and where universal principles determine which distinctive information is specified. It is these underspecified representations, the content of the listener's recognition lexicon, that determine how the speech input is interpreted.

2. Method

2.1. Materials and design

Two sets of materials were constructed, for the Bengali and for the English parts of the study. We will describe first the Bengali stimuli.

The primary set of Bengali stimuli consisted of 21 triplets of Bengali words, each containing a CVC, a CVN, and a CṼC, where each member of the triplet shared the same initial oral consonant (or consonant cluster), and the same vowel (oral or nasal) but differed in final consonant (which was either oral or nasal). An example set is the triplet /kɑp/, /kɑm/, /kɑ̃p/. As far as possible the place of articulation of the word-final consonant was kept constant. The vowels [ɑ, o, æ, ɔ, e] and their nasal counterparts were used.

We also attempted to match the members of each triplet for frequency of occurrence in the language. Since there are no published frequency norms for Bengali, it was necessary to rely on the subjective familiarity judgements of a native speaker (the first author). Judgements of this type correlate well with objective measures of frequency (e.g., Segui, Mehler, Frauenfelder, & Morton, 1982).

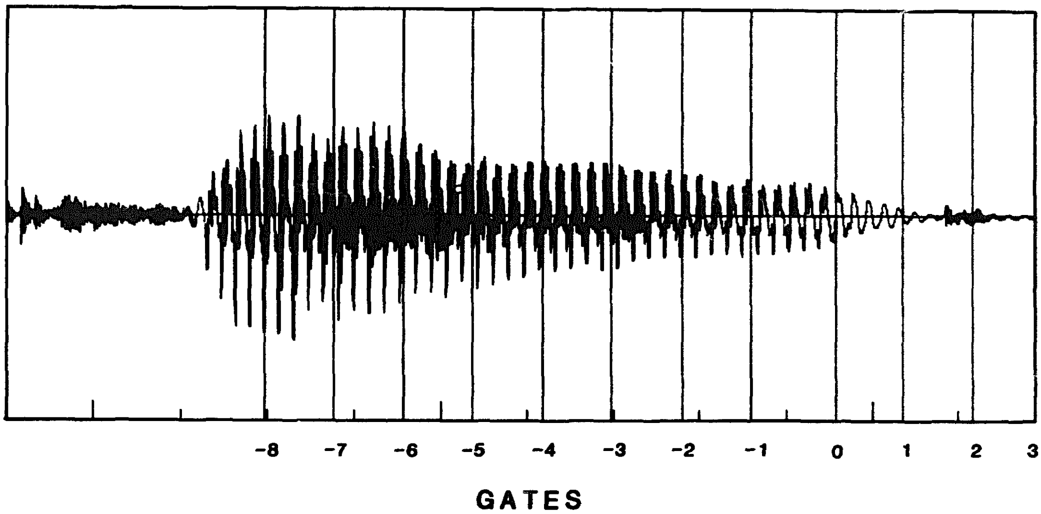
The second set of Bengali stimuli consisted of 20 doublets, containing matched CVCs and CVNs, where there was no word in the language beginning with the same consonant and vowel, but where the vowel was a nasal. An example is the pair /lom/, /lop/, where there is no lexical item in the language beginning with the sequence /lõ/. The absence of lexical items with the appropriate nasal vowels was checked in a standard Bengali dictionary (Dev, 1973). As before, place of articulation of the final consonant in each doublet was kept constant. We used the same vowels as for the triplets, with the addition of [i] and [u].

Given the absence of nasal vowels in English, only one set of stimuli was constructed. This was a set of 20 doublets, matched as closely as possible to the Bengali doublets in phonetic structure. We avoided pairs with high vowels, and used pairs with the same initial and final consonants as the Bengali pairs (in so far as the phonemic inventory of the two languages permitted). The pairs were matched for frequency, using the Kucera and Francis (1967) norms, with a mean frequency for the CVNs of 18.2 and for the CVCs of 23.4.

All of the stimuli were prepared in the same way for use in the gating task. The Bengali and English stimuli were recorded in a sound-attenuated booth, using a Nagra 4.2 tape-recorder and an AKG microphone, by native speakers of the respective languages. They were then digitised at a sampling rate of 20 KHz for editing and manipulation in the Max-Planck speech laboratory.

Each gating sequence was organised as follows. We wanted to be able to look systematically at responses relative both to vowel onset and to vowel offset. The first gate was therefore set, for all stimuli, at the end of the fourth glottal pulse after vowel onset. This first gate was variable in length. The gating sequence then continued through the vowel in constant 40-ms increments until the offset of the vowel was encountered. A gate boundary was always set at vowel offset, with the result that the last gate before vowel offset also varied in length for different stimuli. If the interval between the end of the last preceding gate and the offset of the vowel was less than 10 ms (i.e., not more than one glottal pulse), then this last gate was simply increased in length by the necessary amount. If the interval to vowel offset was more than 10 ms, then an extra gate of variable length was inserted. After vowel offset the gating sequence then continued in steady 40-ms increments until the end of the word. Figure 2 illustrates the complete gating sequence computed for one of the English stimuli.

Figure 2. *The complete gating sequence for the English word* grade. *Gate 0 marks the offset of the vowel.*



The location of the gates for the stimuli was determined using a high-resolution visual display, assisted by auditory playback. When gates had been assigned to all of the stimuli, seven different experimental tapes were then constructed. Three of these were for the Bengali triplets, and each consisted of three practice items followed by 21 test items. The tapes were organised so that each tape contained an equal number of CVCs, CVNs, and CṼCs, but only one item from each triplet, so that each subject heard a given initial CV combination only once during the experiment. A further two tapes were constructed for the Bengali doublets, again with three practice items followed by 20 test items, with members of each doublet assigned one to each tape. The final two tapes, for the English doublets, followed in structure the Bengali doublet tapes.

On each tape, the successive gates were recorded at 6-s intervals. A short warning tone preceded each gate, and a double tone marked the beginning of a new gating sequence.

2.2. Subjects and procedure

For the English materials, 28 members of the MRC Language and Speech Group subject pool were tested, 14 for each of the two experimental tapes. All subjects were native speakers of British English and were paid for their participation. For the Bengali materials, a total of 60 subjects were tested, 36 for the three triplet tapes, and 24 for the two doublet tapes. No subject heard more than one tape. The subjects were literate native speakers of Bengali, tested in Calcutta. They were paid, as appropriate, for their participation.

The same testing procedure was followed throughout. The subjects were tested in groups of two to four, seated in a quiet room. They heard the stimuli over closed-ear headphones (Sennheiser HD222), as a binaural monophonic signal. They made their responses by writing down their word-choices (each in their own script), with accompanying confidence rating, in the response booklets on the desk in front of them. The booklets were organised to allow for one gating sequence per page, and consisted of a series of numbered blank lines, with each line terminating in the numbers 1 to 10. The number 1 was labelled "Complete Guess" and the number 10 was labelled "Very Certain" (or the equivalent in Bengali).

The subjects began the testing session by reading a set of written instructions, which explained the task, stressing the importance of (1) making a response to every gating fragment and of (2) writing down a complete word as a response every time and not just a representation of the sounds they thought they heard – even if they felt that their response was a complete guess. They were then questioned to make sure that they had understood the task. The three practice items then followed. The subjects' performance was checked after each practice sequence, to determine whether they were performing correctly. The main test sequence then followed, lasting for 30–35 min.

3. Results and discussion

The questions that we are asking in this research concern the nature of the representation onto which the speech input is being mapped. Do the subjects behave as if they are interpreting the input relative to an abstract underspecified representation, or is the relevant representation one which is closer to the fully specified surface form of the word?

To answer these questions, the subjects' response sheets were analysed so as to provide a breakdown, for each item, of the responses at each gate. All scoreable responses were classified either as CVCs, CVNs, or C$\tilde{V}$Cs. Three triplets from the Bengali materials and five doublets from the English data had to be discarded. This was because one or more items in these sets were not identified correctly during the gating sequence – usually due to problems in recognising the initial consonant of the item. The subsequent analyses, therefore, are based on 17 Bengali triplets, 20 Bengali doublets, and 15 English doublets. We begin with the main sets of results, for the Bengali triplets and the English doublets.

3.1. Bengali triplets

The results are presented in two ways. Figure 3 gives the mean number of different types of response (CVC, CṼC, or CVN) to each type of stimulus, plotted across the five gates up to the offset of the vowel (Gate 0 in the figure) and continuing for five gates into the final consonant. The top panel of Figure 3 shows the responses to CVN stimuli, the middle panel the responses to CṼC stimuli, and the bottom panel the responses to CVC stimuli. The five gates from Gate −5 to Gate 0 cover a timespan ranging from 170 to 200 ms, depending on the length of the last gate before vowel offset. All items had sufficiently long vowels to allow for at least five gates preceding vowel offset.

The crucial data here are for these first five gates. Once the listener receives unambiguous information about the final consonant, then his response no longer discriminates between alternative theories of representation. What is important is how listeners respond to the vowel *before* the final consonant. To aid in the assessment of this we also include a summary of the responses over the first five gates. Table 2 gives the overall mean percentage of responses of different types to each of the three stimulus types.[9]

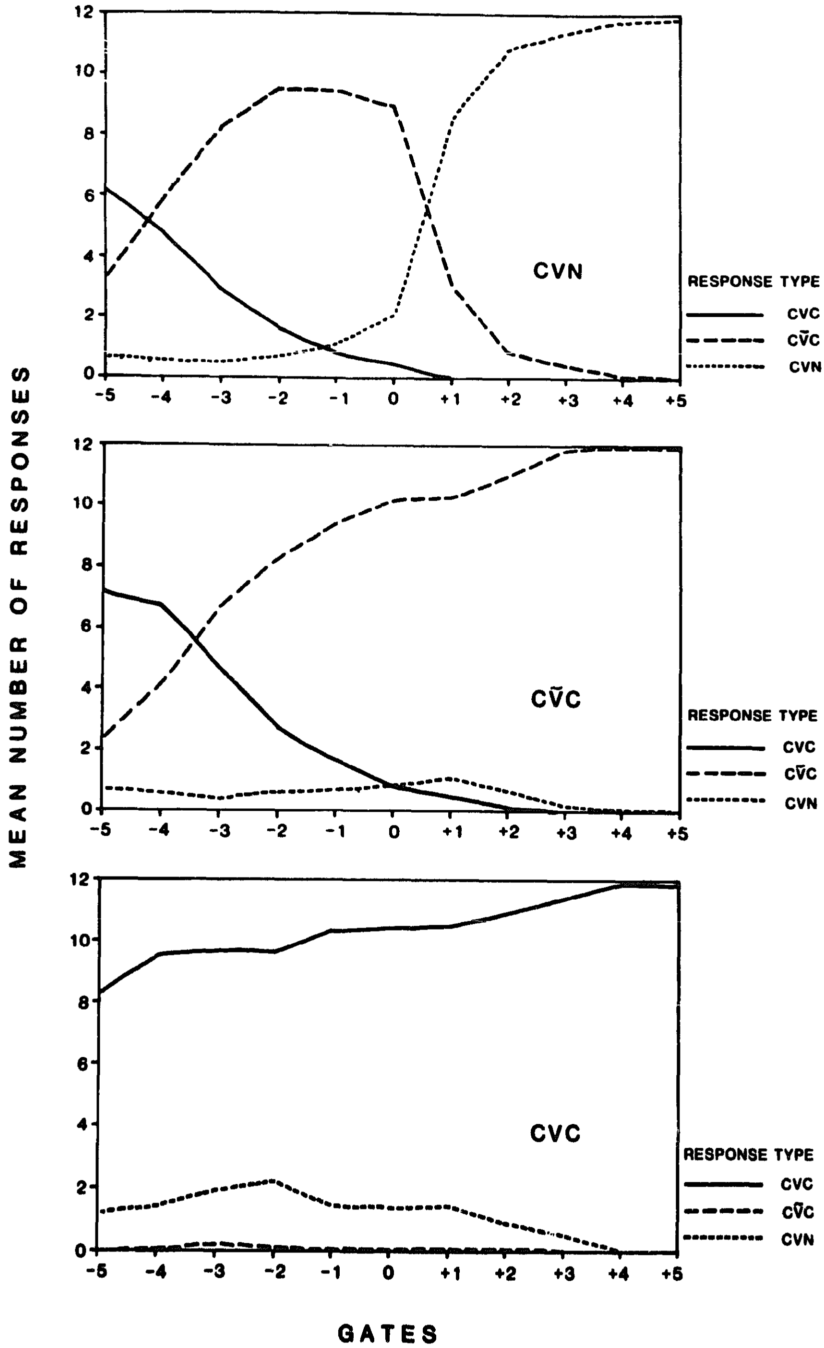
Consider first the listeners' responses to the nasalised vowels in the CVN and CṼC words. For both stimulus types, they interpret the presence of nasalisation as a cue that they are hearing a CṼC and not a CVN. The proportion of CVN responses to both CVNs and CṼCs stays at 5% or less until the last gate before vowel offset. Only at Gate 0, and then only for the

Table 2. *Bengali triplets: percentage responses up to vowel offset*

Stimulus	Type of response		
	CVC	CṼC	CVN
CVC	80.3	0.7	13.4
CṼC	33.2	56.8	5.2
CVN	23.5	63.0	7.9

[9]The rows in these tables do not add up to 100% because not all of the responses were scoreable – especially for the earlier gates, where subjects were more likely not to give a complete word as a response.

Figure 3. *Bengali triplets: mean percentage of different types of response (CVC, CṼC, or CVN) to each type of stimulus, plotted across the five gates up to offset of the vowel (Gate 0) and continuing for five gates into the consonant. The top panel gives the responses to CVN stimuli, the middle panel the responses to CṼC stimuli, and the bottom panel plots the responses to CVC stimuli.*



CVN stimuli, do we start to see an increase in CVN responses, as information starts to come in about the following consonant. At Gate +1, once it becomes clear that the listener is hearing a nasal consonant, there is an immediate switchover from CṼC responses to CVN responses.

The absence of CVN responses to CVN stimuli cannot be attributed to any lack of nasalisation of the vowel in these materials. The presence of the CṼC responses demonstrates that the vowel was perceived as nasalised, and the close parallel between the CṼC response curves for CVN and CṼC stimuli shows that the degree of perceived nasalisation was equivalent for both stimulus types. The overall proportion of CṼC responses was very similar for the CVN stimuli (63%) and CṼC stimuli (57%).

This is a pattern of results consistent with the abstract underspecification hypothesis (see Figure 1), where nasalised vowels are completely matched by the representations of CṼC words in the recognition lexicon, which leads to their dominance over CVN words. CVN words are nonetheless still present in the cohort as candidates, since there is no mismatch between the signal and the vowel quality features underlyingly specified for CVNs, and on a small proportion of trials they surface as gating responses (similar considerations apply to CVC words). On a surface representation hypothesis, in contrast, vowel nasalisation in Bengali should be perceptually ambiguous, matching equally well to the representation of CVNs and CṼCs in the recognition lexicon.

Turning to the CVC stimuli (oral vowels followed by oral consonants), performance here is dominated by CVC responses. Already at Gate −5 the proportion of CVC responses is higher than for the CVN or CṼC stimuli, and remains fairly steady, at around 80%, for the next five gates. Consistent with this, there are essentially no CṼC responses at all. In striking contrast is the relatively high frequency of CVN responses over the first five gates. Listeners produce more than twice as many CVN responses to CVC stimuli as they do to either CVN or CṼC stimuli.

This is hard to explain on a surface representation account. If CVNs are represented, like CṼCs, as containing a nasalised vowel followed by a nasal consonant, then neither CVNs nor CṼCs should be produced as responses to oral vowels. There should be no reason for CVN responses to be more frequent to oral vowels than to nasalised vowels, nor should there be any asymmetry between CVN and CṼC responses.

On the underspecification hypothesis, the pattern of results for CVC stimuli follows directly from its claims about feature specification in the recognition lexicon (see Figure 1). These claims rule out CṼCs as responses to CVCs, but permit CVNs. In fact, as far as the access system's goodness-of-fit computation is concerned, it is just as appropriate to give CVNs as responses to oral vowels as it is to give CVCs. The fact that there is, nonetheless, a preponderance of CVCs, presumably reflects the distributional facts of the language. If words with underlying oral vowels are more commonly CVCs than CVNs, then this would lead to a preference for CVC responses when oral vowels were heard.

The difficulty in evaluating this possibility is that there are not, as far as we know, any extensive language statistics available for Bengali. Chatterjee (1975) reports a count of phoneme frequencies, based on a small written corpus, in which nasal vowels are very infrequent relative to their oral counterparts. What we cannot establish from this is the relative frequency, however, of CVNs as opposed to either CVCs or CṼCs, and these are the proportions that are crucial.

As a preliminary remedy of this, we carried out a sampling of the distribution of monosyllabic CVC, CVN, and CṼC words in a dictionary of Bengali (Mitra, 1968). Randomly sampling one-third of the possible combinations of word-initial CVs allowed by the language, we examined a total of 277 monosyllables. Of these, 67% were CVCs, 17% were CṼCs, and 16% were CVNs. In so far as this allows us to generalise to the language as a whole – and assuming that type frequency correlates with token frequency – there are two points we can make.

The first concerns the imbalance in the distribution of CVC words and CVN words. This corresponds closely to the imbalance in listeners' responses to the triplet CVCs. In the language sample, 81% of the underlyingly oral vowels occur in CVCs (as opposed to CVNs), and this matches the 80% CVC responses to CVC stimuli over the first five gates. The second point, returning to the CVN and CṼC stimuli, concerns the possibility of a frequency bias affecting responses to nasalised vowels. Since CṼC words are not significantly more common than CVN words, there is no reason to attribute to a frequency bias the strong preference for CṼC responses when a nasalised vowel was heard. In terms of the listener's linguistic experience, vowel nasalisation seems to signal, with approximately equal probability, nasal vowels preceding oral consonants and oral vowels preceding nasal consonants. This balance in the distribution of the two types of monosyllable held not only for the language as a whole, but also for the specific word-initial CVs and CṼs that we used in the triplet stimulus set.

Before we move on to the English data, there is one further issue specific to Bengali that we need to deal with. This concerns the possible role of orthographic factors in accounting for the results. In Bengali script, nasal vowels are marked in the orthography by a special diacritic, known as the *chandrabindu* (Chatterjee, 1975; Klaiman, 1987). It is possible to make the argument, therefore, that since literate Bengali speakers know that nasal vowels are written with the nasal diacritic, they will have a strong preference, when they hear a nasalised vowel, to give as a response a word written with the diacritic. In contrast, when they hear an oral vowel, the orthographic pressure works in the opposite direction, to rule out any responses which contain the nasal diacritic.

There are a number of problems with this account. The first is simply that it misrepresents the perceptual experience of the subjects in the gating task. The subjects' consistent report was that when they heard a consonant followed by a nasalised vowel, they interpreted this as a nasal vowel followed by an oral consonant. If they were hearing the sequence [bã] from the CVN *ban*, then they were convinced that they were hearing the CṼC *bãk*. And since they knew how to spell correctly in Bengali, they wrote down the CṼC they thought they were hearing using the nasal diacritic. If orthography is playing a role here, it does not seem to be one of actually determining what the response was. Rather, the listeners heard the gated fragments as being instances of a particular word-type, and the orthography only came in later.

A second problem is that it is not the case that the diacritic is always used for underlying nasal vowels. The orthographic diacritic is simply left out in the redundant context where there is a contiguous nasal consonant available, both for underlying nasal vowel segments as well as oral ones. Thus, underlying nasal vowels (stem finally) when followed by a nasal consonant would not carry a diacritic, whereas the same vowel would be marked as nasal when an oral consonant follows. The verb /chõ/ "touch" illustrates this point. /chõ+ʃ/ (second person, familiar) carries the diacritic, while in /chõ+n/ (second person, honorific) there is no nasal marker on the vowel in spite of the fact that the underlying vowel is nasal. Therefore, the argument that the predominance of underlying nasal responses are due to the presence of the special diacritic to mark nasalisation cannot be true since these vowels themselves are not always written with their orthographic marker – the marking is conditioned by what follows.

A third, and major problem is that an orthographic explanation of the results seems to presuppose the truth of something like our phonologically based claims about lexical representations. The fact that the orthography does distinguish nasal vowels preceding oral consonants from nasalised oral vowels preceding nasal consonants is itself evidence for a distinction in the mental representation of the two types of phonetically nasal segments. So while the presence of the nasal diacritic in the orthography might serve to accentuate, or make more explicit, the distinction between underlyingly nasal and oral vowels, it cannot be seen as itself being the source of this distinction. In all languages, orthography reflects the underlying structure of the language, not the other way around; (for a related discussion on orthography see Chomsky & Halle, 1968, p. 49).

Finally, the orthographic account cannot apply to the English data, since English orthography does not mark vowel nasalisation in the same way. It is to these English results that we now turn.

3.2. English doublets

Figure 4 plots the responses across gates to the English materials, showing the number of responses of different types to the two stimulus sets, with CVN stimuli in the upper panel and CVC stimuli in the lower (note that the English

Figure 4. *Mean percentage of different types of responses to the two English stimulus sets across gates, Gate 0 marking the vowel offset. Responses to CVN stimuli are plotted on the upper panel and to CVC stimuli in the lower panel.*

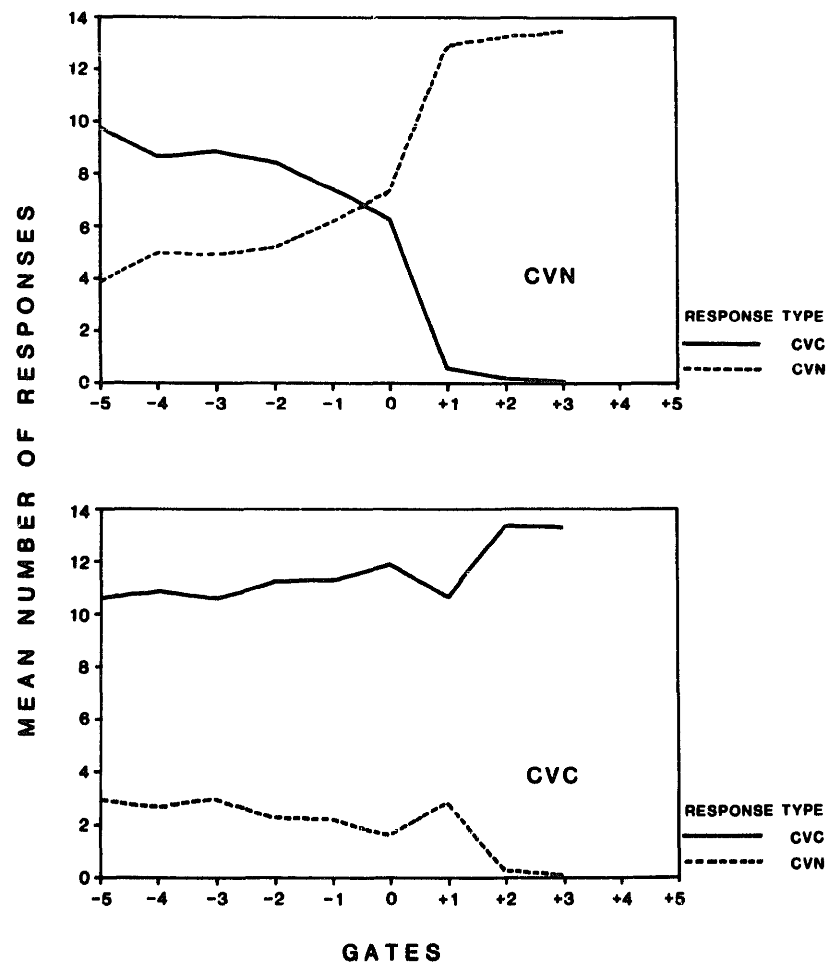


Table 3. *English doublets: percentage responses up to vowel offset*

Stimulus	Type of response	
	CVC	CVN
CVC	83.4	16.6
CVN	59.3	40.7

scores are out of a total of 14, as opposed to 12 for Bengali). Table 3 summarises the overall percentage of responses of each type for the five gates up to vowel offset.

For the CVN stimuli, there is already a relatively high proportion of CVN responses at Gate −5, indicating an early onset of nasalisation. These responses increase steadily to vowel offset. This is as predicted by both accounts. On the surface representation hypothesis, CVN words are represented with nasalised vowels, and can begin to be discriminated from CVC words as soon as vowel nasalisation can be detected. On the abstract underspecification hypothesis, vowel nasalisation in English is interpreted as evidence about the properties of the following segment, so that as soon as nasalisation is detected listeners will start to produce responses where the following segment is a nasal consonant (cf. Fowler, 1984; Warren & Marslen-Wilson, 1987).

Turning to the CVC stimuli, the overall pattern of results parallels the results for the Bengali triplets. There is the same overall proportion of CVC responses, and the number of CVN responses, at 17%, is similar to the 13% found for the Bengali CVCs. This is a pattern which is problematic for a surface representation account but predictable on the underspecification hypothesis, for the same reasons we discussed above. Here also, we find that the relative proportions of CVC and CVN responses reflect the distributional properties of the language. For a sample of 3058 monosyllables in English, 14.8% were CVNs – closely paralleling the 16.6% CVN responses to English oral vowels.

The underspecification of oral vowels in the recognition lexicon is reflected in the uninformativeness of the absence of nasality in guiding lexical choice (for a language where there are no underlyingly nasal vowels). Hearing more of an oral vowel in English does not significantly increase the number of CVC responses or decrease the number of CVN responses. The slight drop off over the first five gates (Figure 4) reflects the appearance of cues to the place of

Table 4. *CVN responses to English CVC stimuli: place effects across gates (percentage response)*

	Gates							
	−5	−4	−3	−2	−1	0	+1	+2
Correct place	12.0	14.5	15.5	11.5	13.5	10.0	21.0	1.5
Incorrect place	9.5	5.0	6.0	5.0	2.5	1.5	0.0	0.0

articulation of the following consonant, rather than the accumulation of cues to orality. As we have shown elsewhere (Warren & Marslen-Wilson, 1987, 1988), place of articulation cues start to affect lexical choice about 50 ms before vowel offset. Some of the CVN responses produced to CVC stimuli at the earlier gates do not share place of articulation with the CVC being heard (for example, giving *bang* as a response to *bad*). It is these responses that drop out as vowel offset approaches, as Table 4 illustrates. This table lists the CVN responses to CVC stimuli, sorted according to the correctness of the place of articulation of the response. The lack of change in correct place responses over the five gates to vowel offset (Gate 0) emphasises the uninformativeness of the absence of nasality. The listener will only stop producing CVNs as responses when it becomes clear that the following consonant is also oral.

This brings us to a further aspect of the results, namely, the *increase* in CVN responses to CVC stimuli at Gate +1, after vowel offset. This is visible in Figure 4, and comes out very clearly in Table 4, where the percentage of CVN responses with correct place increases sharply at Gate +1, and falls back immediately afterwards. This is a very consistent pattern, with only three out of the 15 CVCs showing a drop in CVN responses at Gate +1.

The source of this effect is easily traced. Most of the CVC stimuli ended in voiced stops or affricates. This meant that Gate 0, set at vowel offset, fell at the onset of the period of closure preceding the release of the final consonant. For almost all these stops, Gate +1, occurring 40 ms later, also fell before the onset of the release burst – an example of this can be seen in Figure 2. The consequence of this is that the extra signal information the listeners acquired as they went from Gate 0 to Gate +1 was normally a stretch of pre-voicing; 40 ms of vocal murmur. This does not provide definite information about the manner of articulation of the consonant being heard.

What this means, then, is that listeners who have heard a word like *crowd* or *trade* as far as Gate +1 will have in their possession certain information about the properties of the final consonant. This information limits the possibilities to words ending with voiced consonants, most probably with a specific place of articulation. Looking at the change in gating responses at Gate 1, we see that this is sufficient to exclude a variety of different possibilities – including the possibility that what the listener is hearing is an open CV, such as the word *tray*. But it is not sufficient, however, to exclude the possibility that what the listener is hearing is a CVN. Nor is there any reason why it should. The pre-release vocal murmur is evidence that a voiced consonant is being heard, and this includes the class of nasal consonants in English. The possibility that the final consonant is nasal is only conclusively banished by the information in Gate +2, containing an oral rather than a nasal release.

This is a pattern of behaviour that is difficult to explain on a surface representation hypothesis. It is also difficult to explain on any hypothesis which does not allow for some form of featural – or at least non-segmental – mapping onto an underspecified representation at the lexical level.

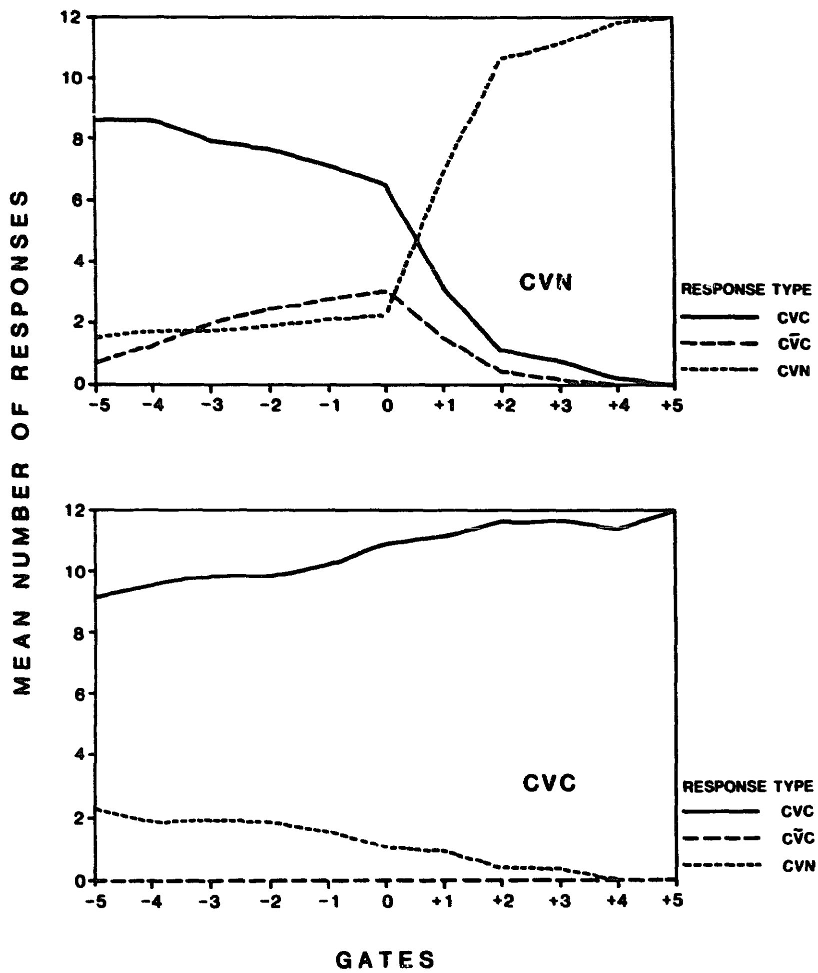
An important aspect, finally, of the results for the English doublets is that they provide evidence for the generality of the claims we are making here. Despite the contrasting phonological status of nasality in the Bengali vowel system as opposed to the English, both languages appear to treat oral vowels in the same way, and with similar consequences for the ways in which the speakers of these languages are able to interpret the absence of nasality in a vowel. Although vowel nasalisation has a very different interpretation in Bengali than in English, leading to exactly opposite perceptual consequences, the presence of an oral vowel leads to very similar ambiguities for listeners in both languages. This is because, in both languages, CVNs are underlyingly oral, and because, in both languages, oral vowels are underspecified in the recognition lexicon.

3.3. Bengali doublets

The Bengali doublets provide an additional test of the two representational hypotheses. These were the stimulus sets composed of Bengali CVCs and CVNs, where there was no CṼC in the language that shared the same initial consonant and vowel. Figure 5 gives the results across gates, showing the number of responses of different types to the two sets of stimuli, with the CVN stimuli in the upper panel and the CVC stimuli in the lower panel. Table 5 summarises the overall mean percentage of responses of each type for the five gates leading up to vowel closure.

The CVC stimuli elicit the same response pattern as we found for the triplets. There are no nasal vowel responses, an average of over 80% CVC responses, and the same relatively high percentage of CVN responses, reach-

Figure 5. *Bengali doublets: mean percentage of different types of response (CVC, CṼC, or CVN) plotted across gates, Gate 0 marking the offset of the vowel. The upper panel gives responses to CVN stimuli and the lower panel the responses to CVC stimuli.*



ing nearly 15%. As we argued earlier, this is consistent with the abstract underspecification hypothesis (see Figure 1). On the surface representation hypothesis, these CVN responses to words containing oral vowels can only be regarded as mistakes.

The CVN stimuli elicit a quite different response pattern. Compared with the responses to the triplet CVNs (Table 2), there is an increase over the first five gates both in CVC responses (from 24% to 65%) and in CVN responses (from 8% to 16%). In addition to this, listeners also produced a significant number of CṼC responses (averaging 17%), despite the fact that no CṼC

Table 5. *Bengali doublets: percentage responses up to vowel offset*

Stimulus	Type of response		
	CVC	CṼC	CVN
CVC	82.6	0.0	14.7
CVN	64.7	17.0	15.6

items were available to them in the lexicon of the language. In fact, for Gates −3 to 0, they produced more CṼC responses than they did CVN responses.

Instead of producing the CVN that was lexically available, the listeners produced as responses CṼCs that were phonologically closely related to the consonant–vowel sequence they were hearing. They either produced real words, whose initial consonant or medial vowel deviated minimally from the actual stimulus, or else they produced "nonsense" words.[10]

This pattern was uniform across subjects and across stimuli. All 24 of the subjects made at least some CṼC responses to the CVNs, and CṼC responses were made, across subjects, to 15 out of the 20 of the CVN stimulus items.

This striking reluctance to produce a CVN response, even when the input is apparently unambiguous, seems inexplicable on a surface representation hypothesis. If the vowels in CVNs are represented in the lexicon as [+nasal], then why does a listener not produce a CVN as a response when he hears a nasalised vowel – and when, indeed, the lexicon of the language does not permit it to be anything else? In contrast, the abstract underspecification hypothesis provides a plausible account of the entire pattern of responses to the doublet CVNs.

First, there is the increase, relative to the Bengali triplet data, in CVC and CVN responses. The information in the signal matches equally well to the vowel representations for CVNs and CVCs. But, unlike the triplet CVN situation, there is no longer a better fitting competitor – namely, a lexically represented CṼC which matches the nasalisation in the signal in addition to the other vowel quality features. This makes the competitor situation for the CVN doublets more similar to the CVC situation, where both CVC and CVN

[10]Responses were defined as "nonsense words" if they were not lexical items in standard spoken Bengali. Many of the responses classified in this way were potential real words in other dialects of Bengali, or even in other languages (such as Hindi) likely to be known to some of the subjects.

words are equally good matches to the vowel information, and where distributional preferences seem to determine the ratio of CVC and CVN responses.

The effects of the absence of a nasal competitor is also reflected in the different patterns of CVC responses over gates to nasal stimuli. For the triplets, where there is a nasal competitor lexically available, CVC responses were 28% in total, dropping off sharply from 48% at Gate −4 to 5% at Gate 0 (see Figure 3). This drop off reflects the increasingly strong cues to nasalisation in the last two gates before vowel offset. This pattern contrasts with the 65% CVC responses to nasal stimuli in the doublets, where there is no lexically available nasal competitor sharing the same vowel qualities. Here the percentage of CVC responses at Gate −4 is 72%, and this drops off relatively little over gates to 55% at Gate 0 (see Figure 5).

Note, furthermore, that this small decrement over gates in CVC responses is not because of an increase in CVN responses. These stay essentially unchanged from Gate −4 (at 15%) to Gate 0 (at 19%). The increase in nasal responses comes from the CṼC responses, going from 10% at Gate −4 to 25% at Gate 0. The perceptual representations of CVN and CVC vowels do seem to be truly indifferent to the presence or absence of nasalisation in the signal.

3.4. Variability in nasalisation

Before moving to the general discussion, we need to consider a different explanation for our results – especially for those involving CVN words – than the one we have argued for so far. This alternative account is based on the argument that there is an intrinsic variability in the phonetic realisation of derived nasalisation, especially in comparison with nasalisation that comes directly from the underlying form. Nasalisation in CVN words is not an intrinsic property of the vowel being produced. It spreads, instead, from the following segment. There is therefore the opportunity, the argument goes, for a variability (or gradience) in the degree of this spreading – and, therefore, in the degree of resulting nasalisation – which does not occur for underlying nasals.

On a surface representation hypothesis, the listeners' representation of a given word-form will reflect their experience with the phonetic realisation of this item. If there is variability in the phonetic expression of derived nasalisation, then there will be a less rigid relationship between the presence or absence of nasalisation for CVNs than for CṼCs. This would have two consequences for listeners' responses to phonetically oral or nasal vowels. Vowel nasalisation will be a more reliable indicator of an underlying nasal vowel than of derived nasalisation. The listener should therefore prefer CṼC words to CVN words as responses to phonetic nasalisation. Conversely, for phonetically oral vowels, the absence of nasalisation would more reliably exclude CṼC words as possible responses than CVNs. The listeners will have heard at least some proportion of CVN words with oral vowels, and this will be reflected in their responses to CVC stimuli. They will be much less likely, however, to produce CṼC words as responses to phonetically oral vowels.

There are a number of problems with this account. There is, first, no evidence that derived nasalisation is more variable in its phonetic expression than underlying nasalisation. The implied contrast here is between the view of nasal assimilation as a phonological rule, applying as an obligatory part of the realisation of a given word-form, as opposed to the view that it is a phonetic implementation rule, which may have a gradient or variable output (cf. Liberman & Pierrehumbert, 1984). In this respect, nasalisation seems to pattern with other assimilation rules which behave like phonological processes rather than gradient phonetic effects. Kiparsky (1985), for instance, characterises the assimilation rule which voices final obstruents in English (dog[z] but not dock[s]) as phonological, on the grounds that it applies whether or not listeners monitor their speech – speakers do not have the option of producing an [s] following [dog]. In any case, as Kiparsky also points out, a process can be gradient in its articulatory expression but still be perceived categorically. A well-known example of this is VOT (voice onset time), where perception is categorical over a wide range of VOT values on either side of the category boundary.

The second point is that even if there were gradience in the expression of vowel nasalisation, and even if this gradience was greater for derived than for underlying nasals, this would still not explain the results. Variability in derived nasalisation would presumably take the form of variability in the timing with which the vowel started to become nasalised. In comparison to words with underlying nasal vowels, CVNs would be less likely to be nasalised early in the vowel. If this statistical gradient were reflected in listeners' representations of CVN words, then what would this predict for performance?

Consider, first, the nasalised vowels in Bengali, where surface nasalisation can be either derived or underlying. For the triplets, the variability account predicts more CVN responses later in the vowel. If the reason that listeners do not produce CVNs early in the vowel is because of the lower probability that CVN vowels will be nasalised at this point, then their production of CVN responses should increase later in the vowel, as the probability increases that the perceived nasalisation is associated with a CVN. This effect should be especially strong for the Bengali doublets, since here there is no competition from words with underlying nasal vowels. But we found no sign of such an increase. For the doublets, in fact, the most visible increase over the final gates preceding the vowel was in the false CṼC responses.

For oral vowels, in both Bengali and English, the variability account predicts a decrease in CVN responses as the end of the vowel approaches. The likelihood that the listener is hearing a CVN with an oral vowel is greatest earlier in the vowel. The presence of CVN responses at the earlier CVC gates would simply reflect the fact that some CVNs, in the listeners' experience, had been produced with oral vowels until relatively late in the vowel. As more of the vowel was heard, these responses should decrease, since it is becoming increasingly unlikely that what the listener is hearing is a CVN with late nasalisation. If anything, however, it is the opposite pattern that we observe. For the English CVCs, in fact, we see a marked increase in CVN responses at a very late point in the word (see Table 4). This is quite inconsistent with the variability story.

We can reject, therefore, the variability hypothesis as an explanation for the results here. Not only is there no clear evidence that derived nasals are more variable than underlying nasals in their phonetic expression, but even if they were, this would make the wrong predictions about the detailed pattern of results.

3.5. Summary and overview

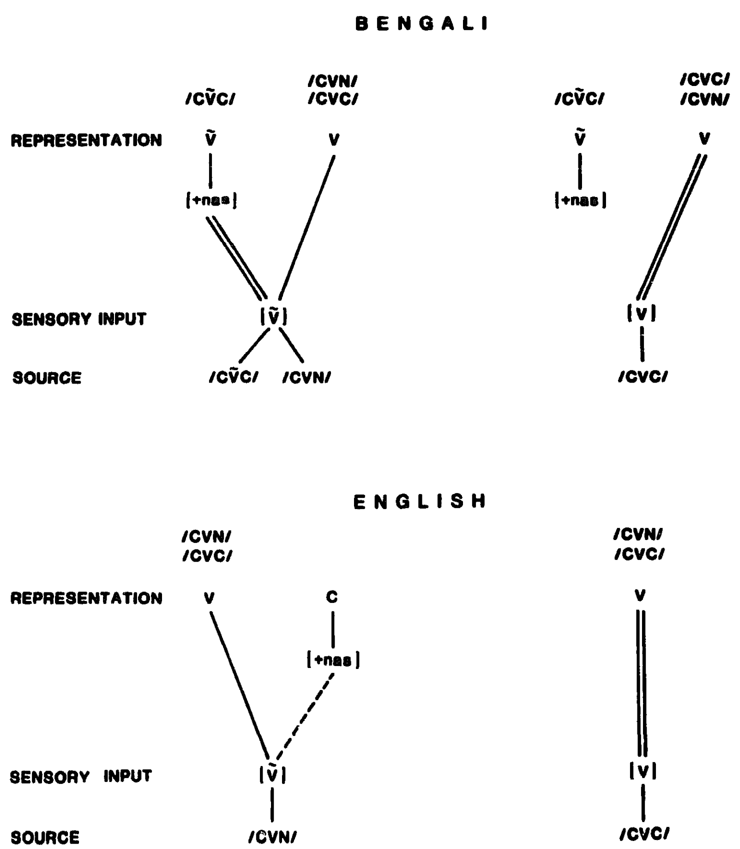
Before moving on to the general discussion, we will summarise the main points so far. The crucial findings are the following:

(1) The strong preponderance of nasal vowel (CṼC) responses to stimuli with nasalised vowels (CṼCs and CVNs) in Bengali.
(2) Almost no nasal vowel (CṼC) responses to stimuli with oral vowels (CVCs) in Bengali.
(3) More CVN responses to CVNs than to CVCs in English, but more CVN responses to CVCs than to either CṼCs or CVNs in Bengali.
(4) The same pattern of responses to stimuli with oral vowels (CVCs) in both languages. Both CVN and CVC responses are given, in proportion to their distribution in the two languages.

These results are very much along the lines predicted earlier in this paper (see section 1.5). The asymmetry in the responses to nasal and oral vowels, the difference in responses to CVN stimuli in Bengali and English, and the similar response patterns to CVC stimuli in both languages can only be explained if we assume an underspecified representation of the sort we have been advocating, rather than a fully specified surface representation. Figure 6 gives a complete overview, illustrating the properties of lexical form representations in the recognition lexicon for both languages, together with the corresponding processing interpretations. The figure shows how listeners match particular types of sensory input to particular representations. The top half of the figure covers Bengali and the bottom half English. On the left is the representation and matching of phonetically nasal stimuli, and on the right the phonetically oral stimuli.

The sensory input is given with its corresponding source – vowel nasalisation in Bengali can come from either CṼC words or CVN words, while in English its only source is CVN words. In describing the listeners' interpretation of these various inputs, the double lines show a complete match between the information in the signal and the representation in the recognition lexicon. The single lines indicate the cases where the match is not complete, but

Figure 6. *Processing and representation of oral and nasal vowels in Bengali and English. Double lines indicate a complete match between input and representation, single lines indicate an incomplete match but without mismatch, and dotted lines indicate linking to the nearest available appropriate feature (see text).*



where there is no positive mismatch, because of the effects of underspecification – this is what permits words with oral vowels to be given as responses to both phonetically oral and nasal vowels. The absence of any connection, in contrast, indicates the cases where a specified feature in the lexical representation does mismatch with information in the signal – this is why CṼCs are not given as responses to oral inputs. The dotted line, finally, indicates the situation for English CVNs, where vowel nasalisation is perceptually linked during lexical access to the following nasal consonant.

This pattern of representations and processing interpretations is consistent not only with the claims about the content of lexical representations that were the direct target of this research, but also with our supporting claims about the processing architecture of the system – in particular, for the claim that featural information is mapped directly onto lexical representations without any intervening process of segmental labelling. We now turn to a discussion of the implications of these results for the issues of abstractness and variability which motivated our original approach to lexical access and representation.

4. General discussion

The crucial point about our research here is that it makes it possible to systematically link the investigation of how sensory information is used over time to an explicit and detailed theory of what information is actually represented in the mental lexicon. If the recognition lexicon only contains abstract underspecified information, then what consequences does this have for the detailed operations of lexical access and selection – where we understand this to be the mapping of featurally specified information, simultaneously and in parallel, onto all representations for which the input provides a match? In particular, what are the consequences for the system's treatment of phonologically conditioned variability in the surface realisation of a given word-form?

4.1. Abstractness, underspecification, and variation

We will focus here on the effects of assimilatory processes, which are not only cross-linguistically very widespread, but which also have been central to our research. The process that we investigated involved the spread of the nasal feature to an oral vowel from a following nasal consonant. In both English and Bengali a vowel can surface as nasal or oral depending on the phonological properties of the following consonant. In Bengali, for example, the phonetic manifestation of a stem vowel such as /kʰɑ/ "eat" will be nasal when preceding a nasal consonantal affix and oral otherwise. The stem vowel will be heard as [ɑ̃] in /kʰɑ+n/ (second person, honorific) and as [ɑ] in /kʰɑ+ʃ/ (second person familiar). This is a *neutralising* process, since it neutralises an underlying distinction between oral and nasal vowels. The sequence [kʰɑ̃] is therefore ambiguous, since it might be derived from a real /ɑ̃/, as in words like [kʰɑ̃rɑ] "hatchet". In English, of course, assimilation of nasality from a following consonant (as in words like [bĩn]) is *allophonic*, since there is no nasality contrast in the language for vowels. There are, however, plenty of assimilatory processes in English which are neutralising – for example, the assimilation of place of a nasal consonant to the following consonant (as in cases like *hand bag* > [hæ̃mbæg]).

How does the access process deal with these types of phonologically induced variation? In our model, the problem is resolved relative to the representation of word forms in the recognition lexicon, in ways which depend on the absence of predictable information from the lexicon, on whether the surface feature involved corresponds to phonologically marked or unmarked underlying values, and on the processing consequences of these factors.

4.1.1. Neutralising variation

We consider first the situation where the system encounters a surface feature which corresponds directly to an underlying marked value, and where there is also a matching lexical entry – as in the [kʰɑ̃] example. This will always be interpreted as evidence for the word containing the specified feature. Vowel nasality cannot be (and is not) interpreted in Bengali as coming from a following nasal consonant. This type of neutralising variation – the fact that the nasalised vowel in [kʰɑ̃] might be derived from the oral /ɑ/ – is, in effect, dealt with by ignoring it as a possibility. This means that sequences like [khã] will be treated by the system as if they are unambiguous. The cost associated with this is that listeners will initially misanalyse nasalisation which is derived from underlying nasal vowels. There are two reasons why this is not a serious problem for the system.

The first of these derives from the way the cohort model works. When a nasalised vowel is heard – say, the sequence [kʰɑ̃] – this maps perfectly onto the lexical representation for a real word like [kʰɑ̃dʒ] "groove", and this is the listener's preferred response. However, it also fits the lexical representation for the stem kʰɑ with its underlying oral vowel. Vowel quality information matches equally well for both entries, and the absence of a match for nasality should not affect the goodness-of-fit computation for the stem kʰɑ. The reason the underlyingly nasal word is initially preferred is because it provides a more complete match to the lexically relevant information in the sensory input (see Figure 6).[11]

But this is only a preference. In particular, the failure to match the nasality of the vowel does not mean that the words with oral vowels are ruled out as candidates. There is no reason why they should be, on current views of the cohort model (Marslen-Wilson, 1987, 1989), since they do not mismatch the stimulus in any way. Thus, although there will be a strong preference for the underlying nasal interpretation, the word-candidates corresponding to the alternative interpretation will remain active in the system until the nature of the following consonant becomes clear. If the following consonant is nasal, no back-tracking or re-computation is then needed.

The second reason for the workability of the strategy is because of the constraints that it imposes on what phonological processes of assimilation can be permitted to do. Neutralising assimilatory processes should never change an underlyingly marked value to an unmarked value on the surface.[12] If this were to happen, then it would cause the correct candidate to be rejected in the recognition process. We saw earlier that words with underlying nasal vowels are not given as responses to phonetically oral vowels. Nasality (a marked feature) can spread to a neighbouring oral vowel, but orality cannot spread to a neighbouring nasal vowel. By the same token, assimilatory processes should never shift from one marked value to another marked value. Velar place, for example, should not assimilate to labial place (assuming these to be the marked values, and coronal place the unmarked value).

In summary, phonological processes should never create a situation where evidence for a feature on the surface is inconsistent with the specified features in the lexical entry intended by the speaker. If they were to do so, then the listener would necessarily misidentify, or fail to recognise, the word being heard.

So far we have discussed the case where assimilatory processes generate a surface feature which matches a feature underlyingly specified in a lexical entry. What happens when the marked surface feature does not find a lexical match? As far as the correct candidate is concerned, the outcome is exactly

[11]The preference for underlying nasal responses here and elsewhere in the results (for example, non-word and real word responses in the Bengali doublets) might also be explained in terms of some system of priorities among different features, such that the feature nasal overrides the other qualitative features of the target vowel. Only a systematic study of the relationship between different features can tell us if there is a hierarchy which plays a role in lexical access. Whether there is any relationship between the hierarchies reflected in models of feature geometry (cf. Clements, 1985) and processing is a topic for future research.

[12]This holds, in general, for the productive application of phonological assimilatory rules which are characteristically feature-spreading in nature.

the same. Consider the Bengali doublets. In Bengali, where vowel nasalisation is distinctive, the appearance of nasality on the vowel can only be interpreted as a fact about that vowel, and not as a cue to the following consonant. This remains true whether or not an appropriate lexical item is available. The actual perceptual outcome under these conditions will reflect the properties of cohort-based competitive processing. Because there is no lexical item that fully matches all the vowel quality information, no single competitor will have a clear superiority until the consonant is heard. If listeners are required to make a response while still hearing the vowel, they will either select some candidate that is a good partial match, or else devise a non-word that fully matches the available phonetic information.

Note, however, that when the final consonant is found to be nasal, the correct CVN can be identified in exactly the same manner as when there is a CṼC competitor. The CVN is an active candidate, and can be immediately identified without backtracking.

4.1.2. Allophonic variation

So far we have discussed variation that involves the neutralisation of underlying contrasts. What happens when the assimilatory process is allophonic? There are two cases to consider here. The first is the case exemplified in our research by nasal assimilation in English, where underlyingly oral vowels acquire nasality from a following nasal consonant. The feature [nasal] is indeed distinctive in English, but only for consonants. There is no oral/nasal distinction for English vowels, and this is why vowel nasalisation is allophonic: it does not neutralise an underlying contrast.

This means, first of all, that the presence or absence of nasality is irrelevant to the goodness-of-fit computation for any vowel in the language, and therefore for any word containing this vowel. There is no possibility here of interpreting vowel nasalisation as a property either of the vowel in the correct candidate or in any of its competitors.

Secondly, however, because nasal *is* distinctive for consonants in the language, the nasal feature is assigned perceptually to the following segment, restricting it to the natural class of all nasal consonants. This is what we see in the responses to the English CVN stimuli. Allophonic variation involving distinctive features in the language is therefore immediately informative for the listener in a way that neutralising variation will not be.

The second type of allophonic variation which we should also mention, although it is not assimilatory in nature, involves prosodically conditioned effects which do not use distinctive features in the language. As we noted earlier, phonological processes in several languages are sensitive to prosodic structures like syllables and feet. Nevertheless, the abstract representation of a lexical item in the recognition lexicon does not contain any prosodic structures; they are predictable and, therefore, assigned by rule. In our model, therefore, neither the structures, nor the associated featural information, should be relevant for lexical access. Consider the well-known example of syllable-initial aspiration of voiceless stops in English. This is a predictable and regular allophonic process. Yet neither the syllable structure nor the aspiration is present in the recognition lexicon. How does the system deal with this?

Unlike vowel nasalisation (and parallel phenomena), aspiration in English is not an assimilation phenomenon. There are no aspirated segments in the language, so that aspiration cannot be spread from a nearby segment. As far as the lexical access system is concerned, aspiration can only be an enhancing feature for the class of voiceless stops; its presence emphasizes the voiceless character of the consonant. In the recognition lexicon itself, of course, the voiceless consonants will always be unspecified for aspiration. A voiceless stop has various phonetic attributes which change in different contexts, and all of the resulting allophonic variants should match equally well to the representation in the recognition lexicon.

Note that we are not claiming that prosodic structures have no role in language comprehension. In the normal language situation, words form part of syntactically organised strings, and the prosodic structure of these strings controls the assignment of stress and intonational contours, and may also have other functions in parsing and interpretation. Nonetheless, the processing system involved in accessing the recognition lexicon is insensitive both to these structures, and to the cues to these structures in the speech signal. Church (1987) and Cutler (1989), for instance, make use of allophonic cues and stress for segmentation purposes, but they too do not claim that this information is present in the lexicon. To reiterate, the system is sensitive only to information coded in the abstract representation in the recognition lexicon, and this does not contain any predictable information, whether this be segmental or prosodic.

4.1.3. Processing aspects

So far we have discussed the problem of variation primarily with reference to the theory of lexical representation, looking at the consequences of abstractness and underspecification for the treatment of neutralising and allophonic variation in the surface string. Implicit in this discussion, however, has been the fundamental congruency between our basic assumptions about processing mechanisms and our basic assumptions about representation.

The theory of abstract representation assumes that all items have a unique representation. Surface similarities resulting from phonological assimilatory neutralisations are not present in the lexicon, so that there are no underlying ambiguities to confuse identification. This matches the basic assumption of the cohort model, that perceptual processing is contingent and competitive. This allows items to still be correctly identified, despite variations in their surface realisation, so long as this variation maintains the underlying distinctiveness of the item relative to its close competitors.

The crucial advantage of a cohort-like recognition process is that the match between the input pattern and the lexical form representation does not have to be a complete one; it has to be sufficient to discriminate the activation level for the form in question from the activation level of other possible words. The major constraint here is that the correct item should enter into the word-initial cohort. The mapping between input patterns and lexical entries is highly directional (Marslen-Wilson & Zwitserlood, 1989), and if the onset of the word is sufficiently distorted or degraded it will not enter into the cohort, and will not, therefore, be treated as a candidate for recognition. If phonological processes are indeed constrained in the ways we suggested earlier, then even if there is neutralising variation word-initially, this will not create conflicts between surface features and features specified in the recognition lexicon, and will not, therefore, prevent the correct candidate from entering the cohort of active candidates.[13]

A final aspect of the model's lexically based approach to the problem of variation – and this is not something we have discussed here – is its use of contextual constraints. These are the constraints on lexical identity that derive from the utterance and discourse context in which the current item is being processed. These constraints do not operate "top-down", to affect the basic form-based processes of access and selection. But they do operate to select, at a subsequent stage of the process, between possible candidates. Such constraints will not override the perceptual choices that emerge from the primary form-based process, but they will enable the system to discriminate among candidates when the sensory information is truly ambiguous or insufficient (Marslen-Wilson, 1989). This allows phonologically highly reduced forms to be nonetheless identifiable under the appropriate conditions of speaking.

[13]A well-cited example of word-initial neutralization is consonant mutation in the Celtic languages. At a first glance, this appears to be a counterexample to our claim. However, all instances of mutation eliminating distinctive contrast in only word-initial consonants are morphologised. For example, initial consonants of certain nouns in Welsh undergo lenition when preceded by *ei* "his" (Willis, 1986). This process changes /p, b/ to /b, v/: *pen* "head", but *ei ben* [iben] "his head"; *brawd* [brawd] "brother", but *ei frawd* [ivrawd] "his brother". The alternation is not phonologically conditioned.

4.1.4. Overview

In summary, the approach to variation that we are suggesting here has two main aspects to it. First, the assumption of an abstract underspecified representation means that the access process should not be affected by most types of phonologically based variation. Word-forms are not allowed to vary on the surface in ways which would cause them to lose their underlying distinctiveness. Second, the competitive nature of processing at the lexical level, and the possibility for contextual support at subsequent stages of the comprehension process, means that the form-based access and selection process need only deliver a *relative* invariance. The correct word-candidate need only be distinctive relative to its competitors; it need never be distinctive absolutely.

5. Conclusions

If we accept the proposals put forward here, for an abstract underspecified representation of lexical form as the content of the recognition lexicon, operating in the framework of a cohort-based processing model to determine the time-course and outcome of the on-line interpretation of lexical form, then this offers a considerable simplification of many aspects of current research into spoken language.

First, the representation itself is greatly simplified, since it has stripped away from it all predictable and default information. This, in turn, offers simpler and more direct procedures for understanding how to deal with variations.

Secondly, the representation we have shown to be relevant for language perception is one that has usually been assumed – implicitly if not explicitly – to be the kind of representation that is appropriate as the basis for language *production*. Many puzzles and complexities in current thinking about the relationship between perception and production would be greatly simplified if we could assume that they are both run off the same abstract underspecified representation.

Thirdly, the results suggest a close linkage between phonological analyses and the actual mental representations of linguistic knowledge that function in psycholinguistic performance. If phonological claims about the underlying representations of lexical form can be subjected, even if indirectly, to the constraints imposed by the need to explain perceptual data as well as distributional data, and if experimental psycholinguistic research into language processing can exploit – again even indirectly – linguistic claims about the representations that are employed in processing, then this offers the possibility of real progress in understanding the representation and processing of lexical form.

Moving to a quite different issue, the current results bear directly on the current resurgence of strong empiricist approaches to perceptual processing and mental representation – that is, the introduction of connectionist learning models as explanatory devices in the study of language processing (e.g., Rumelhart & McClelland, 1986). Fundamental to this approach is the claim that perceptual representations evolve directly from perceptual experience. Listeners' representations of spoken words, therefore, should directly reflect their experience with what words sound like (e.g., Elman & Zipser, 1988). Since CVNs are heard with nasalised vowels, they should be perceptually represented as having nasalised vowels, and should therefore be perceptually distinct from CVCs in the ways we laid out in the Introduction.

Our results do not support this claim. Listeners do not have available to them, as they process the speech input, a representation of the surface phonetic properties of a given word-form. What determines their performance – not just in CVNs, but in all the word-types we looked at – is not what the word actually sounds like, as a phonetic string, but rather its abstract underspecified representation in the recognition lexicon. These mental representations are structured according to constraints which are not just a projection of perceptual experience.

Finally, we should comment on the specifically linguistic implications of this research, especially as they concern competing views of underspecification (although we should again stress that this research was not intended as a test of any particular linguistic theory). The responses that we obtained to phonetically oral and nasal vowels, in two languages which differ in the phonological status of the feature [nasal], argue for radical underspecification rather than contrastive underspecification. Under the latter view, both plus and minus values of [nasal] would be specified for vowels in Bengali since this feature is used to distinguish segments like [ɑ̃] and [ɑ]. In English, of course, all views agree that the vowels remain unspecified for nasality. Our results show that lack of nasality remains unspecified for Bengali as well, regardless of the fact that nasal and non-nasal vowels contrast in this language. This endorses the radical underspecification assumption that only one value of a feature is present in underlying representation.

When a vowel is not specified for nasality, this means that it is not specified for that feature at all, plus or minus. This raises the further question of the binary nature of the feature nasal. If [−nasal] plays no role, can one assume that this feature is monovalent? If [nasal] is monovalent, segments can be specified as being [nasal], but there can be no specification of [−nasal] (cf. Mester & Ito, 1989; Van der Hulst, 1989, for discussion on monovalent features).

Phonologically, both [−nasal] and [+nasal] have been shown to spread in harmony processes (for example in Guarani; Goldsmith, 1976) and therefore one could argue that this feature is not inherently monovalent. Our results are compatible with either assumption ([nasal] or [+nasal]) about the way nasality is specified in underlying representations.

Recent literature in phonology and phonetics discusses underspecification issues in underlying representations as well as in surface phonetic representations (cf. Archangeli, 1988; Keating, 1988). To our knowledge, however, this paper is the first attempt to empirically study the effects of underspecified representations in comprehension, investigating the listeners' interpretation of surface nasality as a function of the presence and absence of nasality in the phonological representation. Here, a word of caution is called for. We can confidently make strong claims about underspecification of nasality (and its processing consequences). Where other features are concerned, they ought to behave in a similar fashion if the phonological status of the features in the languages examined have equivalent properties. But to establish these properties will require careful phonological analysis.

Our concern here, in summary, has been to break the ground for a genuinely psycholinguistic model of language comprehension, providing a unified account of representation and process in natural language. The proposals spelt out in this paper require abstract representations of lexical form, interacting directly with the on-line interpretation of sensory information, and where the properties of these abstract representations are determined by the linguistic structure of the language in question. A combined psychological and linguistic approach seems inescapable if we are to understand language as a psycholinguistic phenomenon.

References

Archangeli, D. (1984). *Underspecification in Yawelmani phonology*. Doctoral dissertation, Cambridge, MA: MIT.

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, 5, 183–207.

Chatterjee, S.K. (1975). *Origin and development of the Bengali language*. Calcutta: Rupa.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Church, K. (1987). Phonological parsing and lexical retrieval. In U. Frauenfelder & L.K. Tyler (Eds.), *Spoken word recognition*. Cambridge, MA: MIT Press.

Clements, G.N. (1985). The geometry of phonological features. *Phonology Yearbook*, 2, 225–252.

Clements, G.N. (1987). Phonological feature representation and the description of intrusive stops. In *Papers from the Parasession on Autosegmental and Metrical Phonology*. Chicago: Chicago Linguistic Society.

Clements, G.N. (1989). *A unified set of features for consonants and vowels*. Manuscript, Cornell University.

Cutler, A. (1989). Auditory lexical access: Where do we start? In W.D. Marslen-Wilson (Ed.), *Lexical representation and process*. Cambridge, MA: MIT Press.

Dev, A.T. (1973). *Students' favourite dictionary*. Calcutta: Dev Sahitya Kutir.
Elman, J.L., & Zipser, D. (1988). Learning the hidden structure of speech. *Journal of the Acoustical Society of America, 83*, 1615–1626.
Ferguson, C.A., & Chowdhury, M. (1960). The phonemes of Bengali. *Language, 36*, 22–59.
Fowler, C.A. (1984). Segmentation of coarticulated speech in perception. *Perception and Psychophysics, 36*, 359–368.
Frauenfelder, U., & Lahiri, A. (1989). Understanding words and word recognition. In W.D. Marslen-Wilson (Ed.). *Lexical representation and process*. Cambridge, MA: MIT Press.
Frazier, L. (1987). Structure in auditory word recognition. In U. Frauenfelder & L.K. Tyler (Eds.), *Spoken word recognition*. Cambridge, MA: MIT Press.
Goldsmith, J. (1976). *Autosegmental phonology*. Doctoral dissertation, MIT.
Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics, 28*, 267–283.
Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry, 20*, 253–306.
Keating, P. (1988). Underspecification in phonetics. *Phonology, 5*, 275–292.
Kiparsky, P. (1985). Some consequences of lexical phonology. *Phonology Yearbook, 2*, 85–138.
Klaiman, M.H. (1987). Bengali. In B. Comrie (Ed.), *The world's major languages*. London: Croom Helm.
Klatt, D.H. (1989). Review of selected models of speech perception. In W.D. Marslen-Wilson (Ed.), *Lexical representation and process*. Cambridge, MA: MIT Press.
Kucera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
Levin, J. (1985). *A metrical theory of syllabicity*. Doctoral dissertation, Cambridge, MA: MIT.
Liberman, A.M., & Mattingley, I.G. (1985). The motor theory of speech revisited. *Cognition, 21*, 1–36.
Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R.T. Oehrle (Eds.), *Language sound structure*. Cambridge, MA: MIT Press.
Marslen-Wilson, W.D. (1984). Function and process in spoken word-recognition. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes*. Hillsdale, NJ: Erlbaum.
Marslen-Wilson, W.D. (1987). Functional parallelism in spoken word-recognition. In U. Frauenfelder & L.K. Tyler (Eds.), *Spoken word recognition*. Cambridge, MA: MIT Press.
Marslen-Wilson, W.D. (1989). Access and integration: Projecting sound onto meaning. In W.D. Marslen-Wilson (Ed.), *Lexical representation and process*. Cambridge, MA: MIT Press.
Marslen-Wilson, W.D., & Welsh, A. (1978). Processing interactions during word-recognition in continuous speech. *Cognitive Psychology, 10*, 29–63.
Marslen-Wilson, W.D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance,15*, 576–585.
Mascaro, J. (1976). *Catalan phonology and the phonological cycle*. Doctoral dissertation, Cambridge, MIT.
Mester, R.M., & Ito, J. (1989). Feature predictability and feature specification: palatal prosody in Japanese mimetics. *Language, 65*, 289–293.
Mitra, S.C. (1968). *Century dictionary*. Calcutta: New Bengal Press.
Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.
Pisoni, D., & Luce, P.A. (1987). Acoustic-phonetic representations in word-recognition. In U. Frauenfelder & L.K. Tyler (Eds.), *Spoken word recognition*. Cambridge, MA: MIT Press.
Pulleyblank, D. (1983). *Tone in lexical phonology*. Doctoral dissertation, Cambridge, MA: MIT.
Rumelhart, D.E., & McClelland, J.L. (1986). PDP models and general issues in cognitive science. In D.E. Rumelhart & J.L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1. Foundations*. Cambridge, MA: MIT Press.
Segui, J., Mehler, J., Frauenfelder, U., & Morton, J. (1982). The word frequency effect and lexical access. *Neuropsychologia, 20*, 612–627.

Stevens, K.N. (1986). Models of phonetic recognition II: An approach to feature-based recognition. In P. Mermelstein (Ed.), *Proceedings of the Montreal Satellite Symposium on Speech Recognition, Twelfth International Congress on Acoustics.*

Tyler, L.K., & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception and Psychophysics, 36*, 217–222.

Van der Hulst, H. (1989). Atoms of segmental structure: Components, gestures, and dependency. *Phonology, 6*, 253–284.

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

Warren, P., & Marslen-Wilson, W.D. (1988). Cues to lexical choice: Discriminating place and voice. *Perception and Psychophysics, 43*, 21–30.

Willis, P. (1986). *The initial consonant mutations in Welsh and Breton.* Bloomington: Indiana University Linguistics Club.

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