LANGUAGE AND SPEECH, 1992, 35(1, 2), 137–152

THE PHONOLOGICAL REPRESENTATION OF [VOICE] IN SPEECH PERCEPTION*

ALLARD JONGMAN
Cornell University
JOAN A. SERENO
Cornell University
MARIANNE RAAIJMAKERS
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 msec 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 obstruent following the vowel differed in voicing. A vowel length continuum ([at] to [a:t]) 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: /zat/–/za:d/, and /stad/–/sta:t/. Our question was whether the vowel category boundaries would be different in pairs like /zat/–/za:d/ as compared to /stad/–/sta:t/. 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 surface 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.

INTRODUCTION

In recent years, a number of variables have been shown to affect the recognition of spoken words. These variables include lower-level acoustic cues, lexical cues such as word frequency, and higher-level influences such as syntactic and semantic context (e.g., Frauenfelder and Tyler, 1987; Marslen-Wilson, 1989). In this paper, we examine the role of a different sort of linguistic variable, namely, the phonological representation of a word. In particular, this study concerns the relationship between phonology and phonetics in research involving the representation and recognition of words. The work of John Ohala has time and again raised the issue of an intimate involvement of phonetics and phonology, supporting the view that the disciplines should be integrated rather than being kept entirely autonomous (e.g., Ohala, 1990). We approach this issue from a related but somewhat different perspective. We believe that the phonetic information of a word available to a listener is to a large extent interpreted with respect to the phonological representation of the relevant entries in the listener's lexicon, especially when the phonetic information is ambiguous. If indeed the phonological structure of lexical entries influences a listener's identification, acoustic information cannot be interpreted independently of phonological representation. We will present experimental data suggesting that phonetic information is interpreted differentially by a listener depending on phonological representations.

Little phonetic or psycholinguistic research has been devoted to the study of the underlying phonological representation of lexical form. Instead, much research has dealt with the unit of speech perception or lexical access: Is it the syllable, the segment, or the feature? (See, for example, Pisoni and Luce, 1987.) Implicit in this approach is the assumption that once this issue has been resolved, the process of speech perception and word recognition will be fairly straightforward: One simply has to extract the relevant units from the speech signal and concatenate them to recognize the intended words. We refer to this approach as a surface approach, since it deals with surface phonetic phenomena. In contrast, we posit a view which assumes that speech perception and word recognition are determined by the underlying phonological structure of the language system as well as the surface phonetic representation (i.e., by the speech signal).

As an example, consider the final consonant cluster in the word 'hand' (Lahiri and Marslen-Wilson, 1991). In the context 'hand you', the final [d] and the following glide [j] may become [dʒ], resulting in [hændʒu]; in the context 'hand me', the [d] may be deleted and [n] may become [m], resulting in [hæmi]; finally, in the context 'hand care', [d] may be deleted and [n] may become [ŋ] yielding [hæŋkɛr]. The same underlying lexical form /hænd/ can surface with four different final segments: [d], [dʒ], [m], and [ŋ]. This example shows that simply extracting and concatenating individual segments without taking into account the underlying representation may not always lead to access of the intended word.

The present research is an attempt to explore the nature of phonological representations and to determine whether listeners make use of them in the processing of speech. Our experiment focuses on the representation of voicing in word-final obstruents. Like many other Germanic languages, Dutch has a phonological contrast in voicing which is neutralized on the surface in syllable-final position due to a rule of syllable-final devoicing (Trommelen and Zonneveld, 1979; Booij, 1981).[1] For example, the two Dutch words 'pad' ("toad") and 'kat' ("cat") differ in terms of the underlying representation of the final consonant, with 'pad' (/pad/) ending in a voiced stop and 'kat' (/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' ([pat]) and 'kat' ([kat]) ending in a voiceless [t]. The morphological process of plural formation, as exemplified by related pairs such as 'pad' – 'padden' ("toads", [padən]) and 'kat' – 'katten' ("cats", [katən]), suggests that the underlying representation of the dental consonant is /d/ in the first pair and /t/ in the second pair.[2] 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 (Dinnsen, 1984; Dinnsen and Charles-Luce, 1984), and Polish (Slowiaczek and Dinnsen, 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, Fourakis and Iverson (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 Blumstein, 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

1 The presence of syllable-final devoicing does not necessarily mean that all syllable-final consonants surface as voiceless. Additional phonological processes such as assimilation can voice syllable-final consonants that had first been devoiced due to the devoicing rule (Trommelen and Zonneveld, 1979; Slis, 1986). In the present paper, all syllable-final consonants were also in utterance-final position, so that they surfaced as voiceless due to the rule of syllable-final devoicing.

2 The rule of syllable-final devoicing does not apply to words such as /padən/, since the medial /d/ is resyllabified as the onset of the second syllable (/pa.dən/).

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]) – 'taak' ("task", [ta:k]). 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; Sharf, 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", [ra:tən]) – 'raden' ("councils", [ra:dən]). These plural nouns are formed by adding the plural suffix [ən] to the stem. In these cases, all vowels (both long and short) are lengthened by approximately 20–30 msec when followed by a medial voiced consonant (Slis and Cohen, 1969b; Koopmans-van Beinum 1980; Nooteboom and Cohen, 1984; Slis, 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 (Slis 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 judged 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, Eimas, and Liberman, 1962; Nooteboom and Doodeman, 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; Connine 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 [d–t] phonetic categorization task, Ganong obtained more voiced 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 'dask–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 Shinn, Blumstein, and Jongman (1985) 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 voicing 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 voiced 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' (/zat/, [zat], "drunk") – 'zaad' (/za:d/, [za:t], "seed"); and 'stad' (/stad/, [stat], "city") – 'staat' (/sta:t/, [sta:t], "state"). Long [a:] 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 (Nooteboom, 1974). Furthermore, perception experiments

have shown that [a:] can successfully be shortened to obtain 100% [a] responses (Nooteboom and Doodeman, 1980). Notice that these word pairs contrast in terms of the combination of vowel length and underlying final consonant voicing. For 'stad–staat', the short vowel precedes the underlying voiced stop and the long vowel precedes the underlying voiceless 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 /stat/, /sta:d/, /zad/, and /za:t/.

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 MD211N 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 [a:] and digitally shortening it in 12 steps. Using both visual and auditory criteria, the long vowel [a:] 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 excising 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 [t] 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 [t].

Two experimental continua were constructed by appending the [at]–[a:t] continuum to different initial consonants, the [z] (144 msec) from 'zaad' and the [st] (244 msec) from 'stad', to create a 'zat–zaad' and a 'stad–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: /rat/ "rat", /rad/ "wheel", /ra:t/ "honeycomb", and /ra:d/ "council". A 'rat/rad–raat/raad' continuum was created by splicing the [r] from 'rad' (109 msec) onto the [at]–[a:t] 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.

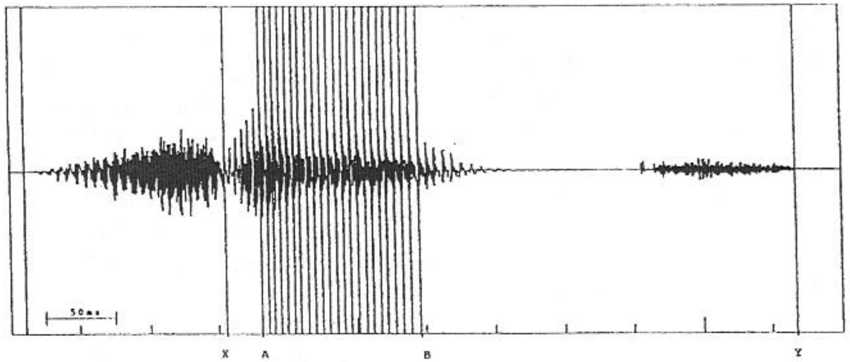

Fig. 1. Waveform of the test stimulus [za:t]. X and Y demarcate the entire VC portion, while A and B demarcate the steady-state vowel used to create the [a:t–at] continuum. Each vertical line during the steady-state vowel marks a pitch period. The vowel continuum was made by simultaneously excising a pitch period from each end of the steady-state vowel (see text).

Subjects

Fifteen university students from the subject pool at the Max Planck Institute for Psycholinguistics were paid to participate in the experiment. All were native speakers of Dutch and reported no history of speech or hearing disorders. All were naive as to the purpose of the experiment.

Procedure

All subjects were tested individually in a sound-isolated booth. The stimuli were presented to subjects from a DAT recorder (Sony TCD1000) through closed-ear headphones (Sennheiser HD520) at a comfortable loudness level. Each subject was presented with the 'zat-zaad', 'stad-staat', and 'rat/d–raat/d' continua. The presentation of the continua was counterbalanced across subjects. Subjects were told that the continuum endpoints were existing Dutch words and that they were required to identify the vowel of each stimulus word as short or long by pressing one of two clearly marked buttons on a response box placed in front of them. One button was labelled "a" and the other "aa". Preceding each continuum, subjects were given a set of 24 practice items. These practice items consisted of the 12 members of the continuum, repeated twice, in randomized order. The experiment lasted approximately 40 minutes.

RESULTS

The variable of interest was the location of the phoneme boundary: When comparing

TABLE 1

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

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

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', 'stad'), 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

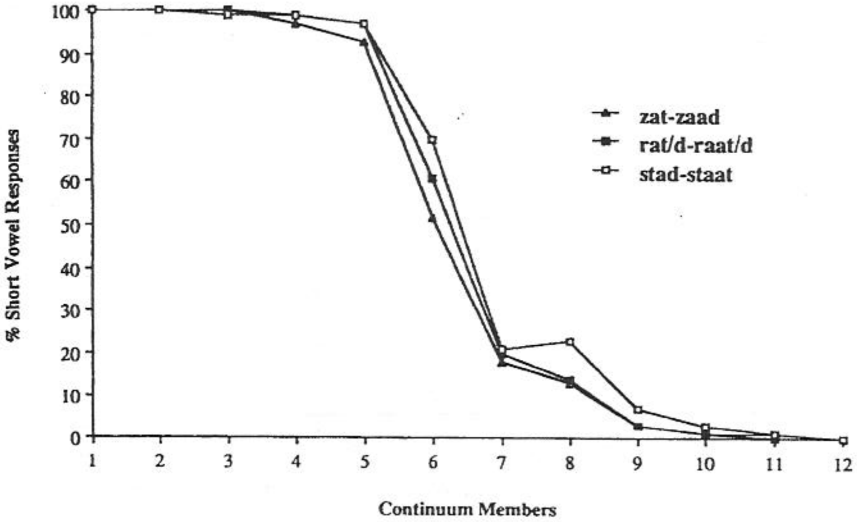

Fig. 2. Vowel identification functions for each of the three continua. Along the horizontal axis, the 12 continuum members are represented, ranging from 1 with the shortest vowel duration (83 msec) to 12 with the longest vowel duration (197 msec).

Another possible result is that there is a systematic and significant difference in phoneme boundary between the 'zat–zaad' and 'stad–staat' continua, since the two continua contrast in terms of their underlying representation. Since all stimulus words end in a phonetically voiceless [t], any difference in vowel categorization could only be due to the underlying voicing of the final consonant and not to the surface voicing. On this view, if the listener uses an abstract phonological representation of the stimulus words, (which ends in either /d/ or /t/), then one might expect that the underlying voicing of the consonant may influence the categorization of vowel length, resulting in a corresponding difference in phoneme boundaries for the two continua.

Table 1 shows the phoneme boundaries for each continuum for each subject. Boundary locations were derived by transforming subjects' responses to *z*-scores and performing a linear regression analysis. Focusing first on the 'zat–zaad' and 'stad–staat' continua, the results of the vowel categorization task show that the identification function for the 'zat–zaad' continuum is significantly different from that for the 'stad–staat' continuum (Figure 2). The identification functions for the two continua are not identical, even though in both conditions listeners heard the same vowel+voiceless consonant continuum. Figure 2 clearly shows that the phoneme boundary for the 'zat–zaad' continuum occurs at a shorter vowel length compared to the 'stad–staat' continuum. A pairwise *t*-test on phoneme boundaries for the 'zat–zaad' and 'stad–staat' continua was statistically significant [$t(14) = 3.44$, $p < 0.004$]. There was a

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' [$t(14) = 1.97$, $p < 0.07$].[3]

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 voiced consonant but as long when followed by an underlyingly voiceless 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 [stVt] with an ambiguous vowel duration: The vowel is longer than the expected duration for [a] and shorter than that for [a:]. Listeners could decide that they heard either [stat] (/stad/) or [sta:t] (/sta:t/). Our results show that, in this situation, listeners decided that they heard [a] as in [stat], 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 [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] (/zat/) or [za:t] (/za:d/). Our results thow that, in this situation, they heard [a:] as in [za:t] rather than [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 /zad/ 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 [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–staat' 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/d–raat/d' continuum is located between those for the 'zat–zaad' and 'stad–staat' continua: The 'rat/d–raat/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 other two continua.

It should be noted that the difference in phoneme boundary between the 'zat–zaad' and 'stad–staat' continua cannot be due to a difference in duration between the onset consonants of the stimulus words. Recall that [z] in the 'zat–zaad' continuum is shorter than [st] in the 'stad–staat' continuum. Slis and Cohen (1969b) 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 [a–a:] boundary to occur at a shorter vowel duration for 'stad–staat', 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 simulus words (in a corpus of 720,000 words) is as follows: 'zat' 156; 'zaad' 9; 'stad' 150; 'staat' 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 ('staat') 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 [a] and [a:], all available knowledge is used to make a decision.

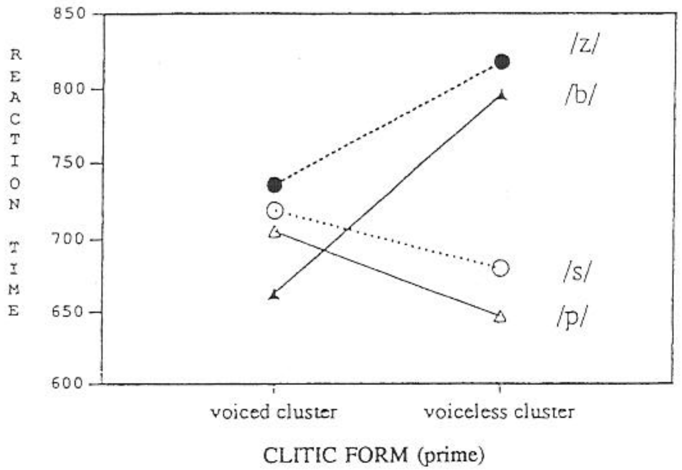

Fig. 3. Reaction times (in msec) to voiced and voiceless clitic forms as a function of their underlying stem-final consonants (/p, b, s, z/). From Lahiri *et al.* (1990).

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", /kœs/) can surface with either a voiceless ([Ikœstər], "I kiss her") or a voiced ([Ikœzdər], "I kiss her") consonant cluster. This alternation also occurs for verb stems ending in an underlying voiced obstruent. 'Kies' (/kiz/, "to choose"), for example, can surface either with a voiceless cluster ([Ikistər], "I choose her") or with a voiced cluster ([Ikizdər], "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 [Ikœstər] or [Ikœzdər] followed by the target [kœs]. Similarly, a listener would hear [Ikistər] or [Ikizdər] followed by [kis]. 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, z/, 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 (/kœs/) and underlyingly voiced (/kiz/) 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.[4] 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 same 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.

REFERENCES

BLUMSTEIN, S.E. (1991). The relation between phonetics and phonology. *Phonetica*, **48**, 108–119.

BOOIJ, G.E. (1981). *Generatieve Fonologie van het Nederlands.* Utrecht: Het Spectrum.

BOOIJ, G.E., and RUBACH, J. (1987). Postcyclic versus postlexical rules in lexical phonology. *Linguistic Inquiry*, **18**, 1–44.

BURTON, M.W., BAUM, S.R., and BLUMSTEIN, S.E. (1989). Lexical effects on the phonetic categorization of speech: The role of acoustic structure. *Journal of Experimental Psychology: Human Perception and Performance*, **15**, 567–575.

CHARLES-LUCE, J. (1985). Word-final devoicing in German: Effects of phonetic and sentential contexts. *Journal of Phonetics*, **13**, 309–324.

CHEN, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, **22**, 129–159.

CONNINE, C.M. (1986). *Modularity and Auditory Word Recognition.* Doctoral dissertation, University of Massachusetts.

CONNINE, C.M., and CLIFTON, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, **13**, 291–299.

DAVIS, S., and SUMMERS, W.V. (1989). Vowel length and closure duration in word-medial VC-sequences. *Journal of Phonetics*, **17**, 339–353.

DENES, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America*, **27**, 761–774.

DINNSEN, D.A. (1984). A re-examination of phonological neutralization. *Journal of Linguistics*, **21**, 265–279.

DINNSEN, D.A., and CHARLES-LUCE, J. (1984). Phonological neutralization, phonetic implementation, and individual differences. *Journal of Phonetics*, **12**, 49–60.

FISCHER-JΦRGENSEN, E. (1964). Sound duration and place of articulation. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, **17**, 175–207.

FOURAKIS, M., and IVERSON, G.K. (1984). On the 'incomplete neutralization' of German final obstruents. *Phonetica*, **41**, 140–149.

FOX, R.A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, **10**, 526–540.

FRAUENFELDER, U., and TYLER, L.K. (eds.) (1987). *Spoken Word Recognition.* Cambridge, MA: MIT Press.

FRY, D.B., ABRAMSON, A.S., EIMAS, P.D., and LIBERMAN, A.M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, **5**, 171–189.

GANONG, W.F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, **6**, 110–125.

GOTTFRIED, T.L., MILLER, J.L., and PAYTON, P.E. (1990). Effect of speaking rate on the perception of vowels. *Phonetica*, **47**, 155–172.

HARNAD, S. (ed.) (1987). *Categorical Perception.* Cambridge, U.K.: Cambridge University Press.

HOUSE, A.S., and FAIRBANKS, G. (1953). The influence of consonantal environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, **25**, 105–113.

JASSEM, W., and RICHTER, L. (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics*, **17**, 317–325.

JONGMAN, A., and SERENO, J.A. (1991). On vowel quantity and post-vocalic consonant duration in Dutch. *Proceedings of the XIIth International Congress of Phonetic Sciences, Vol. 2*, 294–297.

KOOPMANS-VAN BEINUM, F.J. (1980). *Vowel Contrast Reduction: An Acoustic and Perceptual Study of Dutch Vowels in Various Speech Conditions.* Amsterdam: Academische Pers.

KUHL, P.K. (1982). Speech perception: An overview of current issues. In N.J. Lass, L.V. McReynolds, J.L. Northern, and D.E. Yoder (eds.), *Speech, Language, and Hearing, Vol. I* (pp. 286–322). Philadelphia: W.B. Saunders Company.

LAHIRI, A., JONGMAN, A., and SERENO, J.A. (1990). The pronominal clitic [dər] in Dutch: A theoretical and experimental approach. In G.E. Booij and J. van Marle (eds.), *Yearbook of Morphology, Vol. 3* (pp. 115–127). Dordrecht: Foris Publications.

LAHIRI, A., and MARSLEN-WILSON, W.D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, **38**, 245–294.

LAHIRI, A., SCHRIEFERS, H., and KUIJPERS, C. (1987). Contextual neutralization of vowel length: Evidence from Dutch. *Phonetica*, **44**, 91–102.

LUCE, P.A., and CHARLES-LUCE, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal of the Acoustical Society of America*, **78**, 1949–1957.

MARSLEN-WILSON, W.D. (1989). Access and integration: Projecting sound onto meaning. In W.D. Marslen-Wilson (ed.), *Lexical Representation and Process* (pp. 3–24). Hillsdale, NJ: Lawrence Erlbaum.

MCQUEEN, J.M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, **17**, 433–443.

MILLER, J.L., and DEXTER, E.R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 369–378.

NOOTEBOOM, S.G. (1972). *Production and Perception of Vowel Duration*. Doctoral dissertation, Rijksuniversiteit Utrecht.

NOOTEBOOM, S.G. (1974). Some context effects on phonemic categorization of vowel duration. *IPO Annual Progress Report*, **9**, 47–55.

NOOTEBOOM, S.G., and COHEN, A. (1984). *Spreken en Verstaan: Een Inleiding tot de Experimentele Fonetiek*. Assen: Van Gorcum.

NOOTEBOOM, S.G., and DOODEMAN, G.J.N. (1980). Production and perception of vowel length in spoken sentences. *Journal of the Acoustical Society of America*, **67**, 267–287.

OHALA, J.J. (1990). There is no interface between phonology and phonetics: A personal view. *Journal of Phonetics*, **18**, 153–171.

PETERSON, G., and LEHISTE, I. (1960). Duration of syllabic nuclei in English. *Journal of the Acoustical Society of America*, **32**, 693–703.

PISONI, D.B., and LUCE, P.A. (1987). Acoustic-phonetic representations in word recognition. In U. Frauenfelder and L.K. Tyler (eds.), *Spoken Word Recognition* (pp. 21–52). Cambridge, MA: MIT Press.

POLS, L.C.W., TROMP, H.R.C., and PLOMP, R. (1973). Frequency analysis of Dutch vowels from 50 male speakers. *Journal of the Acoustical Society of America*, **53**, 1093–1101.

PORT, R.F., and CRAWFORD, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics*, **17**, 257–282.

PORT, R.F., and O'DELL, M.L. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics*, **13**, 455–471.

RAPHAEL, L.J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English. *Journal of the Acoustical Society of America*, **51**, 1296–1303.

REPP, B.H. (1983). Categorical perception: Issues, methods, findings. In N.J. Lass (ed.), *Speech and Language: Advances in Basic Research and Practice, Vol. 10* (pp. 243–335). New York: Academic Press.

SHARF, D. (1962). Duration of post-stress intervocalic stops and preceding vowels. *Language and Speech*, **5**, 26–30.

SHINN, P.C., BLUMSTEIN, S.E., and JONGMAN, A. (1985). Limitations of context-conditioned effects in the perception of [b] and [w]. *Perception and Psychophysics*, **38**, 397–407.

SLIS, I.H. (1985). *The Voiced-Voiceless Distinction and Assimilation of Voice in Dutch*. Helmond: Wibro.

SLIS, I.H. (1986). Assimilation of voice in Dutch as a function of stress, word boundaries, and sex of

SLIS, I.H., and COHEN, A. (1969a). On the complex regulating the voiced-voiceless distinction, I. *Language and Speech*, **12**, 80–102.

SLIS, I.H., and COHEN, A. (1969b). On the complex regulating the voiced-voiceless distinction, II. *Language and Speech*, **12**, 137–155.

SLOWIACZEK, L.M., and DINNSEN, D.A. (1985). On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics*, **13**, 325–341.

TROMMELEN, M., and ZONNEVELD, W. (1979). *Inleiding in de Generatieve Fonologie*. Muiderberg: Coutinho.

UIT DEN BOOGAART, P.C. (1975). *Woordfrequenties in Geschreven en Gesproken Nederlands*. Utrecht: Oosthoek, Scheltema & Holkema.