

Facilitatory Effects of Vowel Epenthesis on Word Processing in Dutch

Wilma van Donselaar, Cecile Kuijpers, and Anne Cutler

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

We report a series of experiments examining the effects on word processing of insertion of an optional epenthetic vowel in word-final consonant clusters in Dutch. Such epenthesis turns *film*, for instance, into *filəm*. In a word-reversal task listeners treated words with and without epenthesis alike, as monosyllables, suggesting that the variant forms both activate the same canonical representation, that of a monosyllabic word without epenthesis. In both lexical decision and word spotting, response times to recognize words were significantly faster when epenthesis was present than when the word was presented in its canonical form without epenthesis. It is argued that addition of the epenthetic vowel makes the liquid consonants constituting the first member of a cluster more perceptible; a final phoneme-detection experiment confirmed that this was the case. These findings show that a transformed variant of a word, although it contacts the lexicon via the representation of the canonical form, can be more easily perceptible than that canonical form. © 1999 Academic Press

Key Words: word recognition; vowels; epenthesis; Dutch.

No one act of spoken-word recognition is exactly like another. Even the same word spoken by the same person to the same listener in the same room may occur in a different context or against different ambient sound. Variability arising from talker differences and environmental conditions has prompted an enormous volume of research in speech perception and has formed one of the principal issues along which models of speech perception and spoken-word recognition divide: whether at some level of processing invariant cues to sounds and words may be abstracted and represented.

As if the infinite variability offered by talker and environmental factors were not enough, however, the listener's lot is further compli-

cated by variation as a function of the speech context. The precise form of speech sounds differs as a function of the other sounds that surround them; phonemes are not articulated separately, but are coarticulated with other sounds and vary in accord with the characteristics of their neighbours in the speech signal. This process can alter with speech rate, as well, such that at faster rates of speech coarticulation may result in greater contextual effects or effects which extend across a wider neighborhood; thus it adds another way in which variability complicates spoken-word recognition. And finally, as almost the coup de grace, yet further variation is permitted as a function of certain phonological processes. Thus sounds may assimilate to their phonological context, so that in English, for example, a phrase like *hot cakes* may be pronounced with either a /t/ or a /k/ at the end of the first syllable.

One such form of phonologically determined variation is the epenthesis, or insertion, of isolated sounds where no such sound exists in the underlying form of the word. Epenthesis is a mechanism of historical change in word forms; thus the English word *oven*, Dutch *oven*, and German *Ofen* all have an ancestor, without the second (weak) vowel, observable in the Old

Thanks to Arie van der Lugt, Dirk Janssen, Ellen Gerrits, Mark Scholten, Peter van de Pol, Joost Beekman, Mattijn Morren, and Bram de Kruijff for technical assistance. Thanks also to Allard Jongman, Natasha Warner, and two anonymous reviewers for comments on an earlier version of the report. W.v.D. is now with the Netherlands Organisation for Scientific Research (e-mail address: donselaar@nwo.nl). C.K. is now with the Interfaculty Research Unit for Language and Speech, University of Nijmegen (e-mail address: Cecile.Kuijpers@mpi.nl).

Address correspondence and reprint requests to Anne Cutler, Max Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands. E-mail: anne.cutler@mpi.nl. Fax: +31 24 3521213.



Norse form *ofn*. In current English, such insertions can be observed in many casual speech forms, and they can involve either vowels—e.g., a schwa inserted into a word-final consonant cluster, as when *film* is pronounced *filəm*—or consonants—e.g., the [p] that can intrude into the syllable boundary in the word *something*. Such word-internal epenthesis in English, like other intrusions (e.g., the [r] that can occur at the vowel–vowel word boundary in a sequence like *idea of*), do not occur frequently in careful speech.

The situation is different, however, in other languages. There are two ways in which epenthesis can form part of even careful utterances in a language. One is as a result of constraints on permissible sequences; for instance, if a language does not allow consonant clusters, then loan forms with consonant clusters can only conform to the phonology via vowel epenthesis. This may be observed in numerous loan forms from English and other languages into Japanese: glass becomes *garasu*, *Zeitgeist tsaitogai-suto*, MacDonaldd *Makudonarudo*. Here epenthesis is only a meaningful notion to compare such forms with the originals on which they are based: Japanese phonology requires the vowels in these forms, and they are in no sense optional.

Another way in which epenthesis can be rife, however, is for the phonology to allow it as an optional variant. This is the situation with vowel epenthesis in Dutch. Words like *tulp* “tulip,” *werk* “work,” and *film* “film” are routinely pronounced *tuləp*, *werək*, and *filəm*, respectively. Such variants are for the majority of speakers entirely comparable in acceptability to the variants without epenthesis. It is important to stress this point for an audience which knows English, because in this respect the situation in Dutch and English is very different. Thus many speakers of English would recognize the pronunciation of *film* as *filəm* as a possible variant, but most would regard it as a form with low social prestige. In Dutch this tendency is very much less strong—the word *film* also exists in Dutch, and pronunciation as *filəm* is a very widely used optional form. This form may be used more often by some speakers than by others, in some

regional areas than in others, or in some speech contexts than in others, but for many speakers it is unmarked with respect to social prestige. (Those speakers who do perceive a difference in prestige between the forms, however, regard the form without epenthesis as standard and the form with epenthesis as lower in prestige, as in English. We are not aware of any speakers who have the opposite perception.)

Vowel epenthesis in Dutch occurs in coda clusters consisting of liquids followed by another consonant other than /s/ or /t/ (Booij, 1995); because stops and fricatives in syllable-final position in Dutch must be voiceless, the clusters which allow epenthesis thus involve /l/ or /r/ followed by /k,x,p,f,m,n/. Many of these sequences, it should be noted, also occur separated by syllabic schwa at the end of bisyllabic words such as *billijk* (/bɪlək/ “reasonable”) or *knorrig* (/knɔrɔx/ “grouchy”), and indeed it is possible to find minimal pairs of words with epenthetic versus syllabic schwa, such as *wiləg/willig* (“willow/willing”), *hoorən/horen* (“horn/hear”), and *baləg/ballig* (“bellows/snooty”). (Luckily for the listener, however, there are few such pairs, and some of the words involved are rare.) This type of epenthesis in Dutch is not forced by constraints of the phonology; onset and coda clusters are acceptable, indeed common, in Dutch words. There appears to be no pressure to avoid clusters in other optional processes. For instance, nicknames and other word formation processes in Dutch do not avoid clusters—thus someone named *Marcus* can be known as *Marc*, or *Nicolaas* can be *Klaas*, and someone with the function of *directeur* (“director”) may be referred to as the *dirk*.

Spectrograms of words with epenthesis show a clearly vocalic portion separating the two consonants of the cluster. It is important to stress this point, since phonological accounts of epenthetic forms do not necessarily involve insertion of a vowel segment. For instance, accounts of speech production based on gestural analysis (e.g., Browman & Goldstein, 1990, 1992) allow for greater or lesser separation of the various gestural components involved in articulation and explain epenthesis in terms of exaggerated separation of the gestures related to the two

consonants of the cluster, such that for a short interval between the gestures the vocal tract is not constricted. Continued voicing during this period produces a vocalic signal equivalent to that of a centralized vowel, even though the speaker has not actually intended the production of such a vowel; Browman and Goldstein thus refer to such epenthetic vowels as “targetless schwa.” Whether intended by the speaker, however, the vocalic portion is clearly present in the realizations of these words with epenthesis, so that the acoustic difference between, say, *tulp* and *tuləp* is that the latter contains a vowel inserted between the two consonants of the final cluster.

Because of the general acceptability of non-coronal clusters in Dutch, and the fact that optional variants with epenthesis simply coexist with, rather than threaten, the nonepenthesized forms, there has been considerable discussion of why Dutch vowel epenthesis occurs. In general, articulatory accounts have been preferred: Liquids followed by noncoronal obstruents do not share place of articulation, and rapid transition between places of articulation may require considerable articulatory effort (Booij, 1995). Certainly epenthetic forms are closer to the alternation of consonants and vowels, which is the preferred pattern across languages—obligatory in many, illegal in none (Bell & Hooper, 1978; Treiman, Salasoo, Slowiaczek, & Pisoni, 1982). Consistent with the articulatory ease account is the fact that schwa epenthesis is rife in the Dutch spoken by children (Wijnen, Krikhaar, & Den Os, 1994), as well as the finding that the frequency of epenthetic insertion varies with rhythmic context. Thus just as slips of the tongue are more likely to result in an utterance which is rhythmically more regular than the intended utterance would have been (Cutler, 1980), so is epenthesis more likely when it results in a rhythmically more regular output (Kuijpers & Donselaar, 1998).

In the present study, however, we are not concerned with the phonology of epenthesis nor with the explanation of its use in speech production. Instead we focus on the consequences of epenthesis for the listener. Surely, one might imagine, changing the acoustic realization of a

word, whether it makes life easier for speakers, cannot be in the best interests of listeners.

A good deal of recent work has addressed this question with respect to contextually triggered phonological variation (see Cutler, 1998, for a review). Particular attention has been paid to assimilation of place of articulation, as in the *hot cakes* example referred to above. Such assimilation phenomena, this work shows, do not in general result in processing deficits. Thus recognition of a word-initial phoneme (such as the initial sound of *cakes* in the example) is not affected by whether the preceding sound has been subjected to an optional assimilation process (Gaskell & Marslen-Wilson, 1996, 1998; Koster, 1987; Kuijpers & Donselaar, forthcoming; Marslen-Wilson, Nix, & Gaskell, 1995). Only when an unacceptable phonetic sequence has been produced—e.g., by application of an assimilation rule in an inappropriate context—is processing impaired (Gaskell & Marslen-Wilson, 1996; Kuijpers & Donselaar, forthcoming; Otake, Yoneyama, Cutler, & van der Lugt, 1996).

Studies of the role of optional epenthesis in perception have not, as far as we know, previously been undertaken. There have been perceptual studies of the perception of consonant sequences by Japanese listeners which shed some light on the general question of cluster processing with and without vowel insertion. Thus Kashino, van Wieringen, and Pols (1992) compared Dutch and Japanese listeners in a consonant (cluster) identification task. They found that the two listener groups were equally accurate in identifying intervocalic consonants in VCV stimuli. Dutch listeners, however, were significantly more accurate in identifying the consonants of VC1C2V stimuli than Japanese listeners. Kashino et al. therefore concluded that Japanese listeners are unable to perceive consecutive consonants accurately. Similarly Dupoux, Kakehi, Hirose, Pallier, and Mehler (submitted for publication) showed that Japanese listeners could not reliably judge which of a pair of nonwords such as *ebuzo* and *ebzo* they had just heard, although French listeners could perform this task (in contrast, the French listeners could not perform the same judgment with pairs

such as *ebuzo* and *ebuuzo*, with a lengthened medial vowel, whereas the Japanese listeners could). The phonotactics of the language (i.e., the obligatory nature of epenthesis in consonant clusters) apparently lead the Japanese to “correct” deviant forms, in perception as well as in production. A more recent lexical decision study by Mehler and colleagues (personal communication) suggests that the phonological effects are brought into play before word candidates are lexically activated. In a lexical decision task, impossible words in Japanese such as *komgi* were accepted as accurately and as rapidly as their real-word equivalents such as *komugi*, and similarly rejection of nonwords such as *namda* and *namuda* was equivalent. All these results suggest that Japanese listeners rely on their phonological knowledge to construct perceptual representations of spoken words, and their phonological knowledge tells them that consonant sequences must be broken up by vowels.

However, as we saw above, Dutch listeners are not constrained by phonological prohibition of consonant sequences; epenthesis in words like *tulp* is optional. Nor do the optional assimilation effects which have been studied constitute a direct comparison for the case of optional epenthesis. In many cases the assimilation crosses a word boundary (as in *hot cakes*), and the very fact that this occurs in a multiword context means that of course information is available in the context to inform the listeners’ interpretation. But optional epenthesis can apply to a word uttered in isolation. The single-word answer to a question about the Dutch national flower can as easily be *tuləp* as *tulp*. In the experiments reported below, we ask how Dutch listeners cope with this variation. Is either of the variants *tulp* and *tuləp* easier for a Dutch listener to process than the other? Are both forms processed as variants of the same canonical form, such that they activate the same phonological representation, or is there no canonical form because each form is accorded separate storage?

We begin with the latter issue. Separate storage for optional forms could be a simple way to avoid problems of processing. No conversion or

decryption processes would be required; both *tulp* and *tuləp*, *film* and *filəm* would possess access representations, and each representation could be tied to the same meaning information. Neither form would be preferred, neither would have the status of canonical form; each would simply be activated by appropriate input.

One way to address the question of whether access representations are separately stored is via the question of syllabicity. Obviously, adding a vowel between two consonants adds an extra syllable to the word; the optional form with epenthesis has one more syllable than the underlying form without. If access representations are separate, one will have one syllable more than the other. In an earlier experiment (Donselaar, Kuijpers, & Cutler, 1996), we attempted to address this issue using the fragment detection technique (Frauenfelder & Kearns, 1996), comparing response time to detect either the target TUL or the target TULP in *tulp* versus *tuləp*. In two experiments, we found no difference in RT to TUL targets as a function of the input form; a differential effect for TULP targets observed in an initial experiment disappeared when listeners could not rely on simple acoustic matching. This result seems to suggest no difference in the representation accessed by the input *tulp* and the input *tuləp*. However, there is reason to be cautious in drawing such a conclusion from the detection study. On the one hand, the finding is a null result; it does not offer direct evidence for a single representation. And on the other hand, the fragment detection task has produced conflicting findings with respect to whether Dutch listeners are sensitive to the syllable structure of speech input (Zwitserslood, Schriefers, Lahiri, & Donselaar, 1993; Vroomen & De Gelder, 1994), and there is further evidence that apparent syllabic effects are extremely dependent upon the phonetic structure of the input (Donselaar & Stoutjesdijk, 1993; Frauenfelder, Rietveld, & van Til, submitted for publication).

In the present study, therefore, we adopted a completely different technique, one which has recently been demonstrated to give a reliable reflection of listeners’ apprehension of syllabic structure: the syllable reversal technique of

Treiman and Danis (1988; see also Schiller, Meyer, & Levelt, 1997). This technique enables us to study how the input forms *tulp* and *tuləp* are processed; evidence for a single phonological representation for both would be provided not by the absence of an effect, but by parity of subjects' responses given each input form.

EXPERIMENT 1

In the paradigm employed by Treiman and Danis and Schiller et al., subjects listen to bisyllabic words and have to produce the second syllable first and then—after a clear break—the first syllable. Since we were however interested in the representation of words with epenthesis, which could be either monosyllabic or bisyllabic, we adapted the paradigm slightly. Subjects heard a mixture of monosyllabic and bisyllabic words and were asked to reverse the monosyllables phoneme by phoneme (e.g., /tap/ /pat/) and the bisyllables syllable by syllable (e.g., /ho.təl/ /təl.ho/). This combination of auditory input and spoken output minimizes the chance for knowledge of orthographic representation to play a role (Schiller et al., 1997). The experiment included a large number of fillers, either unambiguously monosyllabic or unambiguously bisyllabic. The experimental question was whether words with epenthesis (*tuləp*) would be reversed in the monosyllabic or the bisyllabic fashion.

The experimental words were tested both in forms with epenthesis (*tuləp*) and without (*tulp*). The materials further included pseudo-words, and item type was blocked such that the first half of the experiment comprised pseudo-words and the second half real words.

Method

Stimuli. Sixteen real words with a final cluster of liquids and noncoronals were selected, such that the reversed words were nonsense words. Sixteen pseudo-words were created as well. The reversed pseudo-words were also nonsense words. The stimuli are listed in the Appendix. Each stimulus word was realized twice: with and without epenthesis.

One hundred and twenty-eight real word fillers were selected: 32 bisyllabic words with schwa

in the second syllable (e.g., *bakker* “baker”), 32 initially stressed bisyllabic words with two full vowels (e.g., *kosmos* “cosmos”), 16 monosyllabic words with an initial consonant cluster (e.g., *spek* “bacon”), 32 monosyllabic words with a single initial and final consonant (e.g., *kaap* “cape”), and 16 monosyllabic words with a final consonant cluster not subject to optional epenthesis (e.g., *vals* “false”). All items produced phonotactically legal sequences when reversed according to the instructions. One hundred and twenty-eight pseudo-word fillers were constructed by analogy to the real-word fillers. The real- and pseudo-word sets thus contained equal numbers of monosyllabic and bisyllabic items. Two sets of 26 practice items, representing all the above types, were also created, one for real and the other for pseudo-words.

Procedure. Two experimental sets of stimuli were constructed, and the experimental items (with or without schwa) counterbalanced across these sets. Thus, for instance, *tulp* occurred in set 1, while *tuləp* occurred in set 2. Subjects were randomly assigned to one of these sets and tested in groups of two or three in separate sound-attenuated booths. They were told to listen to the items and to reverse them: They were explicitly instructed to pronounce the item back to front if they believed the item consisted of just one syllable and to change the order of the two syllables if they thought the item had two syllables. Responses were registered by voice key and recorded on DAT. After the subjects had read the instruction, the experiment began with 26 pseudo-word practice items, followed by the block with pseudo-words. After the first block, there was a short break before the second block, which started with a practice series of real words. The interstimulus interval was 3.5 s. The experiment lasted approximately 40 min.

Participants. Forty-eight University of Nijmegen students, all native speakers of Dutch, participated in the experiment, in return for a small payment. Twenty-four heard each stimulus set.

Results and Discussion

The participants generally had little difficulty carrying out the task and failed to reverse on

TABLE 1

Reversal Task: Mean Percentage of Monosyllabic Responses for Real Words and Pseudo-words Realized with Versus without Epenthesis, Experiment 1

| | With epenthesis (%) | Without epenthesis (%) |
|--------------|------------------------|------------------------|
| Real words | (<i>tuləp</i>) 93.75 | (<i>tulp</i>) 98.18 |
| Pseudo-words | (<i>nuləp</i>) 57.81 | (<i>nulp</i>) 98.7 |

only 1% of the trials. The recorded responses were coded as monosyllabic or bisyllabic reversals. Analyses of variance were conducted on the proportion of monosyllabic responses across participants and across items. Table 1 gives the proportion of monosyllabic responses as a function of Word Type (pseudo-word/real word) and Epenthesis (plus/minus).

As expected, real words elicited more monosyllabic responses (96%) than pseudo-words did [78%; $F_1(1,47) = 62.24, p < .001$; $F_2(1,30) = 48.27; p < .001$], and items without epenthesis more monosyllabic responses (99%) than items with epenthesis [76%; $F_1(1,47) = 77.68, p < .001$; $F_2(1,30) = 77.69, p < .001$]. There was also a significant interaction between Word Type and Epenthesis: $F_1(1,47) = 60.57, p < .001$; $F_2(1,30) = 47.26, p < .001$; analysis of the components of this interaction showed that there was no difference in the proportion of monosyllabic responses to real and to pseudo-words without epenthesis (98% versus 99%), but real words with epenthesis received significantly more monosyllabic responses (94%) than pseudo-words with epenthesis [58%; $F_1(1,94) = 53.93, p < .001$; $F_2(1,30) = 47.86, p < .001$].

These results suggest that real words with epenthetic schwa (*tuləp*) are generally held to be monosyllabic. Listeners treated them as monosyllabic items significantly more often than they did the matched pseudowords (*nuləp*). Of course, the latter have no established representation; their syllabic structure can be derived only by analogy with real words. Pseudo-words without schwa (e.g., *nulp*) were expected to be reversed as monosyllables, and this was indeed the case. The prediction for pseudo-words with

schwa epenthesis (e.g., *nuləp*) was less straightforward since some of these pseudo-words (e.g., *nerəg* and *stilək*) had endings which made them resemble real words with suffix syllables—*ig* (/əx/ “ish”) and *-lijk* (/lək/ “like”). These words were therefore ambiguous: they could be perceived either as monosyllables (*nerg*, *stilk*, similar to existing words like *berg*, *spalk*) or as bisyllables (*nerriig*, *stillijk*, similar to words like *knorriig*, *billijk*). The results showed that both of these response options were used. In fact, closer inspection of the pseudo-words with schwa revealed a difference between words with /r/ clusters and /l/ clusters: The former elicited more monosyllabic responses (70%) than the latter (48%). This difference can be explained by the fact that five of the nine /l/ clusters were, with inserted schwa, potential suffixes (*-lg*, *-lk*), but only one of the seven /r/ clusters could have been a suffix (*-rg*). Clearly, then, the listeners were choosing their response alternatives for the pseudo-words on the basis of their knowledge of patterns in real Dutch words; their responses for the real words, however, were chosen on the basis of a canonical representation as monosyllables.

From Experiment 1 we may conclude, therefore, that realizations of real words with schwa epenthesis are represented by listeners as monosyllabic, in essentially the same manner as the schwa-less realizations of the same words. There seems, in other words, to be a unitary representation of words like *tulp* whether they are realized with or without epenthetic schwa; the canonical form of both *tulp* and *tuləp* is *tulp*.

In the remaining experiments, we address the implications of this finding for the perception of words which allow epenthetic schwa. The canonical form is *tulp*, but listeners are quite likely to hear either *tuləp* or *tulp*. Is the canonical form perceptually favored?

We first addressed this question via a simple task: lexical decision (for a review of this task, see Goldinger, 1996). In Experiment 2 listeners heard real words and pseudo-words and classified them as such; we asked whether *tulp* and *tuləp* differed in how quickly and accurately they could be accepted as a known word.

TABLE 2

Lexical Decision: Mean Percentages of YES Responses per Item Type, Experiment 2a

| | With epenthesis (%) | Without epenthesis (%) |
|--|---------------------|------------------------|
| Real word, optional schwa (<i>balk</i>) | 92.56 | 93.45 |
| Nonword, optional schwa (<i>golk</i>) | 0.29 | 0.59 |
| Real word, prohibited schwa (<i>hals</i>) | 1.79 | 85.12 |
| Real word, syllabic schwa (<i>billijk</i>) | 81.85 | 9.52 |

EXPERIMENT 2a

Auditory lexical decision can be conducted either as a “go/no-go” procedure, in which listeners give a response only when they recognize a real word, or as a yes/no procedure, in which listeners respond YES to real words and NO to nonwords. In Experiment 2a we used the go/no-go procedure. Although words like *tulp* or *balk* are real words in both their apparently canonical form without epenthesis and their phonological variant with epenthesis, listeners might respond more rapidly and more accurately to the canonical form. As well as words like *balk*, we included as experimental items in Experiment 2a different realizations (plus/minus schwa) of words that do not allow epenthesis because of their homorganic clusters (e.g., *hals*) and words that have syllabic schwa (e.g., *billijk*). The pseudo-words also had various phonological forms.

Method

Materials. The materials, listed in the Appendix, consisted of 12 words with optional schwa epenthesis (*balk*, *balək*), 12 nonwords with optional schwa epenthesis (**golk*, **golək*), 12 words with prohibited schwa epenthesis (*hals*, **haləs*), and 12 words with syllabic schwa (**bilk*, *billijk*). The number of real words thus equalled the number of pseudo-words. Apart from these 48 experimental items, there were also 16 practice words and 144 filler items, 50% of which were pseudo-words and 50% real words. All items were read by a phonetically trained native female speaker of Dutch in one session and recorded on digital audiotape. The experimental words were realized twice: with and without epenthesis. The words were digi-

tized with a sampling frequency of 20 kHz, and each word’s duration was established, using the ESPS and WAVES+ software. Surprisingly, the real words proved to be on average shorter in duration with epenthesis (mean 573 ms) than without (mean 600 ms).

Procedure. The materials were arranged in two sets, with epenthesis (plus/minus) counter-balanced across sets for the real words and pseudo-words which occurred in two versions. Thus, for instance, *balk* and *ralək* occurred in set 1, and *balək* and *ralk* in set 2. Listeners were randomly assigned to sets and tested in pairs, in separate sound-attenuated booths. The stimuli were presented via Sennheiser headphones. The participants were instructed to decide as quickly as possible whether each stimulus was a real word; they signaled their YES responses to real-word stimuli only, by pressing the single response button. The response window was 1500 ms, and reaction times were measured from word onset. The experimental system used was NESU on a Hermac AT computer. The entire experiment took approximately 20 min.

Participants. Fifty-six Nijmegen University students, all native speakers of Dutch, took part in the experiment in return for a small payment. None had taken part in Experiment 1.

Results and Discussion

The percentage of YES judgements and the response times (in ms) were each subjected to separate analyses of variance across participants and across items. Table 2 shows the proportion of YES responses for each item type. For the real words which allow epenthesis (*balk*), realizations with and without schwa produced exactly equivalent performance. Apparently, both

forms are equally acceptable real words. Similarly, pseudo-words ending with the same kinds of cluster as those in the real-word set (e.g., *golk*) were rejected to an almost exactly equivalent degree in both realizations; both seem thus to be equally nonword-like.

The more interesting cases are the two remaining items sets. For words like *hals*—with homorganic coda clusters of liquids and coronals—epenthesis is prohibited, and the experimental variants with schwa were therefore expected to be considered nonwords. Words like *billijk* have a schwa in the second syllable in their canonical representation; since there is no option to delete this schwa, the variants without schwa were expected to be considered nonwords as well. As Table 2 shows, these predictions were largely correct; however, an unexpected difference emerged between the responses to forms like *haləs* and *bilk*. In judging *haləs*, subjects did not confuse this form with the real word *hals* (the acceptance rate was only 2%). Forms such as *bilk*, however, in which syllabic schwa had been deleted leaving an /lk/ cluster which in principle allows epenthesis, produced 10% real-word responses. This difference between the two types of illegality—*haləs* vs. *bilk*—was significant [$F_1(1,55) = 11.87$, $p < .001$; $F_2(1,11) = 5.76$, $p < .05$].

In other words, listeners judging the lexicality of a spoken item apparently draw on phonological information on the possibility of schwa epenthesis: any /lk/ cluster is a potential /lək/ cluster, whereas a /ləs/ cluster cannot be a /ls/ cluster. Forms like *bilk* led to more confusion than forms like *golk*, since the former but not the latter have a word counterpart with a schwa.

For the items in which one cell received no or very few YES responses, it was not possible to analyze the effect of presence versus absence of schwa on response times. RTs, however, were analyzed for the *balk* words. Here we were interested to observe whether realizations without schwa, which Experiment 1 had suggested to be the canonical form, would be processed faster than realizations with schwa. Surprisingly, the realizations without schwa were recognized significantly less rapidly (950 ms) than

the forms with schwa [911ms; $F_1(1,55) = 13.66$, $p < .001$; $F_2(1,11) = 6.63$, $p < .05$].

The reaction times and acoustic durations showed a weak but significant correlation ($r = .18$, $p < .01$). Additional analyses of variance were thus carried out, covarying RT with word duration. The effect of epenthesis still held in the analysis by subjects [$F_1(1,54) = 11.04$, $p < .01$] but failed to reach significance in the analysis by items. To explore further the obvious variability in the item set, we compared items with /l/ and /r/ clusters, since this factor had produced some differences in Experiment 1, and Kuijpers and Donselaar (1998) found that speakers realized epenthesis more frequently in words with /r/ clusters than in words with /l/ clusters. RT differences between realizations with and without epenthesis were in fact larger for stimuli with /r/ clusters (887 vs. 957 ms) than for words with /l/ clusters (924 vs. 946 ms), but analyses including this factor showed no main effect of the type of cluster (l/r) and only a marginally significant interaction between epenthesis and type of cluster in the analysis by subjects [$F_1(1,55) = 3.67$, $p = .06$].

In fact Experiment 2a, which tested several different kinds of items, necessarily contained only a restricted number of each kind, and the power of the items analyses was thus very limited. In Experiment 2b, therefore, we attempted to replicate the most surprising aspect of the present results, namely, the RT advantage for words with epenthetic schwa. Despite the fact that, as Experiment 1 indicated, listeners from the population in question consider the forms without epenthetic schwa to be more canonical, they find the forms with epenthetic schwa easier to process. Experiment 2b investigated only this factor of the presence versus the absence of epenthetic schwa in words which allow epenthesis, but investigated it with three times as many items of this type as in Experiment 2a; further, we used the yes/no form of the lexical decision task.

EXPERIMENT 2b

Method

Materials. Thirty-six monosyllabic Dutch words which allowed epenthesis were chosen;

they are listed in the Appendix. There were 12 items with /l/ cluster and 24 with /r/ cluster; 34 of the 36 did not contain embedded words. A further 70 real words were chosen to serve as filler, practice, and warmup items, and 106 pseudo-words were constructed. Thirty-six of these pseudo-words were patterned to resemble the experimental word set and were realized both with and without epenthesis. Many of the remaining 70 words and 70 pseudo-words also contained consonant clusters. There were in this case no items of the types *hals*, *haləs*, *bilk*, *billijk*.

All items were recorded by a male native speaker of Dutch; the experimental words and nonwords were again realized both with and without epenthesis. The items were digitized and measured as in Experiment 2a. Again the word realizations with epenthesis were a little shorter (mean 557 ms) than the realizations without epenthesis (mean 570 ms), though the difference did not reach significance. The nonwords with epenthesis were significantly shorter (mean 554 ms) than the same nonwords realized without epenthesis [mean 598 ms; $t(35) = 3.01$, $p < .01$]. The materials were again arranged in two sets, with the presence of epenthesis counterbalanced.

Procedure and participants. The procedure was as in Experiment 2a, except that listeners had two response buttons and were instructed to press the button labeled YES for real-word items and the button labeled NO for nonwords. Twenty-four native Dutch-speaking students from the same population, none of whom had participated in the earlier experiments, took part, again for a small payment; 12 heard each materials set.

Results and Discussion

The RTs for correct YES responses were again subjected to separate analyses of variance across participants and items. The mean RT to words with epenthesis was 829 ms and to the same words without epenthesis 862 ms. This difference was significant: $F_1(1,23) = 9.13$, $p < .01$; $F_2(1,35) = 5.51$, $p < .03$. The correlation between RT and word duration was again weak but significant ($r = .16$, $p < .01$). A further

analysis was carried out in which RT was co-varied with word duration; the RT difference remained significant: $F_1(1,22) = 7.45$, $p < .02$; $F_2(1,34) = 5.88$, $p < .03$. The error rate was also lower (12%) for words with epenthesis than for words without epenthesis (15%), though this difference was not statistically significant. We conclude that the processing advantage for words with vowel epenthesis is robust and replicable.

The mean RT of NO responses to the nonwords with epenthesis was 889 ms, without epenthesis 928 ms, which was also significant: $F_1(1,23) = 16.79$, $p < .001$; $F_2(1,35) = 13.37$, $p < .001$. However, this effect was not maintained in the analysis of covariance with item duration: $F_1 < 1$, $F_2(1,34) = 4.14$, $p = .05$. The error rate was 2% for each form. The presence of an epenthetic vowel therefore does not enable pseudo-words to be rejected more rapidly or accurately. It is clear, however, that it does confer an advantage on the recognition of real words.

Interestingly, this result is not an indication of a general processing advantage for noncanonical forms, since Kuijpers, Donselaar, and Cutler (1996) report the opposite result for words with optional schwa deletion; listeners were faster at recognizing the standard form *tafereel* (in which the medial syllable contains schwa) than the phonological variant *taf'reel* (with the medial vowel deleted). The same result was observed by Racine and Grosjean (1997) for French words such as *semaine* spoken as *s'maine*. One possible explanation of our present finding would therefore simply invoke facilitation of processing at the phonetic level: alternating consonant-vowel sequences (as in *balək*, *tafə'reel*) may be in principle easier to process than consonant sequences (as in *balk*, *tafreel*). Acoustic characteristics of consonants vary as a function of their position in the syllable: Consonants in a coda show stronger coarticulation with vowels than consonants in onsets, with liquids being particularly sensitive to coarticulation (Rietveld & Frauenfelder, 1987). Dutch liquids in a coda (as in *balk*) are easily vocalized (Reenen, 1987) and may thus be less easy to perceive than liquids in onset/ambisyl-

labic position (as in *balək*); in the latter case, the liquid is followed by a vowel, which is known to facilitate consonant perception (Liberman, Delattre, Cooper, & Gerstman, 1954; van Son & Pols, 1995). Such acoustic advantages for consonant–vowel sequences have been referred to in explaining effects in phoneme detection, for instance, slower detection of segments within clusters in English (Treiman, Salasoo, Slowiaczek, & Pisoni, 1982), slower detection of word-initial consonants in French when a following vowel is elided (Matter, 1986), and faster detection of consonants in onsets than in codas in Dutch (Vroomen & De Gelder, 1999).

However, there is also another possible explanation, which involves the possible role of epenthetic forms in normal speech processing. For words pronounced in isolation, as in a lexical decision experiment, word beginnings and endings are quite clear. However, words usually occur in connected speech, and beginnings and endings are then much harder to establish. Especially for sequences of consonants, it may be difficult to determine whether all consonants belong (as a coda cluster) to the same word or form part of different words. In Dutch (and English), for instance, an /rk/ sequence can either be a consonant cluster in a word coda or, alternatively, /r/ and /k/ can belong to different words. It may be that schwa epenthesis in clusters provides listeners with information relevant to this segmentation process.

Indeed, an epenthetic schwa can mark a sequence of consonants as a coda cluster. Although epenthesis can occur across syllable boundaries (as in *werəken* “to work”) speakers in Kuijpers and Donselaar’s (1998) production experiments rarely epenthesized across syllable boundaries. An epenthetic schwa may therefore mark a sequence of consonants as a coda cluster. It is thus possible that epenthesis helps processing of words in context even more than in isolation. For example, the string -V1rkV2- should, according to phonological rules (Maximum Onset Principle) more likely be segmented into -V1r plus kV2- than into -V1rk plus V2-. If the presence of schwa indicates the existence of a final cluster -rk, as in -V1rəkV2-, this schwa could help in correctly segmenting

the input into -V1rk plus V2-. In our next experiment, therefore, we addressed effects of schwa epenthesis in context. We used the word spotting task (see McQueen, 1996, for a review) since this task, which requires listeners to segment words out of nonsense contexts, is specialized for the study of the segmentation of continuous speech.

EXPERIMENT 3

In the word-spotting task, listeners hear a list of nonsense words, some of which contain embedded real words. Their task is to press a button if they hear any real word and then to pronounce the word that they have spotted. For example, if subjects hear the nonsense word *vorkrauk*, they would be expected to detect the real Dutch word *vork* (“fork”). As this example shows, the nonsense words are constructed from a real word plus a minimal context, thus allowing study of the effects of adjacent context on word recognition. In Experiment 3 we investigated whether words like *vork* would be spotted more rapidly when presented in a form with epenthesis than when no epenthesis was applied; that is, we were able to check whether the facilitatory effects of epenthesis which we observed for lexical decisions on isolated words in Experiment 2 would reappear with this recognition-in-context task.

In constructing materials for a word-spotting study, researchers are usually at pains to ensure that each item contains only a single embedded word; listeners’ responses during the experiment are taped and checked to ensure that they in fact spot the intended words. *Vorkrauk* contains no other words than *vork*; *vor*, *krau*, *rauk*, and so on are not words of Dutch.

In Experiment 3, however, we included a subset of items which contained two embedded words; one-third of the items were CVCC words, potentially allowing epenthesis, in which the first CVC was also a real word. An example is *volk* (“people”), which contains the word *vol* (“full”). These items were included to address the issue of whether an epenthetic vowel serves to signal to the listener that the liquid preceding the vowel belongs in a cluster with the consonant following the vowel; if so,

we expect that listeners will respond with the CVC word less often when an epenthetic vowel is present than when there is no epenthesis.

In Experiment 3 we also included a manipulation of the minimal context to address the same issue. In the example *vorkrauk* above, /kʁ/ is a possible syllable onset, so that the nonsense item could be syllabified as *vor* plus *krauk*. In *vorkpeep*, however, the sequence /kp/ could not be a cluster; there must be a syllable boundary between /k/ and /p/. It is known that word spotting is facilitated by clear boundaries adjacent to the embedded word (McQueen, 1998); therefore, we expect that *vork* will be easier to spot in *vorkpeep* than in *vorkrauk*. If, however, the presence of epenthetic schwa indicates a final cluster, then the effect of the boundary ambiguity may be weakened or nullified. In other words, we would in this case observe an interaction between the epenthesis and the boundary factors, such that RT to spot *vork* would be more different in *vorkrauk* versus *vorkpeep* than in *vorəkrauk* versus *vorəkpeep*.

Finally, since this boundary clarity comparison was applied both to the unambiguous (*vork*) and to the ambiguous embedded words (*vol/volk*), we might expect that the tendency to choose a CVC (*vol*) over a CVCC (*volk*) word would be greater with an unclear boundary (*volkroop*) than with a clear boundary (*volkmoop*), but that this difference also would be attenuated when the CVCC word was spoken with an epenthetic vowel.

Method

Materials. Twenty-four Dutch nouns like *vork* were selected—CVCC in structure, with the final cluster subject to epenthesis, and containing no further embedded words. Twelve further nouns like *volk* were selected—CVCC in structure, with the final cluster subject to epenthesis, and containing an embedded CVC word. Two contexts were constructed to follow each of the 36 stimulus words; one context began with a consonant which could form an onset cluster with the final consonant of the CVCC word (unclear boundary context); the other context began with a consonant which could not form the second part of an onset cluster (clear

boundary context). All items, and their contexts, are listed in the Appendix.

Sixty-four further nonsense items were constructed for use as fillers, practice, and warmup items. Of these, 48 were similar in structure to the experimental stimuli but contained no embedded words: 16 CVC(ə)CCVC pseudo-words with the first syllable ending with a consonant cluster containing /l/ or /t/ and 32 CVC(ə)CCVC pseudo-words which did not contain /l/ or /t/. The remaining 16 were CVC(ə)C real words without /l/ or /t/ in coda position, realized again with a CVC following context.

The material was arranged in four lists. The 36 experimental items were realized either with or without an epenthetic vowel, with epenthesis (present/absent) and boundary (clear/unclear) counterbalanced across lists. All material was read by a phonetically trained female native speaker of Dutch, in a single session, and recorded on digital audiotape. The words were digitized with a sampling frequency of 20 kHz using Entropics X-Waves software. The duration of the CVC and CVCC portions of all experimental stimuli was measured. The unambiguous words (*vork*) did not significantly differ in duration when read with epenthesis (mean duration 406 ms) versus without (405 ms), but the CVC portion was shorter in the realizations with epenthesis (210 ms) than in the realizations without (282 ms). The same was true for the ambiguous words (*volk*), which averaged 389 ms with epenthesis and 394 ms without, while the CVC portions averaged 205 ms with epenthesis and 266 ms without.

Procedure. Participants were tested individually in sound-attenuated booths. They heard the material, beginning with a 16-item practice set, over Sennheiser headphones. They were instructed to listen carefully to each item, to press a button as quickly as possible if the item contained a real word, and then to speak the word that they had spotted (or the one that first came to mind if there were more than one alternative). Vocal responses were recorded on a Sony DTC 55 ES DAT recorder. Responses were timed from a signal, inaudible to the subjects, aligned with item onset. Stimulus presentation and response timing and storage were

under control of a personal computer running the NESU experimental control software.

Participants. Forty-eight native speakers of Dutch, students at the University of Nijmegen, took part in the 15-min experiment in return for a small payment. None had taken part in Experiments 1 or 2. Twelve listeners heard each of the experimental lists.

Results

Subjects' oral responses were checked for their correctness and coded as CVC or CVCC responses; responses which did not correspond to a word actually embedded in the stimulus item were removed from the data set, as were responses made more than 2500 ms after onset of the target word. RTs were adjusted for the measured word durations to give response times from embedded word offset and were then subjected to separate analyses of variance across subjects (F_1) and items (F_2). Results for the unambiguous (*vork*) and ambiguous (*vol/volk*) items were analyzed separately.

Unambiguous items (vork). The overall detection rate for these items was 90%. Mean RTs and miss rates per condition are shown in Table 3. Analyses of variance revealed a significant effect of Epenthesis, such that embedded words realised with epenthesis (*vorək*) were detected more rapidly (mean RT, 520 ms) than the same words realized without epenthesis (*vork*; mean RT, 615 ms): $F_1(1,44) = 31.08, p < .001, F_2(1,91) = 9.7, p < .005$. There was no effect of boundary clarity in the RTs and no interaction between these two factors. Analyses of the miss rates showed no effect of epenthesis and

TABLE 3

Word Spotting: Mean RT (ms) from Word Offset (and Mean Miss Rates in Parentheses) for Unambiguous Items, Experiment 3

| | With epenthesis | Without epenthesis |
|---|-----------------|--------------------|
| Clear boundary (e.g., <i>vor[ə]kpeep</i>) | 516 (7.1%) | 632 (6.9%) |
| Unclear boundary (e.g., <i>vor[ə]krauk</i>) | 525 (10.5%) | 598 (14.2%) |

TABLE 4

Word Spotting: Mean RT (ms) from Word Offset and Percentage of Each Type of Response for Ambiguous Items, Experiment 3

| | With epenthesis | | Without epenthesis | |
|---|-----------------|-------|--------------------|-------|
| Clear boundary (e.g., <i>vol[ə]kmoop</i>) | | | | |
| CVC | 796 | 8.3% | 647 | 4.4% |
| CVCC | 648 | 75% | 629 | 93% |
| No response | | 16.7% | | 2.7% |
| Unclear boundary (e.g., <i>vol[ə]kroop</i>) | | | | |
| CVC | 783 | 26.4% | 676 | 31.9% |
| CVCC | 672 | 59.7% | 658 | 47.2% |
| No response | | 13.9% | | 20.9% |

only a small effect of boundary clarity, significant in the subjects analysis only—fewer errors were made when the boundary between word and following context was clear (7%) than when it was unclear [12.4%; $F_1(1,44) = 14.8, p < .001; F_2(1,91) = 1.92, p > .1$].

Ambiguous items (volk). The overall detection rate for the ambiguous items was 86.5%. On 68.75% of trials a CVCC response was made and on 17.75% of trials a CVC response. Mean rates of each type of response, miss rates, and RTs are given in Table 4.

There was no effect of epenthesis on the overall rate of CVC versus CVCC responses; CVC responses were made on 17.4% of trials with epenthesis and 18% of trials without. As predicted, fewer CVC responses were made when the boundary between the CVCC word and the following context was clear (6.35%) than when that boundary was unclear (29.2%): $\chi^2(1) = 59.68, p < .001$. However, the interaction of this effect with epenthesis was not significant.

RTs for CVC and CVCC responses were analysed separately. No effects reached significance. As with the unambiguous items, fewer missed responses occurred in items with clear boundaries (9.7%) than in items with unclear boundaries (17.4%), but again this effect was significant only in the subjects analysis [$F_1(1,44) = 5.48, p < .025; F_2(1,43) = 1.73, p > .1$].

Discussion

The main finding of Experiment 3 is a clear replication of the facilitatory effect observed in Experiment 2: words with final clusters are easier to perceive when they are realized with an epenthetic vowel. *Vork* was perceived significantly more rapidly in *vorəkrauk* and *vorəkpeep* than in *vorkrauk* and *vorkpeep*. The further results provided no evidence for a cohesive effect of epenthesis in signalling the presence of a cluster. Our boundary clarity manipulation did not itself result in a significant effect either on RTs or miss rates, and there was no sign of the expected interaction between epenthesis and boundary in which effects of boundary clarity would be attenuated for items with epenthesis. There was also no sign that epenthesis led to ambiguous items (*vol/volk*) being more likely to be interpreted as CVCC or that epenthesis interacted with boundary clarity in determining the type of response to these items.

We therefore reject the hypothesis that epenthesis signals cluster cohesion. The alternative hypothesis outlined above interprets the facilitatory effect of epenthesis on word recognition in terms of simple phonetic enhancement: specifically, the liquid in the final cluster is easier to identify when it is followed by an epenthetic vowel than when it is followed directly by the consonant with which it forms a cluster, and in consequence recognition of the cluster, and by extension of the word as a whole, is speeded.

In the next and final experiment we test this latter hypothesis, via a task in which listeners' recognition of the liquid consonants in words like *tulp* and *vork* is directly assessed. Phoneme detection (see Connine & Titone, 1996, for a review) measures listeners' RT to detect specified phoneme targets; we use this task to measure how quickly listeners can detect /l/ in *tuləp* versus *tulp* and /r/ in *vorək* versus *vork*.

EXPERIMENT 4

Method

Materials. The materials were the 36 experimental items with unambiguous boundaries used in Experiment 3, plus 1 further item (*durfpuut*) which had been used in a pilot experiment

for Experiment 3 and had been recorded at the same time. The items with ambiguous boundaries were not used because many of them contained one of the target phonemes /l, r/ in the context. All items occurred in two forms: with and without epenthesis, making a total of 74 test items. These were divided into two sets of 37, sets A and B, each set containing one version of each item pair, with the presence versus absence of epenthesis approximately balanced across sets.

Each set was augmented with 58 further items, also taken from Experiment 3, which served as warmup and filler items. Some of these contained occurrences of the specified target in other positions than that (immediately following the first vowel) in which the target always occurred in test items. Some filler items contained embedded words and others did not. The fillers were chosen such that the numbers of YES and NO responses, and the numbers of L and R targets, were, as nearly as possible, equal in the experiment as a whole and in sets A and B. Some fillers occurred in both sets A and B, on occasion with different targets. The first 3 items in each set were warmup items.

The items were measured and the duration of the portion between item onset and onset of the target phoneme ascertained. The onset of the /r/ or /l/ was determined by a combination of visual and auditory examination. The pretarget portion (the initial CV, e.g., of the *tu-* in *tulp*) differed significantly in duration across the sets with epenthesis (155.8 ms) and without [180.6 ms; $t(36) = 4.11, p < .001$].

Participants. Twenty-four Nijmegen University students, with normal hearing, took part in the experiment and were paid for their participation. All heard both sets A and B, without a pause between sets; 12 heard the sets in A-B and 12 in B-A order.

Procedure. Subjects were tested individually or in groups of up to three; they were seated in sound-attenuated booths containing a VDU screen and two response buttons labeled YES and NO. The experiment began with a seven-item practice session. The spoken items were presented over headphones at a rate of one item every 3600 ms. Prior to each spoken item, a

target specification—L or R—appeared on the screen. The subjects decided whether the sound represented by the letter on the screen was present or not in the spoken item and signaled their response by pressing the appropriate response key. Response timing was initiated by a signal, inaudible to the subjects, aligned with the onset of each item and was terminated by a key press response. Stimulus presentation and response timing and storage were controlled by a microcomputer running NESU experimental control software.

Results and Discussion

The error rate was low (3.7%) and errors were not further analyzed. Response times were adjusted by the measured duration between item onset and target onset to give responses from the latter point. Mean response times for each subject and each item were computed and separate analyses of variance across subjects and across items calculated. The mean detection time for the phoneme targets in the items with epenthetic vowel was 699 ms, and the mean detection time in items without epenthesis 731 ms. This difference was significant [$F_1(1,23) = 8.78, p < .005$; $F_2(1,35) = 5.29, p < .03$]. There was no significant difference in response latency to L versus R targets, and although the epenthesis effect was in fact larger for R targets than for L targets, the interaction between the epenthesis and L/R factors did not reach significance.

Thus the detection of a liquid consonant was significantly easier when it was followed by an epenthetic vowel than when it was followed by the consonant with which it nominally formed a cluster. (Note that, as reported for Experiment 3, the overall duration of the words with and without epenthesis was virtually identical. The durations of the initial CV were, however, significantly longer for items without epenthesis than for items with epenthesis. As a consequence, measuring response time as we did, i.e., from target phoneme onset, provides the most conservative measure of the difference between the two conditions. Responses measured from item onset, word offset, or item offset show a

significantly larger advantage for the phonemes followed by epenthetic vowels.)

GENERAL DISCUSSION

In four experiments we have investigated how an optional phonological adjustment, epenthesis of a vowel between the consonants of a coda cluster, affects the recognition of spoken Dutch words. Our results indicate that this alteration of the canonical form of words, far from adversely affecting word recognition, actually facilitates it. In speeded-response tasks—lexical decision and word-spotting—listeners' word-recognition responses were consistently faster when words like *tulp*, *film*, and *werk* were presented in their variant form with an epenthetic vowel than when they were presented in their canonical form without it.

As we pointed out in the Introduction, one simple way in which the language processor could avoid the problem of optional word-form variation would be to have separate representations of the two possible forms. However, our first experiment strongly suggested that the two forms of words like *tulp*, *werk*, and *film* are not represented separately. In this study with the word-reversal task, listeners treated the bisyllabic forms with epenthesis exactly as if they were the monosyllabic forms without. Thus the two forms did not activate alternative access representations; they both activated the same representation, and that representation appeared to be monosyllabic. Thus it appears that the canonical representation of words like *tulp* is the monosyllabic form. The results from the two speeded-response tasks therefore show that the form which in this respect is clearly the canonical one is NOT the form which is easiest to process: lexical activation by the noncanonical form, the form with epenthesis, is actually quicker. This, we argued, is simply because epenthesis acts to enhance perceptibility at the phonetic level: specifically, the liquid in a liquid-obstruent cluster is, as our fourth experiment showed, easier to detect when it is followed by an epenthetic vowel.

No previous studies of the effects of phonological adjustments on spoken-word recognition have addressed exactly the case with which we

have been concerned in our studies of epenthesis. Studies of assimilation phenomena have investigated the effects of violating obligatory sequential constraints (Gaskell & Marslen-Wilson, 1996, 1998; Kuijpers & Donselaar, forthcoming; Otake et al., 1996), or the effects of optional assimilations which occur across word or morpheme boundaries, as in *hot cakes* or *kaasboer* “cheesemonger” (Gaskell & Marslen-Wilson, 1996, 1998; Koster, 1987; Kuijpers & Donselaar, forthcoming; Marslen-Wilson, Nix, & Gaskell, 1995), or the effects of the presence of assimilation in foreign language input (Koster, 1987; Otake et al., 1996). As we described in the Introduction, this research has revealed that processing is hindered by violations of obligatory constraints, but is neither facilitated nor hindered by the application of a rule which produces a phonological variant; *hot cakes* is understood no more nor less efficiently whether pronounced in its canonical form or in its variant form with /t/ replaced by /k/.

One study of a related effect revealed, indeed, an indirect facilitatory effect of an optional variant. Lahiri, Jongman, and Sereno (1990) compared the recognition of spoken Dutch words like *kies* (the imperative form of *kiezen*, “choose”) preceded by a priming phrase *ik kies haar* “I choose her”; this latter form can be spoken in optional variants with the final consonant of the verb voiced or unvoiced. The voiced form in fact is an assimilation which overrides and thus violates the otherwise general rule of syllable-final devoicing in Dutch; it is also not the citation form of the first person form of the verb. Nonetheless, Lahiri et al. found that the imperative target was more strongly facilitated by the prime with the voiced consonant. In this form, the prime more closely corresponds to the infinitive form of the verb, which has a medial voiced consonant; Lahiri et al. argued that this variant was closer to the underlying phonological form and thus more effectively contacted the relevant lexical representation.

The present study, however, is the first to examine the effect of an optional within-word phonological adjustment on the recognition of the words containing that adjustment, spoken in

isolation, and to demonstrate that such phonological variation can have a facilitatory effect on processing of the word itself. Vowel epenthesis in Dutch significantly speeds the recognition of the words in which it has applied.

We do not believe that this result suggests that word processing in Dutch differs from word processing in other languages. The findings regarding assimilation phenomena run exactly parallel whether the materials, and the listeners, are Dutch (Koster, 1987, Kuijpers & Donselaar, forthcoming), English (Gaskell & Marslen-Wilson, 1993, 1996), or Japanese (Otake et al., 1996). Nor do we believe that an explanation of the finding should be sought in the fact that epenthesis applies within rather than across words. In an earlier study (Kuijpers et al., 1996) we found preliminary indications that word recognition may be rendered somewhat more difficult by optional vowel deletion (*taf'reel* for *tafereel*, in Dutch; analogous cases exist in many other languages, such as English *fam'ly* for *family* or French *gal'rie* for *galerie*, and similar inhibitory effects have been reported for French: Racine & Grosjean, 1997).

Instead, we suggest that the facilitatory effects are specific to vowel epenthesis. As we pointed out in the Introduction, it is frequently claimed that epenthesis leads to an increase in articulatory ease. This claim is buttressed by the strong preference across languages for alternating consonants and vowels in word forms and by the finding (Kuijpers & Donselaar, 1998) that epenthesis is more likely if it makes the utterance as a whole more rhythmically regular. However, as pointed out above, it does not explain why there is no obvious bias against clusters in other aspects of the phonology—e.g., why Dutch allows the formation of nicknames which include clusters where the canonical form had none. Consider, however, that speakers are motivated to insert epenthetic vowels largely by the very perceptual facilitation which we have discovered; that is, they do it for the listener. Vowel epenthesis would join a range of such speech production effects in which communication is facilitated by the speaker taking the listener's interests into account (Cutler, 1987; Lindblom, 1988), and the

contradictions raised by an explanation solely in terms of articulatory ease are no longer so significant.

There may of course be more than one reason for listeners' greater ease in processing words like *werk*, *tulp*, and *film* with epenthesis rather than without. However, our results have cast doubt on some possible explanations. Thus we do not believe that durational differences between realizations of words with and without epenthesis have significant effects. Although adding an epenthetic vowel lengthens a word by one segment, it does not necessarily increase its acoustic duration, and indeed, our lexical decision items in Experiments 2a and 2b were slightly shorter with the epenthetic vowel than without, while the word-spotting items in Experiment 3 showed no overall difference between the two forms. Correlations between response time and overall word duration and CVC duration were, however, low and/or insignificant.

Likewise, the results of the word-spotting study argued against a role for epenthesis in signaling an underlying cluster. This result is perhaps not so surprising, given that, as we pointed out earlier, a large proportion of CVCC words in Dutch contain other words, with CVC structure, embedded within them: *vol* "full" in *volk* "people," *hal* "hall" in *half* "half," and so on. Epenthesis in these words cannot remove the competition from the embedded words, since the epenthetic schwa is effectively indistinguishable from a schwa representing an adjectival inflection -e (*volle*), or a noun plural inflection -en (*hallen*), or a verb infinitive ending -en; all are pronounced as schwa. Only in the case of an embedded word which cannot be inflected could an epenthetic schwa have a possible effect of removing competition.

We should note that effects of competitor availability may in fact be observable in our own results. Specifically, interword competition (known to influence word-spotting performance: McQueen, Norris, & Cutler, 1994) could have caused the absence of a facilitatory effect of epenthesis in the ambiguous word-spotting items. Although there is competition in *volkmoop*, such that *vol* competes with *volk*,

there is even stronger competition in *volkmoop*, in which both *vol* and *volle* compete with *volk*, and this increased competition effect acts against the epenthesis effect. Note that this explanation does not require that inflected forms like *volle* have a lexical representation independent of that of their base forms such as *vol*; this may indeed be the case, but the effective competition by *volle* may also simply arise because the added schwa makes *vol* remain a viable competitor for longer.

We believe, however, that the principal reason for the facilitatory effects that we have observed is the simple one of increased phonetic processing ease. We have shown that /r/ and /l/ are easier to perceive when followed by a vowel than when followed by a consonant. Preceding a vowel they are effectively in syllable onset/ambisyllabic position and are hence realized differently from the way they are spoken in a syllable coda (Reenen, 1987). Despite the considerable evidence that listeners can derive information about a coda consonant from a preceding vowel (Whalen, 1984, 1991; Lahiri & Marslen-Wilson, 1991; Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, in press), there is, as we pointed out earlier, evidence that consonants in a singleton onset are easier to identify than the same consonants in a cluster or in a coda. This asymmetry could also well underlie the widespread preference, across languages, for consonant-vowel alternation and the predominance of CV syllables as the most common syllable type across languages.

In the Introduction we noted that, from the speakers' point of view, schwa epenthesis may not arise via insertion of a segment as such, but simply via relaxation of the gestures corresponding to articulation of the consonant cluster. The results of Experiment 1, in which words with epenthesis were treated as monosyllabic, are in accord with such a unitary approach to both realizations of words such as *tulp*. And in this, Dutch schwa epenthesis is of course not unique. Ladefoged (in press), for instance, discusses a similar case in Scottish Gaelic, which contains pairs of words such as *balg* and *ballag*, both of which are pronounced [palak]; despite this, differences in intonation patterns applied to

the two words show that the former word (which, as its spelling suggests, arose from an earlier monosyllabic form) is considered by speakers to be monosyllabic, while the latter is held to be bisyllabic. A bisyllabic surface form can realize a monosyllabic underlying form, just as vowel deletion can produce a monosyllabic surface realization of an underlying bisyllabic form.

Our results, concerning as they do listeners' perceptions of the alternative forms, do not directly speak to the viability of the gestural account of how epenthesis arises. That issue is more properly addressed via phonetic studies of the realization of segments in clusters with and without schwa epenthesis. As we have pointed out, the realization of /r/ and /l/ in Dutch (as in many other languages) differs in prevocalic versus postvocalic position. For example, just as English has "clear /l/" in prevocalic position but "dark /l/" in postvocalic position (Gimson, 1970), so in Dutch may postvocalic /l/, but not prevocalic /l/, become vocalized (Reenen, 1987). If /l/ which is prevocalic by virtue of epenthesis is similar to other prevocalic /l/'s, this implies that schwa epenthesis may actually change the nature of the /l/ articulation rather than simply change its timing, which may pose a challenge to the gestural account as described above.

From the listener's point of view, however, *tuləp* contains a vowel between the consonants of the coda. Yet *tuləp* still remains *tulp* and indeed is easier to recognize as *tulp* than *tulp* itself. As we have argued, this facilitatory effect arises in large part because the perception of otherwise relatively difficult liquid consonants is eased. Certainly vowel epenthesis in Dutch constitutes a noteworthy case in that the variant form of a word, although it makes contact with the lexicon via the same route as the canonical form, is easier to process than that canonical form of which it is in essence a transformation.

APPENDIX

A. Stimuli for Experiment 1 (Reversal)

Words. merg ("marrow") pulp ("pulp") tulp ("tulip") galg ("gallows") kerk ("church") turk ("Turk") merk ("mark") kalk ("chalk") spalk ("splint") tolk ("interpreter")

kelk ("calyx") nerf ("grain") korf ("basket") larf ("larva") kalf ("calf") golf ("wave").

Pseudo-words. nerg palp nulp kalg lerk gork terk ralk stilk nolc selk perf porf marf nalf pulf.

B. Stimuli for Experiment 2a (Lexical Decision)

Real word, epenthesis possible. velg ("rim") helm ("helmet") volk ("people") verf ("paint") norm ("norm") pulp ("pulp") balk ("beam") walm ("smoke") galg ("gallow") dorp ("village") perk ("bed") tolk ("interpreter")

Real word, epenthesis prohibited. veld ("field") held ("hero") volt ("volt") vers ("fresh") nors ("surly") puls ("pulse") balts ("display") wals ("waltz") Hals ("neck") dorst ("thirst") pers ("press") pols ("wrist")

Real word, syllabic schwa. billijk ("fair") grillig ("whimsical") knullig ("awkward") lollig ("jolly") mollig ("plump") stellig ("definite") sullig ("soft") wollig ("woolly") knorrig ("grumpy") narrig ("peevisish") warrig ("muddled") prullig ("trashy")

Pseudo-word, epenthesis possible. kelg melm golc zerf borm fulp ralk nalm jalg sorp nerck zolk

C. Word stimuli for Experiment 2b (Lexical Decision)

berg ("mountain") berm ("verge") durf ("nerve") film ("film") golf ("wave") harp ("harp") helm ("helmet") hulp ("help") jurk ("dress") kalf ("calf") kalk ("chalk") kalm ("calm") kelk ("calyx") kerm ("moan") korf ("basket") larf ("larva") melk ("milk") merg ("marrow") merk ("brand") murw ("soft") nerf ("grain") park ("park") scherm ("screen") scherp ("sharp") schulp ("shell") tulp ("tulip") turf ("peat") vlerk ("wing") vorm ("form") warm ("warm") welp ("whelp") werp ("chuck") worm ("worm") wurg ("strangle") zorg ("care") zwerm ("swarm")

D. Stimuli for Experiments 3 and 4 (Word-spotting, Phoneme Detection)

| Unambiguous words | Context (unclear/clear boundary) |
|-------------------|-------------------------------------|
| worp ("throw") | luup/toos |
| zorg ("care") | laus/feup |
| hulp ("help") | roop/muik |
| vork ("fork") | raut/peep |
| harp ("harp") | loet/giem |
| kerk ("church") | luig/feen |
| merk ("brand") | rien/taag |
| berk ("birch") | roek/goep |
| scherf ("shard") | reum/maaf |
| kalk ("chalk") | luip/teif |
| golf ("wave") | roon/saum |
| hark ("rake") | laut/teif |
| larf ("larva") | ruip/meef |
| merg ("marrow") | laam/faus |
| nerf ("grain") | luip/keeg |
| kelk ("calyx") | ruuf/taap |
| melk ("milk") | rauf/geep |

| Unambiguous words | | Context (unclear/clear boundary) |
|-------------------|---------------------|-------------------------------------|
| berg | ("mountain") | laap/taap |
| tulp | ("tulip") | leug/foos |
| jurk | ("dress") | ruuf/teuf |
| park | ("park") | reem/tuup |
| kwark | ("curd") | loem/goep |
| werf | ("shipyard") | roop/kief |
| smurf | ("smurf") | loem/pee |
| Ambiguous words | | |
| vol-volk | "full"-"people" | raut/moop |
| bel-belg | "bell"-"Belgian" | leig/teeg |
| pul-pulp | "tankard"-"pulp" | roop/fiet |
| dor-dorp | "dry"-"village" | luup/kief |
| tol-tolk | "top"-"interpreter" | ruuk/peug |
| wil-wilg | "will"-"willow" | liek/fees |
| hal-half | "hall"-"half" | rien/suup |
| dol-dolk | "mad"-"dagger" | loes/taum |
| kol-kolk | "collar"-"eddy" | reug/puis |
| wol-wolf | "wool"-"wolf" | loet/koef |
| wol-wolk | "wool"-"cloud" | reem/taaf |
| vel-velg | "skin"-"rim" | laug/peim |

REFERENCES

- Baayen, R. H., Piepenbrock, R., & Rijn, H. van (1993). *The CELEX lexical database (CD-ROM)*. Philadelphia: Linguistic Data Consortium, Univ. of Pennsylvania.
- Bell, A., & Hooper, J. B. (1978). *Syllables and segments*. Amsterdam: North-Holland.
- Booij, G. (1995). *The phonology of Dutch*. Oxford: Clarendon Press.
- Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 341-376). Cambridge: Cambridge Univ. Press.
- Browman, C. P., & Goldstein, L. (1992). "Targetless" schwa: An articulatory analysis. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 26-56). Cambridge: Univ. Press, Cambridge.
- Connine, C. M., & Titone, D. (1996). Phoneme monitoring. *Language and Cognitive Processes*, **11**, 635-645.
- Cutler, A. (1980). Syllable omission errors and isochrony. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 183-190). The Hague: Mouton.
- Cutler, A. (1987). Speaking for listening. In A. Allport, D. G. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production: Relationships between listening, speaking, reading and writing* (pp. 23-40). London: Academic Press.
- Cutler, A. (1998). *The recognition of spoken words with variable representation* (pp. 83-92). Proceedings of the ESCA Workshop on Speech Sounds of Spontaneous Speech, Aix-en-Provence.
- Donselaar, W. van, Kuijpers, C., & Cutler, A. (1996). *How do Dutch listeners process words with epenthetic schwa?* Proceedings of the Fourth International Conference on Spoken Language Processing (Vol. 1, pp. 149-152). Philadelphia.
- Donselaar, W. van, & Stoutjesdijk, O. (1993). The acoustic/phonetic factors behind the "syllable effect." In M. Everaert, B. Schouten, & W. Zonneveld (Eds.), *OTS Yearbook 1993* (pp. 25-43). Utrecht: LEd.
- Dupoux, E., Kakehi, K., Hirose, K., Pallier, C., & Mehler, J. Epenthetic vowels in Japanese: A perceptual illusion. [submitted for publication]
- Frauenfelder, U. H., & Kearns, R. K. (1996). Sequence monitoring. Spoken word recognition paradigms. *Language and Cognitive Processes*, **11**, 665-673.
- Frauenfelder, U. H., Rietveld, A. C. M., & Til, A. van The phonetic basis of the syllable effect. [submitted for publication]
- Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, **22**, 144-158.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference. *Journal of Experimental Psychology: Human Perception and Performance*, **24**, 380-396.
- Gimson, A. C. (1970). *An introduction to the pronunciation of English*. London: Arnold.
- Goldinger, S. D. (1996). Auditory lexical decision. In F. Grosjean & U. H. Frauenfelder (Eds.), *Spoken word recognition paradigms. Language and Cognitive Processes*, **11**, 559-567.
- Hyman, L. M. (1975). *Phonology. Theory and analysis*. New York: Holt, Rinehart, & Winston.
- Kashino, M., Wieringen, A. van, & Pols, L. (1992). *Cross-language differences in the identification of intervocalic stop consonants by Japanese and Dutch listeners*. Proceedings of the Second International Conference on Spoken Language Processing, (Vol. 2, pp. 1079-1082). Banff.
- Koster, C. J. (1987). *Word recognition in foreign and native language*. Dordrecht: Foris.
- Kuijpers, C., & Donselaar, W. van (1998). The influence of rhythmic context on schwa epenthesis and schwa deletion. *Language and Speech*, **41**, 87-108.
- Kuijpers, C., & Donselaar, W. van. Phonological variation and phoneme identification in Dutch. [forthcoming]
- Kuijpers, C., Donselaar, W. van, & Cutler, A. (1996). *Phonological variation: Epenthesis and deletion of schwa in Dutch*. Proceedings of the Fourth International Conference on Spoken Language Processing, (Vol. 1, pp. 94-97). Philadelphia.
- Ladefoged, P. (in press). Some thoughts on syllables: An old-fashioned interlude. In J. Local & R. Ogden (Eds.),

- Papers in laboratory phonology VI Cambridge: Cambridge University Press.*
- Lahiri, A., Jongman, A., & Sereno, J. (1990). The pronominal clitic [dər] in Dutch: A theoretical and experimental approach. *Yearbook of Morphology*, **3**, 115–127.
- Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, **38**, 245–294.
- Lieberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs*, **68**, 1–13.
- Lindblom, B. (1988). Phonetic invariance and the adaptive nature of speech. In B. A. G. Elsendoorn & H. Bouma (Eds.), *Working models of human perception* (pp. 139–173). London: Academic Press.
- Marslen-Wilson, W. D., Nix, A., & Gaskell, G. (1995). Phonological variation in lexical access: Abstractness, inference and English place assimilation. *Language and Cognitive Processes*, **10**, 285–308.
- Marslen-Wilson, W. D., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, **101**, 653–675.
- Matter, J. F. (1986). *A la recherche des frontières perdues*. Ph.D. thesis, Univ. of Utrecht.
- McQueen, J. M., (1996). Word spotting. *Language and Cognitive Processes*, **11**, 695–699.
- McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, **39**, 21–46.
- McQueen, J. M., Norris, D. G., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 621–638.
- McQueen, J. M., Norris, D. G., & Cutler, A. (in press). Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*.
- Otake, T., Yoneyama, K., Cutler, A., & Lugt, A. van der (1996). The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America*, **100**, 3831–3842.
- Racine, I., & Grosjean, F. (1997). *La reconnaissance des mots en parole continue: Effacement du schwa et frontière lexicale*. Actes des Journées d'Etudes Linguistiques, Nantes.
- Reenen, P. Th. van (1987). The vocalization of /l/ in standard Dutch, a pilot study of an ongoing change. In F. Beukema & A. Hulk (Eds.), *Linguistics in the Netherlands* (pp. 189–199). Amsterdam: Benjamins.
- Rietveld, A. C. M., & Frauenfelder, U. H. (1987). *The effect of syllable structure on vowel duration*. In Proceedings of the 11th International Congress of Phonetic Sciences, (Vol. 4, pp. 28–31), Tallinn.
- Schiller, N. O., Meyer, A. S., & Levelt, W. J. M. (1997). The syllabic structure of spoken words: Evidence from the syllabification of intervocalic consonants. *Language and Speech*, **40**, 103–140.
- Son, R. J. H. van, & Pols, L. C. W. (1995). *The influence of local context on the identification of vowels and consonants*. Proceedings of the 4th European Conference on Speech Communication and Technology, (Vol. 2, pp. 967–970), Madrid.
- Treiman, R., & Danis, C. (1988). Syllabification of intervocalic consonants. *Journal of Memory and Language*, **27**, 87–104.
- Treiman, R., Salasoo, A., Slowiaczek, L., & Pisoni, D. (1982). Effects of syllable structure on adults' phoneme monitoring performance. *Progress Report Indiana University Speech Research Laboratory*, **8**, 63–81.
- Vroomen, J., & Gelder, B. de (1994). *Speech segmentation in Dutch: No role for the syllable*. In Proceedings of International Conference on Speech and Language Processing (Vol. 3, pp. 1135–1138). Yokohama.
- Vroomen, J., & Gelder, B. de (1999). Lexical access of resyllabified words: Evidence from phoneme monitoring. *Memory and Cognition*, **27**, in press.
- Whalen, D. H. (1984). Subcategorical mismatches slow phonetic judgments. *Perception and Psychophysics*, **35**, 49–64.
- Whalen, D. H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception and Psychophysics*, **50**, 351–360.
- Wijnen, F., Krikhaar, E., & Den Os, E. (1994). The (non)-realization of unstressed elements in children's utterances: evidence for a rhythmic constraint. *Journal of Child Language*, **21**, 59–83.
- Zwitsersloot, P., Schriefers, H., Lahiri, A., & Donselaar, W. van (1993). The role of syllables in the perception of spoken Dutch. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **19**, 260–271.

(Received January 5, 1998)

(Revision received December 31, 1998)