THE PHONOLOGICAL REPRESENTATION OF [VOICE] IN SPEECH PERCEPTION

ALLARD JONGMAN
Cornell University
JoAN A. SERENO
Cornell University
MARianne RAASJEMAKERS
Catholic University Nijmegen
and
ADiti LAHIRI
University of Konstanz

This paper examines to what extent phonological representations affect word identification. The contrast under investigation involves the interaction between voicing and vowel length in Dutch. Dutch has underlying contrasts both in obstruent voicing and in vowel length. The voicing contrast is neutralized on the surface in syllable-final position. Also, both long and short vowels are lengthened by some 25 mace when followed by medial voiced obstruents. The present study investigated whether this vowel length cue influenced listeners when hearing stimuli with ambiguous vowel duration in an identical, neutralized consonantal context in which the underlying representation of the consonant following the vowel differed in voicing. A vowel length continuum ([pt] to [nt]) was created and appended to initial consonants to make two pairs of real words. Each pair differed in vowel length with opposite underlying final consonant representations: /æt/ /æsːt/, and /sʌt/ /stʌt/. Our question was whether the vowel category boundaries would be different in pairs like /æt/ /æsːt/ as compared to /stʌt/ /stʌt/. Although the underlying consonant is either voiced or voiceless, the surface word-final consonant for both pairs of stimuli is always voiceless ([t]). If the listener uses a representation with a voiceless consonant to recognize the words, there should be no difference in categorization of the vowel-length continua. The results of a vowel categorization task, however, showed a significant difference in the location of the phoneme boundaries between the two continua, suggesting that listeners' perception seems to be guided by the underlying phonological representation of words.

Key words: phonological representation, voicing neutralization, vowel duration, speech perception, Dutch

* This study was conducted while all four authors were affiliated with the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands. We gratefully acknowledge Henning Reetz for his technical assistance. We also thank Jeri Jaeger and Bruno Repp for their helpful comments. This study formed part of the third author's Master's thesis. Parts of this research were presented at the 21st Annual Linguistics Symposium at the University of Wisconsin-Milwaukee (April 1992) and at the 123rd Meeting of the Acoustical Society of America in Salt Lake City (May 1992).

Address all correspondence to Allard Jongman, Department of Modern Languages and Linguistics, Morrill Hall, Cornell University, Ithaca, NY 14853.
syllable-final devoicing (Trommelen and Zonneveld, 1979; Booij, 1981). For example, the two Dutch words ‘pad’ (“toad”) and ‘kant’ (“cat”) differ in terms of the underlying representation of the final consonant, with ‘pad’ (/pad/) ending in a voiced stop and ‘kant’ (/kat/) ending in a voiceless stop. The phonological representation is also reflected in the orthography. On the surface, however, both forms are identical, with both ‘pad’ (/pad/) and ‘kant’ (/kat/) ending in a voiceless [t]. The morphological process of plural formation, as exemplified by related pairs such as ‘pad’—‘padden’ (“toads”, [padan]) and ‘kant’—‘katten’ (“cats”, [katən]), suggests that the underlying representation of the dental consonant is /d/ in the first pair and /t/ in the second pair. The present experiment addresses the issue of the psychological reality of such representations by investigating the interaction of voicing and vowel length in Dutch using categorical perception methodology.

Many minimal word pairs can be found in Dutch which differ only in the underlying voicing of the final consonant. There is a lively debate as to whether languages which have a phonological rule of syllable- or word-final devoicing do indeed exhibit complete phonetic neutralization. Incomplete phonetic neutralization has been reported for a number of languages, including German (Charles-Luce, 1985; Port and O’Dell, 1985; Port and Crawford, 1989), Catalan (Dinsen, 1984; Dinsen and Charles-Luce, 1984), and Polish (Słowiak and Dinsen, 1985). In these studies, the duration of some or all of the relevant parameters (the vowel preceding the final consonant, voicing during closure, closure, release) were found to differ significantly as a function of the underlying voicing of the final consonant. Specifically, underlyingly voiced stops were typically characterized by longer preceding vowel duration, proportionally more voicing during closure, shorter closure duration, and shorter release.

However, Fauré and Ivensen (1984) and Jassem and Richter (1989) both reported evidence suggesting complete neutralization in German and Polish, respectively. Both these studies ascribe earlier failures to find complete neutralization to methodological problems, such as the use of written words to elicit test items, recording under unnatural conditions, and significant exposure to a second language which does not have a rule of final devoicing. Clearly, more research is needed to settle the issue of complete vs. incomplete neutralization (for a recent overview, see Blumen, 1991).

In Dutch, our measurements on a number of minimal pairs differing only in the underlying voicing of the final stop consonant show no acoustic differences between these two forms. Two male native speakers of Dutch read 10 minimal word pairs in isolation, five pairs containing long vowels and five pairs containing short vowels (a subset of a larger corpus that we are currently analyzing). There was little variability between subjects and among word pairs and we summed across these conditions. We found no difference between words ending in voiced vs. voiceless stop consonants, either in preceding vowel duration (198 msec vs. 197 msec, respectively, for long vowels; 98 msec vs. 102 msec, respectively, for short vowels) or in consonant closure duration (101 msec vs. 102 msec, respectively, for long vowels; 101 msec vs. 99 msec, respectively, for short vowels). These preliminary results for a small data set suggest that Dutch has complete phonetic neutralization in syllable-final position.

Additionally, Dutch has a contrast in vowel length, as illustrated by the pair ‘tak’ (“branch”, [tak])—‘takk’ (“task”, [tak]). In general, long vowels are approximately 100 msec longer than their short counterparts, as shown above (see also Nooteboom, 1972; Jongman and Sereno, 1991). In a CVC context at a normal speaking rate, the short vowel has an average duration of about 100 msec and the long vowel is about 200 msec in duration (Lahiri, Schriefers, and Kuijpers, 1987). The distinction between the vowels [a] and [a:], for example, is primarily cued by a difference in duration (Nooteboom, 1972), since these vowels are known to have relatively small spectral differences (Pols, Tromp, and Plomp, 1973).

Turning now from phonemic vowel length differences to more phonetically-conditioned differences, it is well-known that in a great number of languages vowels preceding voiced consonants are longer than vowels preceding voiceless consonants (e.g., Chen, 1970). In English, this difference in terms of duration is exceptionally large. The vowel in the word ‘bag’, for example, is about 50 msec longer than in ‘back’ (e.g., House and Fairbanks, 1953; Peterson and Lehiste, 1960; Shair, 1962; Luce and Charles-Luce, 1985; Davis and Summers, 1989). Moreover, in perception experiments, vowel length has been shown to be an important cue to the word-final voicing distinction (e.g., Denes, 1955; Raphael, 1972).

Dutch also exhibits a vowel length difference conditioned by the voicing of the following obstruent. However, this difference in vowel length does not occur, for example, in singular nouns, since the final consonant is always voiceless as the result of syllable-final devoicing. In the plural, however, Dutch does provide minimal pairs that contrast in the voicing of the medial consonant, as in the pairs ‘ratten’ (“rats”, [ratən])—‘radden’ (“wheels”, [radən]) and ‘raten’ (“honeycombs”, [raːtən])—‘raten’ (“councils”, [raːtən]). These plural nouns are formed by adding the plural suffix [-an] to the stem. In these cases, all vowels (both long and short) are lengthened by approximately 20–30 msec when followed by a medial voiceless consonant (Sils and Cohen, 1969b; Kooiman-van Beinum 1980; Nooteboom and Cohen, 1984; Sils, 1985). Moreover, in a perception experiment in which subjects were required to adjust the duration of vowels preceding either voiced or voiceless medial consonants, vowels preceding voiced consonants were 25 msec longer than those preceding voiceless consonants (Sils and Cohen, 1969a). As in English, it seems likely that Dutch vowel length is a cue for voicing.

Making use of these facts about voicing and vowel duration, the present experiment investigates whether the categorization of a given vowel with ambiguous duration varies as a function of the underlying voicing of the final consonant of stimuli in which no voicing contrast is manifest on the surface. Members of a vowel continuum were indeed
in a categorical perception paradigm. When presented with a speech continuum, subjects easily identify distinct categories of speech sounds. Although subjects’ ability to discriminate members of different categories is very good, discrimination among members of the same category is very poor (for reviews, see Kuhl, 1982; Repp, 1983; Harnad, 1987). Categorical perception of vowels has been extensively studied, not only for isolated vowels, but also for vowel continua embedded in a C–C context (e.g., Fry, Abramson, Elmas, and Liberman, 1962; Nootboomb and Doedeman, 1980; Gottfried, Miller, and Payton, 1990).

The categorical perception paradigm has also often been used to investigate lexical effects (e.g., Ganong, 1980; Fox, 1984; Connine, 1986; Conline and Clifton, 1987; Miller and Dexter, 1988). In an early study, Ganong (1980) showed significant shifts in the categorization functions of acoustically ambiguous continuum members due to the lexical status of the continuum endpoints. Using a [a−i] phonetic categorization task, Ganong obtained more voice responses in the vicinity of the phoneme boundary in a ‘dash–tash’ continuum (where the voiced endpoint is a lexical item) and more voiceless responses in a ‘dash–task’ continuum (where the voiceless endpoint is a lexical item). The categorization function shifted depending on the lexical status of the target stimuli. Given ambiguous stimuli, there was a tendency to choose words rather than nonwords.

More recent evidence, however, has tempered this claim. Burton, Baum, and Blumstein (1989) and McQueen (1991) have shown that the acoustic-phonetic quality of the stimuli significantly affects the lexical effects occurring in phonetic categorization tasks. When stimulus quality improved, lexical effects disappeared. These studies suggest that acoustic information is not always overridden by lexical influences. (See also Shin, Blumstein, and Jongman: for a similar finding concerning effects of speaking rates.)

The question we want to ask, however, is somewhat different. If the available acoustic information is indeed ambiguous, what is the possible effect of the phonological representation of a word on the listener’s perception? In the present study, a vowel categorization task was used to determine whether listeners differentially categorized a vowel with ambiguous duration when the voice of the final consonant contrasted in the underlying but not in the surface representation. If listeners are sensitive to the fact that vowels preceding medial voiceless consonants are longer than when preceding medial voiceless consonants, the question arises whether this vowel length cue can influence their perception when hearing stimuli with ambiguous vowel duration in an identical, neutralized consonantal context.

**METHOD**

**Materials**

Two pairs of words were used in the experiment: ‘zat’ ([zət], [zæt], “drunk”) – ‘zaad’ (zaːd, [zaːt], “seed”); and ‘stadt’ ([stəd], [stæt], “city”) – ‘staat’ ([stəːt], [stæt], “state”). Long [aː] and short [a] were chosen for a number of reasons. As mentioned earlier, the distinction is primarily cued by a difference in duration, and these vowels show no diphthongization (Nootenboom, 1974). Furthermore, perception experiments have shown that [aː] can successfully be shortened to obtain 100% [a] responses (Nootenboom and Doedeman, 1980). Notice that these words contrast in terms of the combination of vowel length and underlying final consonant voicing. For ‘stadt–staat’, the short vowel precedes the underlying voiceless stop and the long vowel precedes the underlying voiced stop. Conversely, for ‘zat–zaad’, the short vowel precedes the underlying voiceless consonant and the long vowel the underlying voiced consonant. On the surface, of course, the final consonant was always voiceless. In addition, these pairs were chosen such that there were no lexical competitors which differed from these words only in terms of the underlying voicing of the final consonant. Thus, Dutch has no words [stəːt], [stɑːːt], [zɑːd], and [rɑːːt].

The stimulus words were recorded by a female native speaker of Dutch on a DAT recorder (Sony TCD1000) in a sound-isolated booth using a Sennheiser MD211M microphone. They were then digitized on a VAX750 computer at a sampling rate of 20 kHz with a 10 kHz low-pass filter setting.

A vowel length continuum was constructed by taking a long vowel [aː] and digitally shortening it in 12 steps. Using both visual and auditory criteria, the long vowel [aː] from the word ‘staat’ was first identified. The steady-state portion of the vowel, which excluded initial and final transition information, was then labeled. The vowel continuum was realized by simultaneously exciting a glottal pulse from each end of the steady-state vowel (see Figure 1). All glottal pulses were marked at positive zero-crossings. In such a manner, the intact long vowel (197 msec) was shortened in 12 steps of approximately 10 msec to a short vowel (83 msec). The final [a] segment (215 msec) which included the silent closure and release burst from the original ‘staat’ stimulus was then appended to all members of the 12-step continuum. All continuum members thus ended in an identical voiceless [a].

Two experimental continua were constructed by appending the [stəːt]–[zɑːd] continuum to different initial consonants, the [k] (144 msec) from ‘zad’ and the [k] (244 msec) from ‘stadt’, to create a ‘zat–zaad’ and a ‘stadt–staat’ continuum, respectively. The two continua were thus constructed such that the vowel and final consonant of the stimuli were identical.

A third continuum was included using the same speaker and methodology described above. For these stimuli, all possible combinations of vowel length and voicing exist. In Dutch, all four combinations are possible: [rɑːt] “rat”, [rɑːd] “bread”, [rɑːt] “honeycomb”, and [rɑːːt] “council”. A ‘rat/rad–rɑːd/rad’ continuum was created by splicing the [r] from ‘rad’ (109 msec) onto the [stæt]–[zɑːd] continuum described above. In the endpoints of this continuum, both long and short vowels were paired with underlyingly voiced and voiceless consonants.

Three experimental tapes (one for each continuum) were prepared. For each tape, the 12 members of the continuum were repeated 10 times each. The resulting 120 stimulus items were randomized and presented at a fixed rate. From the offset of a stimulus, there was a two second silent interval until the next stimulus was presented. All responses were recorded on a Micromax computer.
Phonological Representation

Table 1
Crossover boundaries (in msec) for each continuum for each subject

<table>
<thead>
<tr>
<th>Subject</th>
<th>zat–zaad</th>
<th>rat/d–raat/d</th>
<th>stad–staat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>140</td>
<td>148</td>
<td>146</td>
</tr>
<tr>
<td>2</td>
<td>135</td>
<td>136</td>
<td>138</td>
</tr>
<tr>
<td>3</td>
<td>133</td>
<td>136</td>
<td>139</td>
</tr>
<tr>
<td>4</td>
<td>137</td>
<td>135</td>
<td>141</td>
</tr>
<tr>
<td>5</td>
<td>133</td>
<td>135</td>
<td>132</td>
</tr>
<tr>
<td>6</td>
<td>136</td>
<td>143</td>
<td>141</td>
</tr>
<tr>
<td>7</td>
<td>139</td>
<td>147</td>
<td>159</td>
</tr>
<tr>
<td>8</td>
<td>143</td>
<td>143</td>
<td>145</td>
</tr>
<tr>
<td>9</td>
<td>134</td>
<td>139</td>
<td>140</td>
</tr>
<tr>
<td>10</td>
<td>134</td>
<td>134</td>
<td>137</td>
</tr>
<tr>
<td>11</td>
<td>134</td>
<td>132</td>
<td>139</td>
</tr>
<tr>
<td>12</td>
<td>142</td>
<td>144</td>
<td>141</td>
</tr>
<tr>
<td>13</td>
<td>135</td>
<td>136</td>
<td>138</td>
</tr>
<tr>
<td>14</td>
<td>140</td>
<td>142</td>
<td>142</td>
</tr>
<tr>
<td>15</td>
<td>145</td>
<td>144</td>
<td>147</td>
</tr>
<tr>
<td>( \bar{X} )</td>
<td>137</td>
<td>140</td>
<td>142</td>
</tr>
</tbody>
</table>

responses to the ‘zat–zaad’ continuum to those to the ‘stad–staat’ continuum, would the locations of the phoneme boundaries differ? Recall that subjects always heard a voiceless word-final consonant. However, for two endpoint stimuli (‘zat’, ‘staat’), the underlying representation of the final consonant is voiceless while for the other two endpoint stimuli (‘zaad’, ‘staaat’), the underlying representation of the final consonant is voiced. The underlying representation of the final consonant is based on phonological and morphological factors. Given these facts, two competing patterns of results are possible. One possible result is that there is no difference in phoneme boundary between the ‘zat–zaad’ and the ‘stad–staat’ continua, since both continua are phonetically identical, that is, they contain the same vowel and same final voiceless consonant. On this view, if the listener uses only the surface representation of the stimuli (which is vowel + [t] for both pairs of words), then one would expect no differential effect of the voicing of the final consonant on vowel length categorization since, in all cases, the
significant 5-msec difference in phoneme-boundary location between the two continua. The average phoneme boundary for the 'zat–zaad' continuum was at 137 msec while that for the 'stad–staat' continuum was at 142 msec.

The 'rat/d–raat/d' function is located between the two other functions. Pairwise t-tests revealed that the location of the phoneme boundary for 'zat–zaad' occurred at a significantly shorter vowel duration than for 'rat/d–raat/d' \( t(14) = 2.62, p < 0.02 \). In addition, there was a strong trend for the boundary location for 'stad–staat' to occur at a longer vowel duration than for 'rat/d–raat/d' \( t(14) = 1.97, p < 0.07 \).

**Discussion**

In a vowel identification task, categorization of vowel duration was seen to significantly vary as a function of the underlying voicing of the syllable-final consonant. In the ambiguous vowel duration region, from an identical vowel+voiceless consonant continuum, listeners perceived the same vowel duration as short when followed by an underlyingly voiceless consonant but as long when followed by an underlyingly voiced consonant. That is, the phoneme boundary for the 'stad–staat' continuum occurred at a shorter vowel duration that that for the 'zat–zaad' continuum. Consider first the 'stad–staat' continuum. In the middle of the continuum, listeners heard [st/\(\ddot{a}\)t] with an ambiguous vowel duration: The vowel is longer than the expected duration for [\(\ddot{a}\)] and shorter than that for [a:]. Listeners could decide that they heard either [st/\(\ddot{a}\)t] (\(\ddot{a}/\text{stad}\)) or [sta:t] ([sta:t]). Our results show that, in this situation, listeners decided that they heard [\(\ddot{a}\)] as in [st/\(\ddot{a}\)t], rather than [a:] as in [sta:t]. Recall that medial voiced consonants are preceded by slightly longer vowels than voiceless consonants. Rather than categorizing an ambiguous vowel as [a:], listeners assumed that the slightly longer [\(\ddot{a}\)] was the result of vowel lengthening preceding a voiced consonant. Consider next the 'zat–zaad' continuum. Again, listeners could decide that they heard either [zat] (\(\ddot{a}/\text{zaad}\)) or [za:t] ([za:t]). Our results show that, in this situation, they heard [a:] as in [za:t] rather than [\(\ddot{a}\)] as in [zat]. In contrast to the 'stad–staat' continuum, listeners concluded that they were not dealing with a slightly lengthened [a:] caused by vowel lengthening preceding a voiced consonant, since there is no lexical alternative [zaad] with a short vowel preceding a voiced final consonant. Listeners, therefore, concluded that this ambiguous vowel duration signals a change in vowel category from [a] to [\(\ddot{a}\)], causing them to respond with [a:] as in [za:t].

The present results indicate that the categorization of ambiguous vowel length appears to be affected by the underlying voicing of the following word-final consonant. Although listeners are hearing phonetically identical continua (vowel+voiceless consonant), there

---

3 Although one subject (\#7) shows unusually large boundary shifts, omitting this subject does not substantially alter the pattern of results: 'zat–zaad' vs. 'stad–staat': \( t(13) = 5.17, p < 0.0002 \); 'zat–zaad' vs. 'rat/d–raat/d': \( t(13) = 2.27, p < 0.04 \); 'stad–staat' vs. 'rat/d–raat/d': \( t(13) = 1.64, p > 0.13 \).
is a significant difference in phoneme boundary location between the ‘zat–zaad’ and ‘stad–staa’ continua. This effect appears to be due to the underlying phonological representation of the final consonant: namely, the underlying voicing of the final consonant seems to influence the categorization of vowel length, resulting in different phoneme boundaries for the two continua. As such, these data provide evidence that listeners actually use the underlying phonological representation in the perception and identification of words.

The present data also show that the phoneme boundary for the ‘rat/r-d–rat/d’ continuum is located between those for the ‘zat–zaad’ and ‘stad–staa’ continua. The ‘rat/r-d–rat/d’ continuum can in some sense be considered as a neutral condition, since phonological bias is absent here. That is, in the ambiguous vowel duration region, no phonological influence can bias the response, since all four combinations of vowel length and underlying voicing are possible. When using a vowel continuum in which the phonological voicing status of the final consonant was ambiguous, the identification function was seen to fall between the two continua.

It should be noted that the difference in phoneme boundary between the ‘zat–zaad’ and ‘stad–staa’ continua cannot be due to a difference in duration between the onsets of the stimulus words. Recall that [z] in the ‘zat–zaad’ continuum is shorter than [st] in the ‘stad–staa’ continuum. Sils and Cohen (1996b) report for Dutch that, as initial consonant duration increases, the duration of the following vowel decreases. A similar finding is reported for Danish by Fischer-Jørgensen (1964). Therefore, if anything, in the present experiment one would expect the [z] boundary to occur at a shorter vowel duration for ‘stad–staa’, which has a longer initial onset, than for ‘zat–zaad’, which has a shorter initial consonant. This is exactly opposite to the results obtained in the present perception experiment. It, therefore, seems unlikely that the present results are simply due to a difference in onset duration between the two continua.

A second possible influence, that of word frequency, can also be ruled out. The frequency of occurrence of the stimulus words (in a corpus of 720,000 words) is as follows: ‘zat’ 156; ‘zaad’ 9; ‘stad’ 150; ‘staa’ 475 (Uit den Boogaart, 1975). If it were simply word frequency that was driving listeners’ responses, we would expect a preference for short vowel (‘zat’) and long vowel (‘staa’) responses for stimuli in the ambiguous vowel duration region. However, our results show the opposite pattern. Responses in this region are determined by the voicing of the underlying consonant rather than word frequency.

A final possibility is that categorization was guided by an intermediate phonetic representation containing some sort of phonetic lengthening preceding voiced consonants. However, this representation never occurs, since the lexical rule of final devoicing applies before lengthening (Booij and Rubach, 1987), thus blocking any vowel lengthening. Instead of some intermediate phonetic representation, it is the combination of the representation of the final consonant as voiced along with the listener’s knowledge or experience that vowels are longer before voiced consonants which causes the observed effect. As soon as the vowel is unambiguously long (or short), there is no difference between the two word pairs. However, when the vowel duration falls in between the expected durations for L and S, all available knowledge is used to make a decision.

Our results indicate that the knowledge used is that of phonetic lengthening before a voiced consonant.

Taken together, the present results show that categorization of vowel duration in a vowel identification task significantly varies as a function of the underlying voicing of the syllable-final consonant. These findings strongly suggest that phonetic perception can be affected by the underlying phonological representation of a lexical item.

Converging evidence for this claim comes from a recent study by Lahiri, Jongman, and Sereno (1990). This experiment involved the processing of verb-clitic constructions in Dutch. In colloquial speech, when the clitic ‘der’ (“her”) attaches to a preceding verb, the verb-clitic combination leads to an optional voicing alternation on the surface. This alternation occurs when the clitic attaches to a host ending in an obstruent. For example, a verb stem ending in an underlying voiceless obstruent (e.g., ‘kus’, “to kiss”, [k[s]i] can surface with either a voiceless ([k[s]iär], “I kiss her”) or a voiced ([k[œ]d]är], “I kiss her”) consonant cluster. This alternation also occurs for verb stems ending in an underlying voiced obstruent. ‘Kiss’ ([k[s]i], “to choose”), for example, can surface either with a voiceless cluster ([k[s]iär], “I choose her”) or with a voiced cluster ([k[œ]d]är], “I choose her”). As a result, for each underlying verb stem, two clitic forms can surface: One that matches the underlying representation in terms of voicing, and another that mismatches the underlying representation in terms of voicing.

Exploiting this unique characteristic, Lahiri et al. (1990) used an auditory lexical decision task with a priming paradigm to determine whether the parsing of the cliticized forms and the eventual recognition of the verbs was affected by either surface phonetic representations or underlying phonological representations of the verb stem. In the
experiment, subjects had to make a lexical decision to a target item which was preceded by one of the two clitic forms as a prime. The prime was either the voiceless or the voiced clitic form, and the target was the imperative form of the same verb. For example, a listener would either hear [kast] or [kæzæ] followed by the target [kas]. Similarly, a listener would hear [kist] or [kæzæ] followed by [kiš]. Results showed a significant interaction between clitic form and obstruent voicing (see Figure 3). For verbs underlyingly ending in the voiceless obstruents /p, s/, responses to the voiceless clitic forms were faster than those to the voiced clitic forms. Similarly, for verbs underlyingly ending in the voiced obstruents /b, s/, responses to the voiced clitic forms were faster than those to the voiceless clitic forms. Although voiceless clitic forms always matched the verb stem in terms of voicing for both underlyingly voiceless (/kæs/) and underlyingly voiced (/kæs/) verbs (due to syllable-final devoicing of the verb stem), listeners had no overall preference for the voiceless clitic forms. That is, they did not prefer a clitic form with a voiceless consonant cluster over one with a voiced cluster. Instead, there was an asymmetry in response latencies to the same imperative form of the verb, depending on whether or not the listener heard the clitic form which matched the underlying phonological representation of that verb.

Together, the present results and those from Lahiri et al. (1990) provide compelling evidence that there is indeed an underlying abstract phonological representation differing from the surface pronunciation which influences the recognition and identification of words. The listener's perception is guided not just by available acoustic information but also by underlying phonological representations of words. In the clitics experiment, lexical decision times were faster when the acoustic prime matched the underlying voicing of the verb stem. In the present vowel categorization experiment, ambiguous phonetic information was not randomly categorized, but instead was interpreted via and influenced by lexical phonological representations. These results suggest that models of perception in phonetics and psycholinguistics must take into account the close link between phonological representations and acoustic-phonetic input.

---

4 Although this paper makes the claims as Lahiri and Marslen-Wilson (1991) with respect to abstractness of underlying representations, the present experiment does not address the issue of underspecification. Rather, the present study was designed to explore if the underlying representation of [voice] differed from what is observed on the surface.
A. Jongman, J.A. Sereno, M. Rastijmakers, and A. Lahiri


