4

Speech perception and lexical representation: the role of vowel nasalization in Hindi and English

JOHN J. OHALA and MANJARI OHALA*

4.1 Introduction

The theory of underspecification of lexical representation maintains that the stored form of words omits any predictable or nondistinctive information (Archangeli, 1988). As a data compression technique, there is no dispute that lexical storage can be minimized via underspecification. However, this leaves open the question of whether underspecification corresponds to how words are represented in speakers' mental lexicons. There have been various attempts to gather empirical evidence on the issue (Davidsen-Neilsen, 1975; Keating, 1988; Cohn, 1990; Choi, 1992; Stemberger, 1991, 1992). A recent attempt of this sort was made by Lahiri & Marslen-Wilson (1991, 1992; henceforth L&MW) and Lahiri (1991) who looked to speech perception as a domain where underspecification might be validated. In this paper we summarize their experiment and report our attempt to replicate it.

L&MW suggested that the underlying lexical representations posited by phonologists "correspond, in some significant way, to the listener's mental representation of lexical forms . . . and that these representations have direct consequences for the way . . . the listener interprets the incoming acoustic-phonetic information" (1992: 229). Lahiri (1991: 385) argued specifically that "the surface structures derived after postlexical spreading do not play a distinctive role in perception; rather, a more abstract underspecified representation determines the interpretation of a phonetic cue."

L&MW investigated Bengali and English listeners' identification of syllables in their respective languages which had had varying amounts of their terminal portions gated out. Of interest was their reaction to the presence or absence of nasalization on the vowels heard without the following consonant. To understand L&MW's predictions, it is convenient to refer to Figure 4.1 (corresponding to Figure 9.1 in L&MW, 1992).

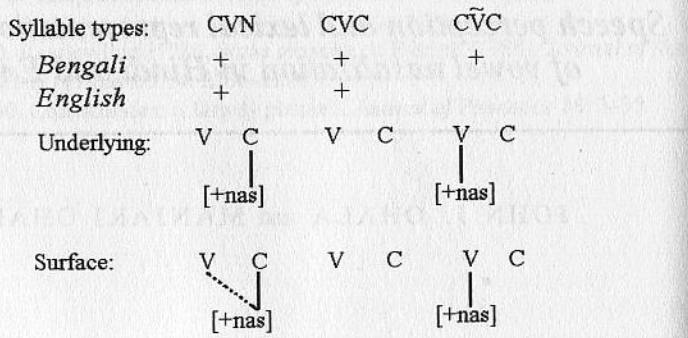

Figure 4.1. Schematic illustration of the representation of VC portion of CVN, CVC, and CṼC syllables in the mental lexicon ("underlying"), and as they appear in speech ("surface"), according to underspecification theory. Pluses ("+") opposite *Bengali* and *English* indicate whether the given syllable type occurs in the language. Vowel nasalization due to postlexical spreading of [+nasal] from a nasal consonant (N) is indicated by the dashed line.

As shown in Figure 4.1, insofar as vowel nasalization is concerned, Bengali has three syllabletypes phonologically: CVC, CVN and CṼC. In the CVC syllable the vowel is regarded as *unspecified* for [nasal] at the underlying (phonological) level and is oral at the surface (phonetic) level. In the CṼC syllable the vowel is [+nasal] at both levels. In the CVN the vowel is said to be unspecified for nasal at the underlying level, but is nasal at the surface due to posited postlexical rules spreading the nasalization from the following consonant onto the vowel (represented by the dashed line in Figure 4.1). What this implies is that although the vowel is uttered with nasalization, this nasalization is non-distinctive. English has only the CVC and CVN syllable types, to the vowels of which the same pattern of vowel nasalization applies as in Bengali.

If the final consonant were spliced off of such syllables and listeners required to guess which word (or type of word) they were excised from, L&MW make different predictions depending on whether the listeners compared the stimuli with the underlying underspecified representation (UR) or with the fully specified surface representation (SR). Table 4.1 summarizes their predictions according to the two competing hypotheses, UR and SR. CV(C), CV(N), CṼ(C) refer to the stimuli having the canonical shapes CVC, CVN, and CṼC with the final consonant spliced off; Beng3, Beng2, Engl2 etc. refer to conditions where the possible responses were any of these three syllable types or just two of them (CVC and CVN). A prediction that a given stimulus would be ambiguous is

indicated by listing all possible response types. For example, according to the UR hypothesis, stimulus CV(C) would be ambiguous between CVC and CVN.

Of the seven possible conditions, the UR and SR hypotheses make identical predictions about one, Engl2 CV(N). They make different predictions about the remaining six, Beng3 CV(N), Beng3 CV(C), Beng3 CṼ(C), Beng2 CV(N), Beng2 CV(C), Engl2 CV(C).

Table 4.1. L&MW's predictions of responses to end-truncated stimuli according to two competing hypotheses, UR and SR.

STIMULUS → LISTENERS' ACCESS ↓	CV(N)	CV(C)	CṼ+(C)
UR	*Beng3*: CṼC	*Beng3*: CVC/CVN	*Beng3*: CṼC
	Beng2: CVC/CVN	*Beng2*: CVC/CVN	
	--------------------	--------------------	
	Engl2: CVN	*Engl2*: CVC/CVN	
SR	*Beng3*.: CVN/CṼC	*Beng3*.: CVC	*Beng3*: CṼC/CVN
	Beng2: CVN	*Beng2*.: CVC	
	--------------------	--------------------	
	Engl2: CVN	*Engl2*: CVC	

4.2 L&MW's experiment

L&MW (1991, 1992) employed a gating paradigm (Cohen & 't Hart, 1964; Grosjean, 1980) in which listeners were presented with a syllable with its final consonant and most of the vowel preceding it gated out, and on subsequent trials with progressively more of the syllable included until finally the entire intact syllable was presented. At each trial listeners were asked to guess what the word was. Subjects were not constrained in their choice of possible guesses except that it was supposed to be a complete word. Up to 11 gates were placed at 40 ms intervals, with some variation to ensure that one gate coincided with the vowel offset, i.e., the onset of the syllable-final consonant. Of interest were the listeners' reactions to those stimuli that did not include any portion of the final consonant, the hypothesis being that when they heard this consonant the character of the syllable would be perceptually evident and their responses would no longer help to differentiate between the competing hypotheses.

L&MW ran three separate conditions as a function of the type of stimuli used: (a) Bengali triplets, in which all three syllable types (CVC, CṼC, CVN) appeared; e.g. /bad/ "difference", /bãd/ "dam", /bãn/ "flood"; (b) Bengali doublets, in which, given the initial CV, the only existing Bengali words would have been CVC or CVN (i.e., where subjects would not have the option of giving CṼC responses); e.g. /lop/, /lom/, */lõC-/; and (c) English doublets involving CVC and CVN syllables; e.g. /dʌb/ *dub*, /dʌm/ *dumb*. Conditions (b) and (c) were designed to permit partial comparison between the Bengali and English results.

L&MW obtained results in the six relevant conditions which they interpret as supporting the UR hypothesis. In the Beng3 CV(C), Beng2 CV(C), and Engl2 CV(C) conditions, they predicted responses split between CVN and CVC. This did occur although the percentages of CVN responses were low (13.4%, 14.7%, and 16.6%, respectively). L&MW characterize these percentages as "relatively high" (1991: 271) and equal to the proportion of the CVN syllable type in the total set of monosyllables of each language. In other words, when hearing the truncated CV(C) stimuli subjects were in effect guessing whether the word was CVC or CVN, and the incidence of guesses matched approximately the incidence of those syllable types in the language.

As predicted by the UR hypothesis, the majority of responses to the Beng3 CV(N) stimuli were CṼC (63%), and the majority of responses to the Beng3 CṼ(C) stimuli were CṼC (56.8%), i.e., the responses in these conditions were not split equally between CṼC and CVN, as would be predicted by the SR hypothesis. In the Beng2 CV(N) condition, the majority of responses were CVC, while some subjects gave CṼC responses (17%), even though the condition was designed to eliminate CṼC as a possible response. On the other hand, the "correct" CVN response was the minority response. L&MW (1992: 244 ff.) remark: "this difficulty in producing a CVN response follows directly from the [UR] hypothesis, where nasalization on the surface is interpreted as a cue to an underlyingly nasal vowel."

In sum, L&MW find multiple points of support for the UR hypothesis in their results.

4.3 A replication

We attempted to replicate L&MW's study (with some modifications) using English and Hindi stimuli and subjects. As far as oral and nasal vowel patterns are concerned, Hindi is virtually identical to Bengali (as described by L&MW).[1]

4.3.1 Method

Some aspects of our experiment differed from that of L&MW. First, our word lists, given in Tables 4.2–4.4, were much shorter than theirs. The Hindi triplet condition contained five near-minimal triplets (L&MW had 21), the Hindi doublet condition four (L&MW had 20), and the English doublet condition eight minimal pairs (L&MW had 20).

Table 4.2. Hindi minimal triplets used in the Hindi triplet condition.

sas	"mother-in-law"	sãs	"breath"	san	"to mix (e.g. flour)"
kʰas	"main"	kʰãs	"cough"	kʰan	"mine (for ore)"
bas	"bad smell"	bãs	"bamboo"	ban	(type of rope)
bʰuk[2]	"hunger"	bʰũk	"bark (of dog)"	bʰun	"roast" (v.)
baʈ	"path"	bãʈ	"distribution"	ban	(type of rope)

Table 4.3. Hindi minimal pairs used in the Hindi doublet condition.

ʧɛʈ	"a month"	ʧɛn	"peace"
t̪is	"thirty"	t̪in	"three"
d̪ud̪ʰ	"milk" (n.)	d̪un	"twice something; name of a train"
d̪ekʰ	"see"	d̪en	"gift"

Second, instead of the open response set used by L&MW, we used a closed response set. The problem with an open response set is that it does not permit any statistical analysis of the results. L&MW suggested that the approximately 15% rate at which CVN responses were given to CV(C) stimuli mirrored the incidence of CVN words in the lexicon. But it is not clear how relevant this figure is; according to the cohort model (Marslen-Wilson, 1984),when listeners are looking for candidate words to match the incoming truncated signal to, they are not searching the *entire* lexicon, but only that subpart of the lexiconthat has the same initial CV sequence. Since in L&MW's experiment subjects were apparently allowed to give polysyllabic words as responses, the possible matches would not have been confined to monosyllables, so it is very difficult to assess what the proportion of CVN syllables was out of the total cohort.

Table 4.4. English minimal pairs used in the English doublet condition.

rube	room
lewd	loon
seed	seen
raze	rain
lead [lid]	lean
ride	Rhine
laud	lawn
seize	seen

The open response set presents other difficulties as well. As noted by L&MW (1991, footnote 9) some of their subjects' responses on the early, more severely gated stimuli could not be scored according to their criteria, and thus approximately 3–5% of the responses had to be discarded. Furthermore, in the Beng2 CV(N) condition, a full 17% of the responses were CṼC, although CṼC was not supposed to be a possible response. Finally, an open response set does not guarantee that in the triplet condition, all three syllable types were always available in each subject's mental lexicon.

For these reasons, we restricted our subjects' answers to each stimulus just to the target triplet or pair. This enabled us to analyze the results statistically. We do not think this modification of the technique used by L&MW materially changed the comparability of the experiments. L&MW endorse Marslen-Wilson's "cohort" model, which characterizes the act of word recognition as an ongoing reduction in the number of possible matches an incoming speech signal has to items in the lexicon. At some point in the identification of words the cohort set must also be reduced to two or three possibilities even in the L&MW experiment. In our experiment *we* provided listeners with that reduced cohort set. In addition we believe that by providing subjects with a restricted response set we avoided the thorny problem of trying to find words which would be of equal frequency.

Third, we modified the way the stimuli were gated. L&MW gated their stimuli into silence. But a relatively abrupt attenuation of the signal could in itself be a potent segmental cue for a voiceless stop, and might lead listeners to interpret, for example, a gated [sã(n)] as /sãp/. To avoid this effect, we gated into wide-band white noise which was 1 dB less in intensity that the peak intensity of the word itself. Furthermore, to avoid transients the amplitude of the speech was attenuated from full scale to zero over 5 ms and the noise increased from zero to full scale in 5 ms. These two ramps overlapped such that the start of the attenuating speech ramp coincided with the start of the increasing noise

ramp. In this way listeners were not presented with a spurious segmental cue and could imagine that the latter part of the signal was simply obscured (Pols & Schouten, 1978).

Fourth, we also had only five gate points (as opposed to L&MW's 11): one coincided with the vowel offset and three others were spaced at 60 ms intervals before that. Vowel offset was determined as that point where there was an abrupt discontinuity in the amplitude and/or spectrum of the vowel. The fifth gate included the whole word.

Fifth, we gathered subjects' responses in a way to avoid a possible "hysteresis effect." As discussed in J. Ohala (1992a), by presenting first the most severely gated version of a word and then in immediate succession the progressively longer versions until the whole word was heard, L&MW's subjects' responses may have been contaminated by a kind of "hysteresis effect," i.e., their judgments on any given stimulus may have been partly influenced by their judgments on an earlier version. In other words, once having made a judgment, subjects may be reluctant to change it (see Fredricksen, 1967). Insofar as vowel nasalization is concerned, the early portions of a vowel in a CVN syllable would be less nasalized than later portions, and so this effect could suppress CVN or CṼC judgments. Evidence that the relatively low number of CVN judgments L&MW obtained with the CV(N) stimuli in the case of the English doublets may be due to hysteresis comes from Ali, Gallagher, Goldstein & Daniloff's (1971) results, which showed that English listeners *could* differentiate between CV(C) and CVN syllables that had had the last one third of the vowel and the final consonant removed. The Engl2 CV(N) condition is not one that L&MW consider capable of differentiating between the UR and SR hypothesis, but a hysteresis effect could have had an influence on the other conditions as well. L&MW (1992: 237) state that previous studies show that such an effect would be negligible. But these cited studies do not give separate results for words of the type that we focus on here, i.e., words differing only in presence or absence of nasalization on the vowel.

Although this methodological issue is not completely resolved, we divided our subjects and our stimuli into five groups such that no group heard more than one gated version of a word type, with the following exception: if a given group heard the most severely gated version of a word, then four tokens later they were presented with the whole (ungated) word. We counted on there being minimal influence from their judgment on the earlier gate on their judgment of the whole ungated word. In any case, the only responses of interest, and those treated statistically, were of those stimuli with gates up to and including the vowel offset.

The speech samples were recorded under high quality conditions, the second author providing the Hindi words and the first author, the English words. They were digitized at 16 kHz after being low-pass filtered at 7 kHz, and gated with added noise using both commercial speech analysis software (CSRE) and custom-made software.

The stimuli were presented to the subjects over headphones via a high quality portable tape recorder. The Hindi subjects (N = 39) were students at the Jawaharlal Nehru University, Delhi, and listened to the stimuli in a quiet room. The English subjects (N = 44) were students at the University of Alberta and listened to the stimuli in a sound-attenuated booth. All subjects were paid for their participation. Although we differentiate between the Hindi triplet and doublet conditions (following L&MW), stimuli from both sets were intermixed and differentiated for the subjects only by virtue of having three or two possible answers, respectively, on the answer sheet. Each group's answer sheets had a randomized order of the possible responses to a given stimulus. Subjects were presented with the answer sheet (in Devanagari, the Hindi script, and English, for Hindi and English subjects respectively), and were told that they would hear some words some of which had been interrupted by noise. They were told to attempt to identify the words from among the choices given on the answer sheet, guessing if necessary.

4.3.2 Results and discussion

Overall results are presented in Tables 4.5–4.7; results for each gate in all conditions (with L&MW's results superimposed) are presented in Figures 4.2 to 4.8.

The one part of the experiment in which the UR and SR hypotheses would, according to L&MW, give identical predictions is the Engl2 CV(N) condition, to which the responses should all be CVN. Curiously, in L&MW's results, CVN was the minority response, but they do not view this as undermining their predictions. In our data the CVN judgments to Engl2 CV(N) stimuli were overwhelmingly correct (see Table 4.7), and as is evident from Figure 4.8, this judgment was largely correct from the earliest truncation point. Our results are congruent with the earlier study by Ali *et al.* (1971). We surmise that this higher rate of correct responses may result from our efforts at reducing a possible hysteresis effect, as discussed above.

The results for the Hind3 CV(N) condition are similar to those of L&MW, in that the majority response was CṼC ($\chi^2 = 35.8$ (2 df), $p < 0.001$). However, among the minority responses, we had a higher rate for CVN than did L&MW (24.4% as opposed to their 7.9%), and this was comparable to the CVC responses (19.9%). In the Hind2 CV(N) condition we found no significant difference between the CVN and CVC responses ($\chi^2 = 0.4$ (1 df), n.s.) although the trend is in the same direction found by L&MW. Neither result provides support for the SR hypothesis, and are rather in accord with the predictions of the UR as formulated by L&MW.

Table 4.5. Results for the Hindi triplet condition: raw numbers (in bold), and percentages (in italics) of response type to stimulus type, up to vowel offset. χ^2 probability levels for each set of responses are shown in parentheses after the stimulus label. For comparison, L&MW's results (presented as percentages) for their Bengali triplet condition are given in square brackets.

		RESPONSE		
		CVC	CṼC	CVN
	CV(C) ($p < 0.001$)	**140** (*71.8*) [80.3]	**38** (*19.5*) [0.7]	**17** (*8.7*) [13.4]
STIMULUS	CṼ(C) ($p < 0.001$)	**28** (*14.4*) [33.2]	**139** (*71.3*) [56.8]	**28** (*14.4*) [5.2]
	CV(N) ($p < 0.001$)	**31** (*19.9*) [23.5]	**87** (*55.8*) [63.0]	**38** (*24.4*) [7.9]

In the three cases where CV(C) stimuli were involved (see Tables 4.5, 4.6 and 4.7) the majority responses were the correct CVC, a result which was highly significant in all cases ($\chi^2 = 133$ (2 df), $\chi^2 = 35.1$ (1 df), $\chi^2 = 109$ (1 df), for Hindi triplets, Hindi doublets and English doublets respectively; $p < 0.001$ in all cases). Clearly, the subjects did *not* find the stimuli ambiguous between CVC and CVN. L&MW state that according to the SR hypothesis there should be *no* CVN responses to CV(C) stimuli. But this is an unreasonable claim; although it can be granted that the unelaborated SR theory would not predict CVN responses, in real life when such perceptual experiments are done one always encounters "noise" in the data due either to unanticipated defects in the stimuli, subjects' inattention, or defects in the theory (because it does not cover all events in the domain to which it pertains). Rather than zero CVN responses, all that can be expected is that there be a statistically significant "tilt" in the incidence of the responses. This is what we have shown.

In the discussion of their results for the Beng3 CV(C) condition, L&MW draw attention to the higher rate of (incorrect) CVN over CṼC responses. In our results, this was reversed: there were 19.5% CṼC responses to 8.7% CVN responses. We have no explanation for this, and do not attach any importance to it since both responses were overwhelmingly in the minority.

Table 4.6. Results for the Hindi doublet condition: raw numbers (in bold), and percentages (in italics) of response type to stimulus type, up to vowel offset. χ^2 probability levels for each set of responses are shown in parentheses after the stimulus label. For comparison, L&MW's results (presented as percentages) for their Bengali doublet condition are given in square brackets. (Note that CṼC responses have no entries for our results, since we obtained forced-choice responses.)

		RESPONSE		
		CVC	CṼC	CVN
STIMULUS	CV(C) ($p < 0.001$)	**115** (*73.7*) [82.6]	[0.0]	**41** (*26.3*) [14.7]
	CV(N) (n.s.)	**82** (*52.6*) [64.7]	[17.0]	**74** (*47.4*) [15.6]

Table 4.7. Results for the English doublet condition: raw numbers (in bold), and percentages (in italics) of response type to stimulus type, up to vowel offset. χ^2 probability levels for each set of responses are shown in parentheses after the stimulus label. For comparison, L&MW's results (presented as percentages) for their English doublet condition are given in square brackets.

		RESPONSE	
		CVC	CVN
STIMULUS	CV(C) ($p < 0.001$)	**278** (*79.0*) [83.4]	**74** (*21.0*) [16.6]
	CV(N) ($p < 0.001$)	**55** (*17.9*) [59.3]	**253** (*82.1*) [40.7]

In the Hind3 CṼ(C) condition, our results were again quite similar to those of L&MW, in that the majority response was CṼC ($\chi^2 = 126$ (2 df), $p < 0.001$). The only difference is that the minority responses were evenly split between CVC and CVN, as opposed to the greater preponderance of CVC responses in L&MW's experiment. We attach no importance to the even split between the minority responses.

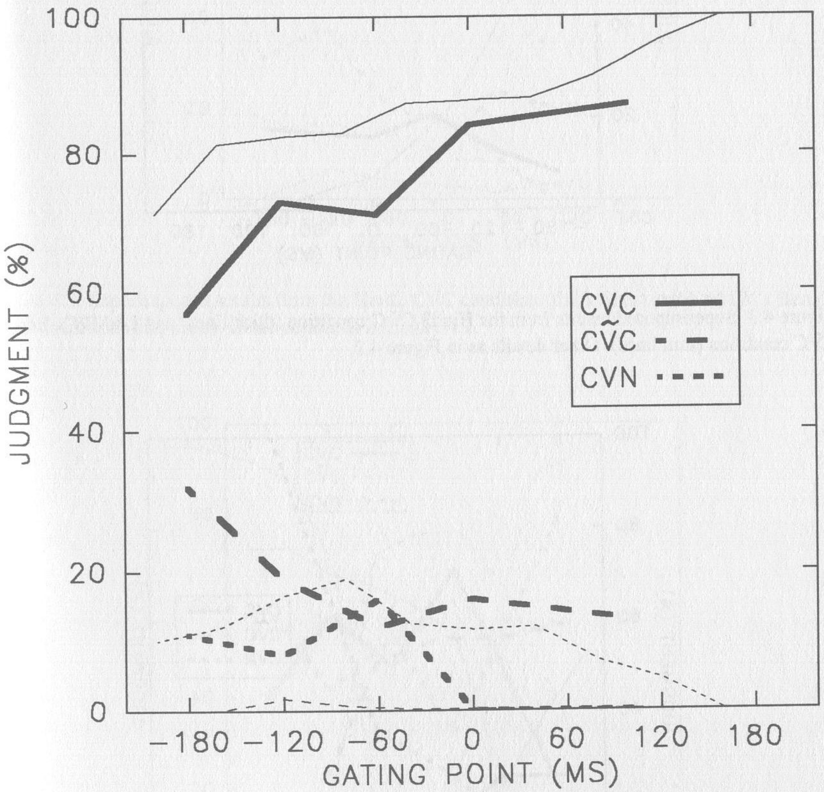

Figure 4.2. Superimposed results from the Hind3 CVC condition (thick lines) and L&MW's Beng3 CVC condition (thin lines). Horizontal axis: gating points (in ms), with respect to vowel offset (= 0 ms); vertical axis: percent judgments in indicated category. Parameters: solid line = CVC responses, medium broken line = CṼC responses, short broken line = CVN responses.

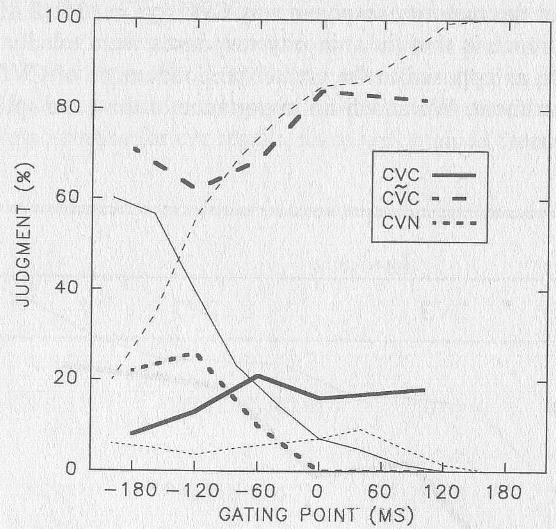

Figure 4.3. Superimposed results from the Hind3 CṼC condition (thick lines) and L&MW's Beng3 CṼC condition (thin lines). Other details as in Figure 4.2.

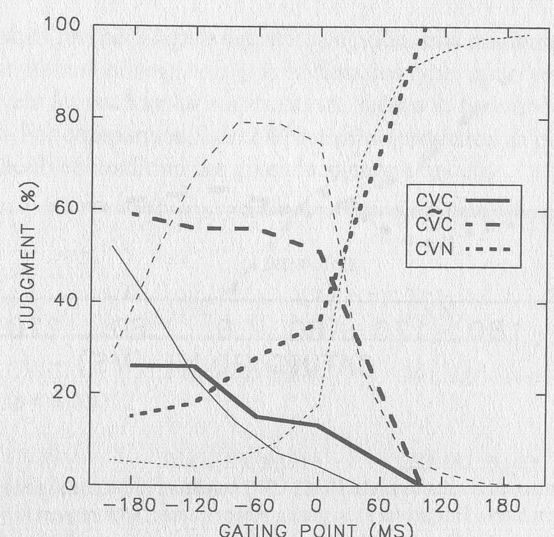

Figure 4.4. Superimposed results from the Hind3 CVN condition (thick lines) and L&MW's Beng3 CVN condition (thin lines). Other details as in Figure 4.2.

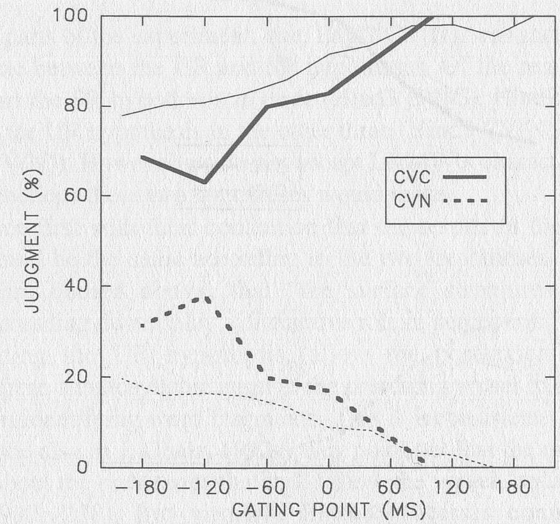

Figure 4.5. Superimposed results from the Hind2 CVC condition (thick lines) and L&MW's Beng2 CVC condition (thin lines). Other details as in Figure 4.2.

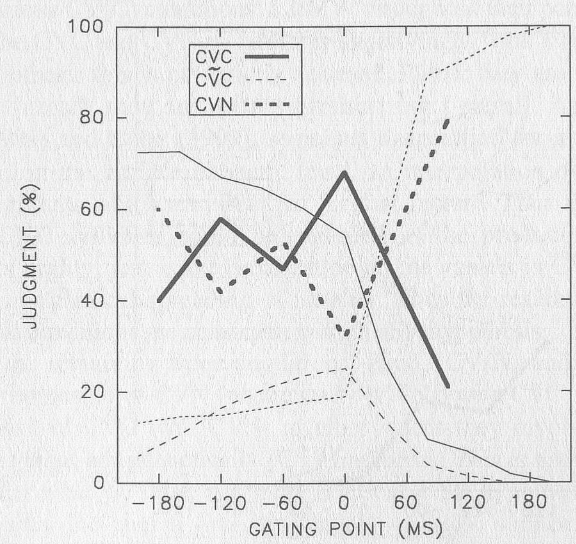

Figure 4.6. Superimposed results from the Hind2 CVN condition (thick lines) and L&MW's Beng2 CVN condition (thin lines). Other details as in Figure 4.2.

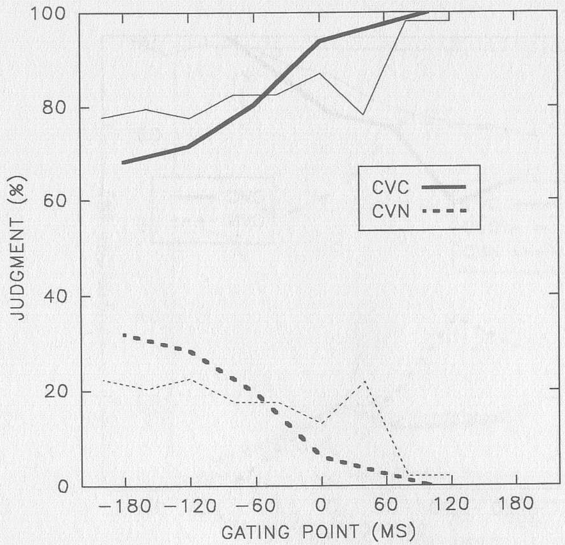

Figure 4.7. Superimposed results from the Engl2 CVC condition (thick lines) and L&MW's Engl2 CVC condition (thin lines). Other details as in Figure 4.2.

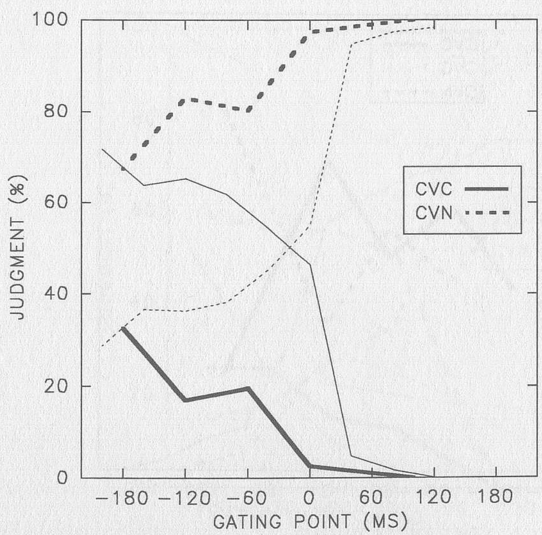

Figure 4.8. Superimposed results from the Engl2 CVN condition (thick lines) and L&MW's Engl2 CVN condition (thin lines). Other details as in Figure 4.2.

4.4 General discussion and conclusion

Of the seven parts of the experiment, one, Engl2 CV(N), was stated to be unable to differentiate between the UR and SR hypotheses. Of the remaining six, our results support the SR hypothesis in three (Hind3 CV(C), Hind2 CV(C), Engl2 CV(C)), and the UR hypothesis in the other three (Hind3 CV(N), Hind3 Cṽ(C), and Hind2 CV(N)). However we do not accept L&MW's characterization of the different predictions these two hypotheses would make.

We disagree first with their contention that the results of the Engl2 CV(N) condition would be the same according to the two hypotheses. In spite of the explicit claim, quoted above, that "the surface structures derived after postlexical spreading do not play a distinctive role in perception" L&MW, while still advocating the UR hypothesis, allow the postlexical spreading of nasalization from a tautosyllabic nasal to the preceding vowel to be accessible to the listener in identifying word fragments. This is inconsistent. To counter this criticism (made also in J. Ohala, 1992a) they point out that the nasalization tells the listener about the *consonant* that will follow the vowel, not the vowel itself (L&MW, 1992: 238). But virtually all the processes commonly labeled "postlexical spreading" have this characteristic. By allowing the listener to access a form exhibiting postlexical spreading they have endorsed the SR hypothesis. Therefore, we consider that our results with the Engl2 CV(N) condition also support the SR not the UR hypothesis. This impacts on the results from the various CV(C) conditions: L&MW throughout their papers refer to the vowels in the CVC and CVN syllables as underlyingly "oral", but according to the UR hypothesis this is not strictly accurate. Rather they are *unspecified* for the feature [nasal]; they are neither [+nasal] nor [–nasal]. As suggested by Keating (1989) and Cohn (1990), segments unspecified for a certain feature value have, on the surface phonetic level, an interpolation of the values of adjacent segments which are specified for that feature. Thus the surface oral vowel in CVC syllables could be regarded as the product of postlexical spreading of orality, just as the nasalization on the vowels in CVN syllables is regarded as postlexical spreading of nasality. Thus the results of four of our experimental conditions are consistent with the SR hypothesis.

Two of the remaining three conditions, Hind3 CV(N) and Hind3 Cṽ(C), involve the confusion of CVN (phonetically [CṽN]) with CṽC, and the lack of any confusion of CṽC with CVN; in other words they involve giving CṽC responses to what are phonetically [Cṽ] fragments. This is not too difficult to explain under what we think is a more reasonable SR hypothesis, namely one which includes the cues a listener requires to make phonemic and lexical decisions. There are two parts to our argument.

First, distinctive speech sounds, as they are manifested phonetically, are not time-invariant. They have a dynamic structure which changes over time. This has always been recognized for diphthongs, affricates, pre- and postnasalized stops, etc. Equally, an important cue for a stop is a release found only at the end of the period of reduced amplitude. M. Ohala (1979, 1983) assigned the feature [distinctive release] to Hindi voiceless aspirated and voiced aspirated stops, as well as affricates to account for certain sound patterns they have in common, in contrast to simple unaspirated stops (see also Steriade, 1993). Similarly, it is well known that the *perceptual cues* for some aspects of segments lie in the "transitions" at their onset and offset. This can have surprising manifestations: Maturi (1991) showed that by cross splicing the final vowels of the Italian words *strada* "street" and *strana* "strange", the results are perceived by listeners as most like *strana* and *strada*, respectively: only the vowels were interchanged but this caused most listeners to hear different consonants! Evidently the presence or absence of what might be considered "non-distinctive" nasalization on the preceding vowel *is* part of what makes a consonant a nasal or an oral segment. We believe that part of what makes a vowel non-distinctively nasal (in Bengali and Hindi) is the presence of a following nasal consonant. This might be regarded as a truism but has important perceptual consequences that bear on the interpretation of the experimental results reported above. And this is the second point in our argument. We illustrate it by an analogy.

Consider the task of identifying capital letters of the Roman alphabet when parts of those letters have been erased. In Figure 4.9, erasure of one side of a letter corresponds to the temporal truncation of part of a speech sound. On the left is a letter whose right side has been erased. Is it "underlyingly" a *C*, *O*, *Q*, or a *G*? Likewise, for the fragment of the letter in the third column: is it "underlyingly" an *F*, *P*, or a *R*? From a strictly logical point of view, we cannot say, of course, but from a perceptual and practical point of view the viewer's inclination is to associate it with that underlying letter to which it bears the greatest *surface* resemblance. In these cases, that means *C* and *F*, respectively. Except under special circumstances, the default strategy in perception is the inverse of WYSIWYG: WYGIWYS ("What You Get Is What You See"). Apparently, in perception, the default strategy is to pay most attention to what is present rather than what is absent but might have been present. (See also McQueen's commentary to this paper; however, see as well Ohala & Shriberg, 1990 and Shriberg, 1992 for special circumstances where the default strategy does not apply and the perceiver "fills in" missing parts of the stimulus array.)

The relevance of these two points to both L&MW's and our studies is this: if the release of a nasal vowel into a nasal consonant is crucial to the differentiation of CṽC and CVN syllables, just as the release of an affricate is essential to differentiate it from a simple stop, then when listeners hear [Cṽ] stimuli, whether these come from CṽC or CVN words, it is not surprising that they associate them with CṽC words. This is equivalent to identifying a *G* with its right half missing as a *C*. (It should be noted that perception of ṽ, i.e., of a

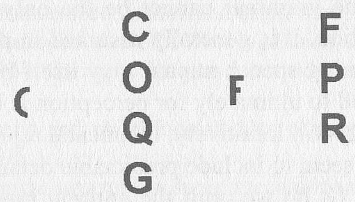

Figure 4.9. Partially erased and fully-specified letter shapes. See text for explanation.

distinctively nasalized vowel, in Bengali and Hindi would presumably not rely as much on the following segment since ṽ can occur in open syllables.)

To sum up our argument, we do not believe that our results or those of L&MW's regarding listeners' judgments on [Cṽ] stimuli necessarily support the UR hypothesis; at best they are compatible with both the UR and SR hypotheses. Obviously more research on this issue would be desirable.

There remains the Hind2 CV(N) condition, where we found no significant difference between the CVC and CVN judgments. This does not support any version of an SR hypothesis that we are aware of. Our expectation was that CVN responses should have predominated; they did not. These stimuli were ambiguous to listeners as L&MW predicted according the UR hypothesis.

By our count, then, of the seven experimental conditions, four support the SR hypothesis, one, the UR hypothesis, and two are a draw.

4.5 Summary

The phonological theory of underspecification maintains that the lexical representation of words in the mental lexicon does not include predictable phonetic details. Lahiri & Marslen-Wilson (1991, 1992) and Lahiri (1991) offer experimental results showing that when accessing words during speech perception listeners attempt to match the incoming signal with such an underspecified representation. We present a replication of their experiment and a re-interpretation of their results which do not fully support their claim.

McQueen, in his commentary to this paper, notes that there are really two hypotheses proposed by L&MW: (a) that the underlying representation is underspecified and (b) that the incoming speech signal is compared with this underspecified form. He notes that our results disproved only hypothesis (b). Hypothesis (a) remains untested and could still be true if a fully specified form was derived from a lexically underspecified form and the incoming signal compared with that. Logically he is correct, but we regard this as an extravagant hypothesis which begs to be pared by Occam's razor. The principal arguments in favor of underspecification are that it is simpler; but "simpler" at what cost?

Counting features in the grammar cannot be the only measure of "cost" (J. Ohala, 1992b). In addition, it is generally assumed in psycholinguistic studies that although the incoming speech stream may itself be transformed in some way, what it is compared to ultimately for perception is the underlying (lexical) representation (Cutler, 1986). Intuitively, the mental representation of things we perceive and recognize seem to include predictable details: in our conception of the shape of the letter *G*, do we omit the entirely predictable left side? Our intuitions tell us "no"; the reader can follow his or her own intuitions, but whatever they are, these intuitions should apply to speech perception as well as letter shapes.

Notes

* We thank Terry Nearey for the use of his laboratory facilities at the University of Alberta, Edmonton, and Anvita Abbi for facilitating the recruiting of subjects at Jawarharlal Nehru University, Delhi. We are also grateful to Aditi Lahiri for sending us the Bengali word list used in hers and William Marslen-Wilson's experiment. In addition, we thank John Ingram, Elisabeth Shriberg, James McQueen, Sarah Hawkins, and Amalia Arvaniti for helpful comments on earlier drafts of the paper. Finally, our thanks to all our subjects both in Delhi and Edmonton.

1 A potential difference that may exist between Bengali and Hindi involves constraints on the types of stops that can appear after long nasalized vowels. As documented by M. Ohala (1983), given the Hindi sequence Ṽ: + STOP, the stop can only be voiceless. Where historically there has been a voiced stop, today there is a sequence of (homorganic) nasal followed by the stop, e.g. Old Hindi [ʧã:da] > Mod. Hindi [ʧãnd]. This point has gone unnoticed by those working on the phonology of Hindi because the orthography fails to reflect it. We do not know whether Bengali has the same constraint or not. The list of Bengali words used by L&MW contains some 20–25% words with final voiced stops among those with distinctively nasal vowels (Aditi Lahiri, personal communication).

2 This word is spelt with a final /k^h/ but in pronunciation the aspiration is neutralized. The word is pronounced as given here.

References

Ali, H. L., T. Gallagher, J. Goldstein & R. G. Daniloff. 1971. Perception of coarticulated nasality. *Journal of the Acoustical Society of America* 49: 538–540.

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

Choi, J. D. 1992. Phonetic underspecification and target interpolation: An acoustic study of Marshallese vowel allophony. *UCLA Working Papers in Phonetics* 82.

Cohen, A. & J. 't Hart. 1964. Gating techniques as an aid in speech analysis. *Language and Speech* 7: 22–39.

Cohn, A. C. 1990. Phonetic and phonological rules of nasalization. *UCLA Working Papers in Phonetics* 76.

Cutler, A. 1986. Phonological structure in speech recognition. *Phonology Yearbook* 3: 161–178.

Davidsen-Nielsen, N. 1975. A phonological analysis of English *sp, st, sk* with special reference to speech error evidence. *Journal of the International Phonetic Association* 5: 3–25.

Fredricksen, J. R. 1967. Cognitive factors in the recognition of ambiguous auditory and visual stimuli. (Monograph.) *Journal of Personality and Social Psychology* 7.

Grosjean, F. 1980. Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics* 28: 267–283.

Keating, P. A. 1988. Underspecification in phonetics. *Phonology* 5: 275–292.

Keating, P. A. 1989. The window model of coarticulation: articulatory evidence. In J. Kingston & M. E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge *University Press*, 281–302.

Lahiri, A. 1991. Anteriority in sibilants. *Proceedings of the XIIth International Congress of Phonetic Sciences,* Aix-en-Provence, 1: 384–388.

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

Lahiri, A. & W. Marslen-Wilson. 1992. Lexical processing and phonological representation. In G. J. Docherty & D. R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody.* Cambridge: Cambridge University Press, 229–254.

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, 125–150.

Maturi, P. 1991. The perception of consonantal nasality in Italian: conditioning factors. *Proceedings of the XIIth International Congress of Phonetic Sciences,* Aix-en-Provence, 5: 50–53.

Ohala, J. J. 1992a. Comments on chapter 9 [Discussion of paper by Lahiri and Marslen-Wilson]. In G. J. Docherty & D. R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody.* Cambridge: Cambridge University Press, 255–257.

Ohala, J. J. 1992b. The costs and benefits of phonological analysis. In P. Downing, S. D. Lima, & M. Noonan (eds.), *The Linguistics of Literacy.* Amsterdam, Philadelphia: J. Benjamins 211–237.

Ohala, J. J. & E. E. Shriberg 1990. Hyper-correction in speech perception. *Proceedings of the 1990 International Conference on Spoken Language Processing*, Kobe, 1: 405–408.

Ohala, M. 1979. Phonological features of Hindi stops. *South Asian Languages Analysis* 1: 79–88.

Ohala, M. 1983. *Aspects of Hindi Phonology*. Delhi: Motilal Banarsidass.

Pols, L. C. W. & M. E. H. Schouten. 1978. Identification of deleted consonants. *Journal of the Acoustical Society of America* 64: 1333–1337.

Shriberg, E. E. 1992. Perceptual restoration of filtered vowels with added noise. *Language & Speech* 35: 127–136.

Stemberger, J. P. 1991. Apparent anti-frequency effects in language production – the addition bias and phonological underspecification. *Journal of Memory and Language* 30: 161–185.

Stemberger J. P. 1992. Vocalic underspecification in English language production. *Language* 68: 492–524.

Steriade, D. 1993. Closure, release, and nasal contours. In M. K. Huffman & R. A. Krakow (eds.), *Nasals, Nasalization, and the Velum*. Orlando, FL: Academic Press, 401–470.