Journal of Phonetics (2001) **29,** 53–87
doi:10.006/jpho.2001.0129
Available online at http://www.idealibrary.com

Perception of epenthetic stops

Natasha Warner

Max Planck Institute for Psycholinguistics, PB 310, NL-6500 AH Nijmegen, The Netherlands and Department of Linguistics, University of Arizona, P.O. Box 210028, Tucson, AZ 85721-0028, U.S.A.

Andrea Weber

Max Planck Institute for Psycholinguistics, PB 310, NL-6500 AH Nijmegen, The Netherlands

In processing connected speech, listeners must parse a highly variable signal. We investigate processing of a particular type of production variability, namely epenthetic stops between nasals and obstruents. Using a phoneme monitoring task and a dictation task, we test listeners' perception of epenthetic stops (which are not part of the string of segments intended by the speaker). We confirm that the epenthetic stop perceived is the one predicted by articulatory accounts of how such stops are produced, and that the likelihood of an epenthetic stop being perceived as a real stop is related to the strength of acoustic cues in the signal. We show that the probability of listeners mis-parsing epenthetic stops as real is influenced by language-specific syllable structure constraints, and depends on processing demands. We further show, through reaction time data, that even when epenthetic stops are perceived, they impose a greater processing load than stops which were intended by the speaker. These results show that processing of phonetic variability is affected by several factors, including language-specific phonology, even though the mis-timing of articulations that creates epenthetic stops is universally possible.

1. Introduction

The acoustic signal a listener must parse when perceiving speech diverges in many ways from the canonical form of the words the speaker intends. Not only does the form of a word vary depending on its environment, and vary across speakers, but also, a given word is often produced in different ways even in the same environment and by the same speaker. Particularly in fast connected speech, there are many phonological or phonetic changes which do not always apply, such as some types of assimilation, lenition, epenthesis, etc. When these changes do apply, they make a word less similar to its canonical form. Listeners must recognize words despite such variability. In this paper, we investigate listeners' processing of a particular type of phonetic variability, epenthetic stops between a nasal and an obstruent. There has been very little work on the perception of epenthetic stops, but because their production is highly variable, they

E-mail addresses: nwarner@u.arizona.edu; andrea.weber@mpi.nl

0095–4470/01/010053 + 35 $35.00/0

provide an interesting case by means of which to investigate how listeners process natural speech. Furthermore, past research has shown that language-specific phonological factors affect how listeners segment the speech stream to locate word boundaries and how they interpret sequences to which obligatory phonological rules apply. Investigation of epenthetic stop perception allows us to determine whether language-specific phonology also influences the processing of phonetic variability.

Epenthetic stops often appear between nasals and following obstruents, both as synchronic variation (1a) and as a diachronic change[1] (1b) (examples from Clements, 1987; Ohala, 1995).

(1a) [dɹɛmt] ~ [dɹɛmpt] "dreamt"
[timstɚ] ~ [timpstɚ] "teamster"

(1b) "empty" < "œmtig" (Old English to Modern English)
"glimpse" (from a form related to "gleam" plus a suffix "-se")
sumo, sumpsi, sumptus (Latin forms of sumo "to take, choose")

Ohala (1995) provides examples from a variety of additional languages, showing that stop epenthesis is not restricted to a few languages or one language area.

The language we chose for our experiment was Dutch. We are not aware of any cases of historical sound change causing epenthetic stops to become part of the underlying representation in modern Dutch, but epenthetic stops do appear as a matter of synchronic variation in the language (2) (examples from Wetzels, 1985).

(2) [hemt ~ hempt] /hemd/ "shirt"
[haŋt ~ haŋkt] /haŋt/ "hangs"

Wetzels (1985) states that the epenthetic stops in these words were written in some earlier stages of Dutch. We do not know of any quantitative data on the frequency of stop epenthesis in modern Dutch natural speech, but it seems to be quite common. In an informal survey of several native Dutch speakers, we found that epenthetic stops were clearly audible in approximately one-half of productions of the words in (2), as well as in words like those in (3). In Section 3 below, we present quantitative data on the frequency of epenthesis in Dutch nonwords.

(3) [laŋs ~ laŋks] /laŋs/ "along"
[zɛiliŋs ~ zɛiliŋks] /zɛiliŋs/ "sideways"
[ziŋt ~ ziŋkt] /ziŋt/ "sings"

1.1. *The articulatory and historical account of stop epenthesis*

The type of epenthesis we investigate takes place either between a nasal and a non-homorganic voiceless stop or between a nasal and any voiceless fricative with a constriction in the oral cavity. Ohala (1995, 1997) reviews a common articulatory explanation of epenthesis: during production of a nasal, there is a complete closure in the oral cavity, and the velum is lowered, allowing air to escape through the nose. During the production

---

[1] By diachronic or historical change, as it concerns epenthesis, we mean cases in which a sequence which at some previous stage in the language contained a nasal-obstruent cluster, which may have been pronounced variably with and without an epenthetic stop, now has the once epenthetic stop in the underlying representation.

of either a stop or an oral fricative, the velum must be raised. If the velum is raised and voicing ceases before the oral closure for the nasal is released, the oral cavity will be completely sealed, and when the oral closure is released, a voiceless burst will result. (This is the same sequence of events as in a nasal-homorganic stop sequence where the stop is not epenthetic, e.g. /mp/, of course. The difference is in whether the stop is a part of the intended phoneme string, and this causes differences in the timing of the stop.) Cessation of voicing and closure of the velum are assumed to occur nearly simultaneously. If the velum closed before the release of the oral closure, but voicing continued, a voiced epenthetic stop would result instead. Such voiced epenthetic stops do occur occasionally, as in "tens" produced as [tʰɛndz] (Fourakis & Port, 1986), but they are less common than voiceless epenthetic stops.

In a nasal-stop cluster, the nasal and stop must be heterorganic for an epenthetic stop to appear: early closure of the velum in a homorganic sequence would only result in a longer stop closure. However, an epenthetic stop can occur between a homorganic nasal and fricative (as in [prɪnts] for "prince"). Stop epenthesis also occurs in nasal-sonorant and /ls/ clusters, but we will not discuss these types here. This articulatory account predicts that the epenthetic stop must always be homorganic with the preceding nasal, since it results from the release of the oral closure for the nasal.

Ohala (1997) gives stop epenthesis as an example of how sound change develops through listeners mis-parsing the variability in the signal. A speaker, at some point before the epenthetic sound change took place, intended the string "œmtig", but produced [mpt] phonetically through mis-timing of the velic closure. Some listener failed to recover the speaker's intention, and interpreted the string as /mpt/, with a phonemically present /p/. This misinterpretation spread through the language, resulting in the sound change to /ɛmpti/ (which is even represented in orthography). In cases of synchronic variation (e.g., "sense" [nts] ∼ [ns]), there is potential for such sound change, but either listeners have not misinterpreted the epenthetic stops, or the misinterpretation has not (yet) spread through the language. Most speakers probably do not have an underlying /t/ in "sense", even if they often produce a [t] phonetically.

1.2. *Previous studies of epenthetic stops*

Several authors have investigated the implications of stop epenthesis for formal phonological theories, and have put forth formal analyses of how epenthesis occurs (Murray, 1967; Anderson, 1976; Piggott & Singh, 1985; Wetzels, 1985; Clements, 1987; Picard, 1987, 1989). However, there have been only a few phonetic investigations of epenthetic stops, and these focus almost exclusively on production rather than perception. Barnitz (1974), in the first quantitative investigation of epenthetic stops we know of, recorded four English speakers producing real words with nasal-fricative sequences, and listened to the recordings (at normal and slow speed) to decide for each production whether an epenthetic stop had been produced. (He did not compare his perceptions to acoustic evidence.) He finds that production of epenthetic stops is less likely if they would create impossible onset clusters: he claims epenthesis is impossible in "constitution", for example, because [kantstətuʃən] would contain an illegal onset cluster /tst/. (Why the epenthetic /t/ must be assigned to the onset is unclear, and Barnitz's interpretation would apparently rule out the attested epenthesis in "empty" as well.) However, his study leaves multiple factors uncontrolled, and cannot separate influences on production and perception.

Ali, Daniloff & Hammarberg (1979) investigate production of nasal-voiceless fricative sequences in English, measuring oral and nasal airflow and oral pressure. They find that when there is a period of silence in the waveform between the end of voicing for the nasal and the beginning of frication noise (an epenthetic stop), nasal airflow often continues throughout this silent period and even into the following fricative, indicating that the velum does not close at the beginning of the epenthetic stop, as in the explanation in Section 1.1. Instead, in their data it appears that the first change in articulation is the cessation of voicing. Because of the abduction of the vocal cords, airflow into the supraglottal vocal tract increases. The velum remains open, so air vents out the nose, creating what is technically a voiceless nasal. Perhaps the velic opening is not large at this point, or the voiceless nasal produces little acoustic effect, as Ali *et al.* report silence during this portion of the signal. During this same portion of the signal, their data show oral air pressure increasing, even though nasal airflow continues. Perhaps the velic opening is too small for all the air to vent through the nose because of the increased airflow through the abducted vocal cords. After this silent period with continuing nasal airflow and increasing oral air pressure, the velum closes, and the oral closure for the nasal is released. It seems from the data shown by Ali *et al.* that velic closure can either precede or follow the oral release. Because oral air pressure has already been increasing despite velic opening, it is often high enough to produce a burst visible in the waveform, even though there is sometimes still some nasal airflow during the burst itself.

Ohala's (1974) aerodynamic data represent only one token (an /ms/ cluster), but it shows nasal airflow ceasing at the same time as voicing for the nasal ends, as in the more typical articulatory explanation of epenthetic stops. A comparison of the nasal airflow traces and waveforms in Ohala (1974) and Ali *et al.* (1979) shows that the velum closes far earlier relative to the release of the oral closure in Ohala's data. The two papers seem to represent different methods of producing epenthetic stops, in which the primary difference is whether the silent portion of the epenthetic stop corresponds to complete closure of the vocal tract, as in normal stops (Ohala's data), or a voiceless nasal which is too low in amplitude to see clearly in the waveform (Ali *et al.*).

Ali *et al.* (1979) also collected preliminary perception data on the speech from their production study by having the three authors judge during the course of data collection whether an epenthetic stop had been produced. However, since the speakers were necessarily wearing airflow masks, the listening conditions were probably dubious. They find that epenthetic stops are more often produced and perceived before /s, θ, f/ than before /ʃ/. They also find that epenthetic stops are rarely perceived at morpheme or syllable boundaries, even though silent gaps are often produced there, and attribute this to a perceptual tendency to discount silent gaps at boundaries.[2] They note that some words with /ns/ were perceived as having an epenthetic stop, although no silent gap was produced. Furthermore, even tokens with nasals followed by /f/ or /θ/ which were not perceived as having epenthetic stops had acoustic silent periods present.

Fourakis & Port (1986) investigate production of epenthetic stops in words ending in /n/ or /l/ plus a fricative, in two dialects of English (South-African and American Midwestern). They show that epenthesis is not purely a result of universal articulatory constraints on coordination of articulations: in some environments, American speakers epenthesize in every token, while South African speakers never do. They also show that

---

[2] Epenthetic stops in "empty" and other such examples must have been heard at some stage in the development of the language, though.

Americans' productions of words such as [prɪnts] from "prints" and "prince" differ: the [t] closure is significantly longer, and the nasal significantly shorter, when the [t] is underlying than when it is epenthetic. Thus, they argue, epenthesis cannot be simply the result of a phonological rule inserting a segment, since phonetic differences in the segment are maintained. They argue that this is a case of incomplete neutralization.

Blankenship (1992) investigates epenthesis in /ns/ clusters in connected English speech, using the TIMIT database. She compares /ns/ clusters to underlying /nts/ clusters, and finds that some form of stop (silent period, burst, or both) is produced more often in /nts/ clusters than /ns/ clusters, although epenthesis in /ns/ clusters is relatively common (26% of all occurrences). She also finds that epenthesis in /ns/ is infrequent if the following vowel is stressed, and probably more common if /ns/ is within a syllable coda than crossing a syllable boundary, but is possible even if the /ns/ cluster crosses a word boundary. She finds that there is only a non-significant tendency for epenthetic [t] to be shorter than underlying /t/, but this failure to replicate Fourakis & Port's (1986) result could stem from the variability of corpus data.

1.3. *Previous work on processing of phonetic variability*

Many studies on psycholinguistic processing of speech variability involve cases of conditioned and predictable phonological variability, for example resyllabification of a word-final consonant to the onset of the following syllable, or obligatory assimilation of a nasal to the place of a tautosyllabic following stop. Although a given morpheme has more than one form, which form is used is determined by the environment. Effects of such phonologically predictable variation on processing are likely to differ from the effects of optional phonetic variation, or free variation, such as stop epenthesis, where nothing in the environment tells the listener which variant to expect.

Some studies have addressed this latter type of variability (see Cutler (1998) for a review of both types). Word-final nasals in English may, but do not always, assimilate to the place of the initial consonant of the following word. This type of assimilation appears to make recognition of the word containing the nasal neither easier nor harder (Gaskell & Marslen-Wilson, 1996; Marslen-Wilson, Nix & Gaskell, 1998). (Several other authors have addressed related questions (Cutler, 1998).) Furthermore, Kuijpers & Donselaar (forthcoming) find that application or nonapplication of an optional voicing assimilation rule in Dutch consonant clusters (/sb/ appearing as [zb] or [sb]) does not affect listeners' speed in detecting the second consonant of the cluster, which conditions the assimilation.

Donselaar, Kuijpers & Cutler (1999) investigate the perception of Dutch words and segments with and without epenthetic schwa. In Dutch, an epenthetic schwa can break up a consonant cluster in some environments, but this epenthesis is highly variable both across and within speakers, and is never obligatory. Donselaar *et al.* find that listeners can recognize words such as /tylp/ "tulip" more easily with epenthetic schwa ([tyləp]) than without, and can also spot the /l/ more easily. They relate this to acoustic cues for liquids in prevocalic *vs.* in coda position.

Thus, optional or free variation, where the same word surfaces in different forms even in the same environment, either does not affect or facilitates listeners' recognition of the word and the surrounding phonemes. Epenthetic stops are a case of such free variation, but they present a very different situation for the listener from assimilation, or from epenthetic schwa. First, they involve the addition of an acoustic event to the signal, rather than a change in the characteristics of an existing sound, as in assimilation. The question of most interest regarding stop epenthesis is whether listeners perceive the additional acoustic event as a stop, or succeed in "factoring it out" (attributing it to unintended phonetic variation). Secondly, in comparison to schwa epenthesis, stop epenthesis (as synchronic variation) seems to be more purely phonetic: it involves mistiming in the coordination of several articulators, and is less likely to involve the phonological insertion of a segment than vowel epenthesis is.

1.4. *Topics and predictions of the current study*

Previous studies have brought up interesting questions about stop epenthesis. The findings of Ali *et al.* (1979) regarding comparison of production and perception lead one to wonder why in some clusters epenthetic stops are perceived when they are not produced, but in others, not perceived even when they are produced. Furthermore, none of the previous studies has confirmed that the epenthetic stops listeners perceive are the ones predicted by the articulatory accounts of epenthesis, namely epenthetic stops homorganic with the preceding nasal. Barnitz' (1974) conclusion regarding phonotactic legality of the cluster that would result from epenthesis and the findings of Ali *et al.* (1979) regarding morpheme or syllable boundaries are preliminary, but both point toward an influence of the phonology of the language on how epenthetic stops are produced and/or perceived. Since language-specific phonology affects other aspects of speech perception (Cutler, Dahan & Donselaar, 1997; Cutler, 1997), it may also affect how listeners interpret epenthetic stops. The finding of Fourakis & Port (1986) that epenthetic stops are acoustically distinct from underlying ones suggests that listeners might not perceive the two types of stops in the same way.

In this study, we investigate the perception and processing of epenthetic stops in nasal-fricative and nasal-stop clusters. The primary focus of the study is perceptual, but we also present acoustic analyses in order to compare production and perception of the epenthetic stops. We use a wider variety of clusters than previous studies have, and we investigate epenthetic stops only in nonsense words, so that knowledge about how to spell a particular item cannot influence either speakers' production of epenthesis or listeners' perception of it. We investigate how often epenthetic stops are perceived as tokens of the stop phoneme, and whether the epenthetic stop perceived is the one predicted by the articulatory account of their production (a stop homorganic with the nasal). We also investigate language-specific influences, and we predict that listeners will be less likely to perceive an epenthetic stop as a token of the stop phoneme if it would violate a syllable structure constraint in their language.

We use two tasks, a phoneme monitoring task and a dictation task, to test the role of processing demands and orthographic knowledge in perception of epenthetic stops. We measure number of responses, both to epenthetic stops and to stops intended by the speaker, for both tasks. We predict that listeners will be more likely to report epenthetic stops in the phoneme monitoring task than in the dictation task, because phoneme monitoring puts listeners under time pressure and emphasizes low-level acoustic cues. We also analyze reaction times from the phoneme monitoring task in order to compare the processing of underlying and epenthetic stops and clarify what demands the processing of phonetic variation places on the listener. Here, we predict that recognition of underlying stops will be faster than recognition of epenthetic stops, because epenthetic stops may not be acoustically as strong as underlying stops, as Fourakis & Port (1986) found.

In the following sections, we will present first the methods involving production of the materials, then information on the frequency with which epenthetic bursts are produced, and then the methods and results for the perception study, along with further acoustic analyses. However, it should be kept in mind that the production and perception studies cannot be entirely separated from each other, since the primary purpose of the acoustic analyses is to allow for comparison of data on perception and production.

2. Materials

In order to investigate the questions discussed above, we created nonwords containing the environment for stop epenthesis as well as, in many cases, a nonepenthetic stop. Using the CELEX database (Baayen, Piepenbrock & Rijn, 1993), we created 14 nonword target items for each of the categories shown in Table I. These categories included monosyllabic items with coda clusters consisting of each of the nasals /m, n, ŋ/ followed by each of the obstruents /p, t, k, s/. (Additional fricatives were not used in order to limit the size of the experiment, and because epenthetic stops seem to be more likely before /s/ than before many other fricatives (Barnitz, 1974; Ali *et al*., 1979).) This resulted in a total of 168 items. Table I shows which stimulus types provide an environment for which epenthetic stops.

The items began with either one or two consonants (never a stop) and had one of the short vowels of Dutch, followed by the nasal-obstruent cluster. All items were nonwords in Dutch, and embedded real words were avoided as much as possible. Matched triplets were used for each intended obstruent. That is, for each column of Table I, there were 14 sets of three stimuli with only the nasal varied (e.g. /flemp, flenp, fleŋp/). Within the limits of the Dutch phoneme inventory and phonotactic constraints, it was not possible, however, to find matched sets of 12 nonword items with only the nasal and following obstruent varied, so items for the various following obstruents did not always have an identical syllable onset and nucleus. All target items are listed in Appendix A.

In order to test the effect of phonology of listeners' language on perception of epenthetic stops, we introduced a phonotactic legality manipulation. Dutch syllable structure allows for a short vowel to be followed by up to two consonants within the syllable, with

TABLE I. Stimulus types. Rows indicate the nasal of the final cluster, while columns indicate the final obstruent of the cluster. Bracketed transcription shows which epenthetic stop (if any) is expected to be possible. Epenthetic stops appear in parentheses. Clusters in which the epenthetic stop would violate a syllable structure constraint if it were phonemically present are italicized

	intended /p/	intended /t/	intended /k/	intended /s/
/m/	(C)CVmp	(C)CVmt [m(p)t]	(C)CV*mk* [*m*(*p*)*k*]	(C)CVms [m(p)s]
/n/	(C)CV*np* [*n*(*t*)*p*]	(C)CVnt	(C)CV*nk* [*n*(*t*)*k*]	(C)CVns [n(t)s]
/ŋ/	(C)CV*ŋp* [ŋ(*k*)*p*]	(C)CVŋt [ŋ(k)t]	(C)CVŋk	(C)CVŋs [ŋ(k)s]

additional coronal consonants possible if they form a syllable appendix (Booij, 1995). Thus, only the items with /t/ or /s/ as the intended obstruent would form phonotactically legal syllables if the epenthetic stop were phonologically present. Tautosyllabic sequences of heterorganic nasal and obstruent are only phonotactically possible in Dutch if the obstruent is coronal. Hence, items with noncoronals as the intended obstruent (e.g. /flemk/) violate a syllable structure constraint even without the epenthetic stop. However, the presence of the epenthetic stop (/flempk/) would violate an additional syllable structure constraint, and we expect listeners to be less likely to respond to epenthetic stops on the basis of this additional violation.

An additional 183 phonotactically possible nonword items, consisting of both mono- and bisyllabic items, were created as fillers. In 124 items, such as /flem/ and /blimzəl/, the sounds /p/, /t/, and /k/ (the targets for phoneme monitoring) did not occur. In the rest of the fillers the target sounds occurred in a variety of positions (e.g. /delp, nist, kiːmər/). The three nasals /m/, /n/ and /ŋ/ occurred 114 times in the fillers. A list of 14 practice items was also constructed with similar materials.

All materials were recorded onto a DAT tape in a soundproof booth and sampled at 16 kHz. Two female native speakers of Dutch who have backgrounds in linguistics but did not know the purpose of the experiment were recorded. Two productions of all items were recorded from one speaker, and one production from the other. For both speakers, one repetition of each of a few items was inadvertently omitted during the recording session. Stimulus items and fillers were presented in random order in Dutch orthography, except that capital "N" was used to mark the alveolar nasal when it occurred before /k/, since the orthographic sequence "nk" represents an assimilated /ŋk/ cluster in Dutch. /ŋk/ was spelled "nk" in the recording materials. The speakers were instructed in producing nasals without assimilating them to the place of the following obstruent, which they did not find difficult, but were otherwise told only to read the items as naturally as possible. Epenthetic stops were not marked in the materials presented to the speakers, and the speakers were instructed to read the sequence as it was spelled in the materials. The speakers did not intentionally produce epenthetic stops, but often did produce them phonetically.

3. Overall production results: presence of epenthetic bursts

Although the primary purpose of this study is perceptual, there are large differences among the final cluster conditions shown in Table I with regard to how often speakers produce epenthetic bursts and how acoustically strong the cues for epenthetic stops are.[3] It is necessary to understand the differences in production among the conditions in order to interpret the differences in perception. We carried out two sets of analyses: one of all productions, and one of just the items used as stimuli for the perception study. For all productions, we evaluated the proportion of productions containing epenthetic bursts. This allows us to determine whether there are general patterns, consistent across more

---

[3] By the "strength" of an epenthetic stop, we mean its strength on the several acoustic measures discussed in Section 5.6.1: epenthetic burst duration, power of the burst, duration of silence before the burst, duration of preceding nasal, and duration of total closure before the release of the following obstruent. These measures correlate with perceptibility of the epenthetic stop, as discussed in Section 5.6.2. Except for nasal duration, a greater value indicates a more perceptible, thus stronger, epenthetic stop. We also consider epenthetic stops which have high values only on some of the acoustic measures to be relatively strong epenthetic stops.

than the one production used as stimuli for the perception experiment, in how often epenthetic bursts are produced in the various nasal-obstruent clusters. A thorough investigation of production of epenthetic bursts would require many more speakers than the two used here, but this small set of data suffices to show that the patterns across cluster conditions are not specific to a single production. By treating the two productions from the speaker who produced the materials twice as if they were from separate speakers, we can also carry out a preliminary statistical analysis. For the items which were used as stimulus items in the perception experiment (see Section 4.1), we performed further acoustic analyses of the epenthetic burst and surrounding sounds. These measurements are primarily useful in comparison with the perceptual results, and are therefore reported in Section 5.6.1. For all acoustic analyses, we used the XWaves/ESPS software.

Ideally, one would like to determine the proportion of items with epenthetic stops, rather than the proportion with epenthetic bursts. However, it is not always possible to determine from acoustic information alone whether a given token was produced with an epenthetic stop. An epenthetic stop burst is sufficient to show that an epenthetic stop was produced. However, even if no epenthetic burst is present, the speaker could still have ceased voicing and closed the velum before releasing the oral closure of the nasal, thus producing an epenthetic stop, if the duration of complete vocal tract closure was too short for sufficient oral pressure for a burst to build up. In this case, there would be an epenthetic stop with an inaudible release. Furthermore, in clusters such as /ŋp/, an epenthetic [k] might be released while the labial closure is already being made, thus preventing the velar burst from appearing in the acoustic record. However, such burstless epenthetic stops might be perceptible based on other cues. Since we cannot determine without articulatory information how many tokens contained epenthetic stops (defined as a period of velic closure while an oral closure at the place of articulation of the nasal is being made), we evaluated instead the number of tokens produced with epenthetic bursts.

For items with /p, t, k/ as the intended following obstruent, a token was considered as having an epenthetic burst if there was a broadband burst-like noise visible in the spectrogram either late in the nasal or during the silence between the nasal and the burst of the intended stop. In order not to discount any epenthetic bursts, even extremely brief and low-amplitude broadband bursts were counted. Fig. 1 shows spectrograms of a clear epenthetic burst, a weak epenthetic burst, and a token with no epenthetic burst, all before intended stops. If the intended obstruent was /s/, there was sometimes a silent period between the cessation of voicing for the nasal and the onset of frication noise, without any broadband burst. This was also counted as an epenthetic burst (or more precisely, as an epenthetic stop for which the instant of release can be located in the spectrogram). However, most nasal-/s/ tokens did have a brief burst just before the frication noise and distinct from it, with broader band energy than the frication noise. Fig. 2 shows a nasal-/s/ stimulus with an epenthetic burst.

The proportion of items with epenthetic bursts in each final cluster condition with the environment for epenthesis, by speaker, is shown in Fig. 3. (A few tokens with ambiguous nasals, as described in Section 4.1, were excluded from this analysis.) There are very few epenthetic bursts in the /np/ condition, relatively few in the /ŋp/ condition, more in the /mt, mk, ŋt/ conditions, and a very high number in the remaining conditions. The percentage of epenthetic bursts produced differs significantly among cluster conditions ($F(8, 16) = 50.32$, $p < 0.001$).

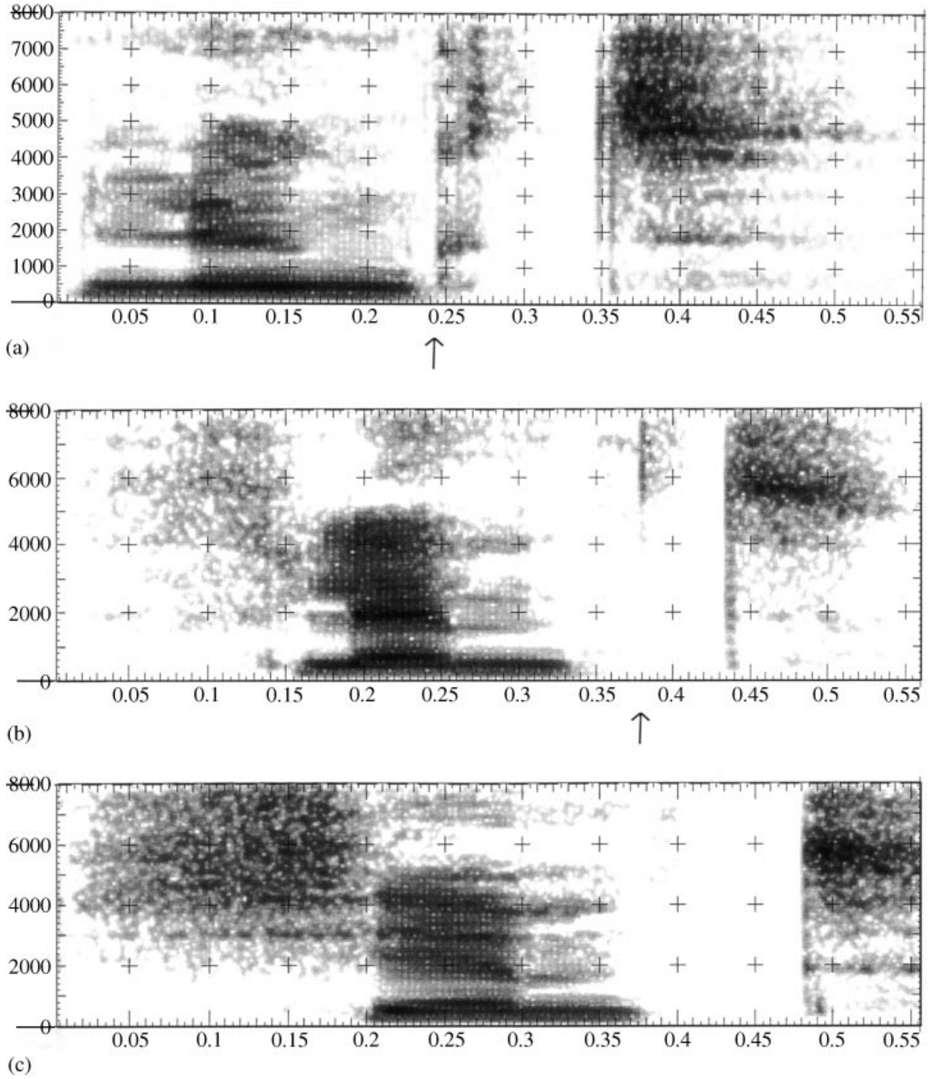

Figure 1. Spectrograms of an item with a strong epenthetic burst (/lyŋt/ [lyŋkt] 1(a)), a weak epenthetic burst (/flemt/ [flempt] 1(b)), and no visible epenthetic burst (/ʃemt/ [ʃemt] 1(c)). In 1a and 1b, the arrow marks the beginning of the epenthetic burst.

A possible reason for some of these differences involves the articulatory origin of epenthetic stops. Although the usual articulatory explanation of epenthetic bursts involves early closure of the velum relative to release of the first oral closure, as described in Section 1.1, there is another sequence of events which could lead to epenthetic bursts. Ohala (1995) shows for /mn/ clusters that if the closure for the /n/ is made before the more anterior closure for the /m/ is released, during the release of the labial, air pressure in the small cavity sealed off between the two closures can drop, causing a weak ingressive burst at the labial release. The same could apply to

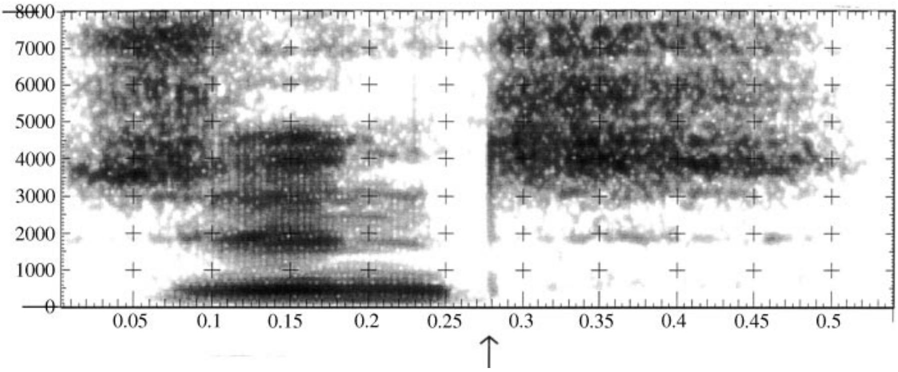

Figure 2. Spectrogram of /zyns/ [zynts], an item with a burst which is distinct from the frication noise. The arrow marks the epenthetic burst.

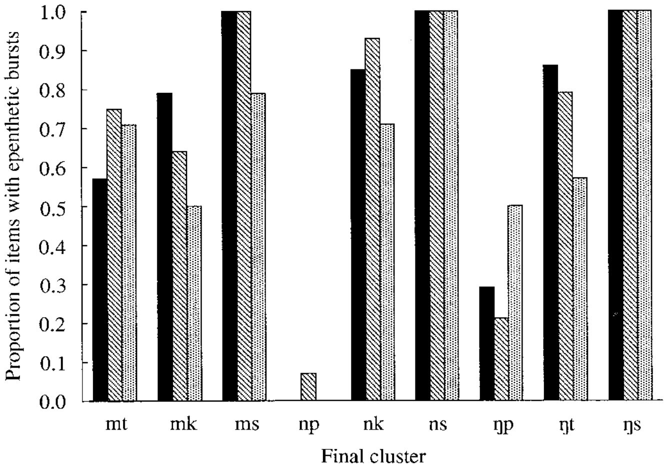

Figure 3. Proportion of items in each cluster in which an epenthetic burst was produced, by speaker. "Speakers" 1a and 1b are the first and second productions by speaker 1, respectively. ■ speaker 1a; ▨ speaker 1b; ▦ speaker 2.

nasal-stop clusters, but only if the closure for the nasal is anterior to the closure for the intended stop. Thus, in the clusters /np, ŋp/, epenthetic bursts can only be produced by the sequence of events described in Sections 1.1 and 1.2, while in many other clusters, there are two types of mistiming which could produce epenthesis. In a cluster with the nasal anterior to the intended stop, the release of the anterior oral closure can create an audible epenthetic burst whether the closure for the following stop is already being made or not. In a cluster with the nasal posterior to the intended stop, though, the oral closure for the nasal must be released before the closure for the following stop is made in order for the burst not to be blocked by the anterior closure. These factors may contribute to the low rate of production of epenthetic bursts in the /np, ŋp/ items.

4. Methods for the perception experiment

4.1. *Perception experiment materials*

Of the two speakers, the recording of the speaker who had produced two tokens of each item was used for the perception experiment, and of that speaker's two productions, the better pronunciation of each item was selected for use as a stimulus. The first author confirmed, by listening to each item without its final obstruent, that the speaker had produced the correct nasal in each item, and not assimilated it to the following intended obstruent. Very few items had ambiguous or incorrect place for the nasals, and in these cases, the other token was used for the stimulus if possible, leaving only a few stimuli with ambiguous place and none with incorrect place. For items with two useable tokens and the environment for stop epenthesis, the production with the stronger epenthetic stop was selected based on examination of the waveform and spectrogram, and listening.

Since the tokens chosen as stimuli could come from either the first or second recording of the speaker, and were chosen to maximize the number of stimuli containing epenthetic bursts, the proportion of stimuli with epenthetic bursts cannot be determined from Fig. 3 above. Fig. 4 shows the proportion of stimulus items in each condition in which an epenthetic burst was produced. (Conditions without the environment for any epenthetic stop are omitted.) In the conditions /ms, nk, ns, ŋt, ŋs/, nearly all items chosen for use as stimuli have epenthetic bursts. In the condition /np/, only one item contained an epenthetic burst.

The target and filler items described in Section 2 were used to construct four lists. Each list contained all 168 experimental stimulus items and 183 fillers in a different pseudorandom order, such that there was always at least one filler without /p/, /t/, or /k/ before a target item.

4.2. *Tasks*

We used two experimental tasks for the perceptual aspect of this study, phoneme monitoring, in which subjects had to press a button as quickly as possible when they heard

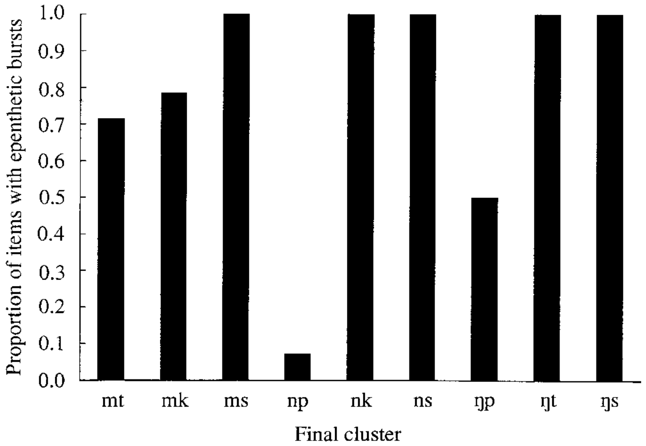

Figure 4. Proportion of stimulus items with an epenthetic burst for each cluster (silent period before onset of frication sufficient for inclusion of nasal-/s/ items). Only items used as stimuli are included, and only clusters with the environment for epenthesis are shown.

a particular sound (/p/, /t/, or /k/), and a dictation task. Phoneme monitoring involves two processes for the subjects, listening to speech and detecting the predetermined target sound. (For an overview of the method see Connine & Titone, 1997.) We used the "generalized phoneme monitoring" procedure (Frauenfelder & Segui, 1989), in which the target phoneme can occur anywhere in the stimulus, rather than at a prespecified position. When listeners monitoring for /t/, for example, hear a stimulus containing an epenthetic [t], they must decide whether the stimulus contained the sound /t/ or not. The number of responses in phoneme monitoring allows us to determine how often listeners perceive epenthetic stops as tokens of the stop phoneme, and thus how often they succeed or fail in recovering the string of segments intended by the speaker. We also measured reaction times in the phoneme monitoring task, which can show whether listeners perceive epenthetic and intended stops as equally good tokens of the stop. Reaction times are assumed to reflect variations in speech processing, which means that longer reaction times are associated with greater processing load. Furthermore, phoneme monitoring puts subjects under time pressure, and encourages them to attend to low-level acoustic information, which may increase their likelihood of misperceiving epenthetic stops as real.

In the second task, listeners transcribed each stimulus item in Dutch orthography to dictation, using a computer keyboard. This task, unlike the phoneme monitoring task, gives the listener ample time for processing the signal. Because subjects must transcribe their responses, it may encourage them to invoke knowledge of what patterns are typical in the orthography of the language. (However, since stimuli are nonwords, orthographic knowledge provides no information about specific stimuli.) The measure for this task is the number of responses containing epenthetic stops.

4.3. *Subjects and procedures*

Seventy-two native speakers of Dutch, primarily students at the University of Nijmegen in the Netherlands, took part in the experiment. They were paid for their participation. In the phoneme monitoring task, 24 subjects monitored for /p/, 24 for /t/, and 24 for /k/. The subjects were tested one or two at a time in separate soundproof booths. The stimuli were presented over headphones, and the NESU experimental software was used to control the presentation of stimuli and to measure reaction times and record dictation responses. Subjects were presented with one of the lists for phoneme monitoring and a different one for dictation.

Subjects first performed the phoneme monitoring task, then the dictation task. For phoneme monitoring, subjects were given written instructions telling them that they would hear Dutch nonwords, and that their task was to press the button in front of them as quickly as possible if they detected the target sound (/p/, /t/, or /k/, depending on which group the subject had been assigned to). The instructions (in Dutch) described the target sound as "the sound 'k'" for example, and gave one example nonword in orthography which contained the target sound (not in a nasal-obstruent cluster). Because Dutch orthography is rather regular, target phonemes can be described orthographically. Subjects were told that not all items contained the target sound. Subjects could press the button with either hand. Each subject heard the practice list first, followed after a short pause by all experimental stimuli and fillers in one of the four pseudo-randomized orders. The phoneme monitoring experiment lasted approximately 18 min.

After a short break, the subjects performed the dictation task. They were given written instructions telling them that they would hear the same nonwords in a different order, and that after each item, they should type the item using a computer keyboard. Subjects were allowed up to 16 s to type their responses, and correct typing errors if necessary, before the next item was presented automatically. However, if they finished sooner, they could go on to the next stimulus immediately. They were given the same 14 item practice list, followed by a brief pause and the experimental list. The dictation task lasted approximately 30 min, depending on the performance of individual subjects, and there was one break.

4.4. *Data analysis*

For the phoneme monitoring task, the measures are number of responses to the target sound and reaction time. When subjects responded to experimental stimulus items (either to an intended or to an epenthetic version of the target sound), reaction time was measured from the beginning of the burst of the relevant target sound. That is, for a k-monitoring subject hearing /flemk/, reaction time was measured from the burst of the /k/. However, for a k-monitoring subject responding to /fleŋp/, reaction time was measured from the beginning of the epenthetic burst if there was one, and from the end of voicing for the nasal if there was not (if the listener responded even though no epenthetic burst was present). Only reaction times either to intended stops or to stops in the appropriate environment for epenthesis were analyzed, so in the rare cases of other responses (e.g., a k-monitoring subject responding to /flemt/) reaction times were not analyzed, although number of responses was. Reaction times shorter than 100 ms or longer than 1500 ms were treated as errors, and counted as nonresponses.

In the dictation task, the number of responses containing the stop (whether intended or epenthetic) was counted. Data from the dictation task was only analyzed for presence of the phoneme for which the subject had been monitoring in the phoneme monitoring task. Although subjects were not told to pay attention to any particular sound during the dictation task, and their responses do provide information about their perception of all stops, the experience of having just monitored for a particular sound in the phoneme monitoring task could affect their responses to that sound *vs.* other epenthetic sounds in the dictation task. For the dictation task, each listener's response to each item was counted as either a positive response or a nonresponse. If the dictation response contained the target sound (the target from the phoneme monitoring task), this was treated as a positive response. For example, if a subject who had previously monitored for /p/ heard /fremp/ and responded "fremp", or heard /flimt/ and responded "flimpt", these were counted as positive responses. (Note that although /mt/ is an environment for stop epenthesis, this does not mean that an epenthetic stop was necessarily phonetically present.)

Because of the open response methodology and because subjects were not given any instructions about how to transcribe their responses except for being told to use Dutch orthography, several factors had to be considered in calculating the number of responses. For t-monitoring subjects, "d" in the response (e.g., "lund" for the stimulus /lynt/) was counted as a positive response as well as "t" because of Dutch final devoicing. (We would have counted "b" as a positive response to /p/, but this never occurred. "d" for "t" occurred 29 times.) The letter "g" was not accepted as a positive /k/ response, since Dutch orthographic "g" is not realized as a stop. Any occurrence of a letter representing the target phoneme, if located somewhere after the vowel in the response, was counted. All of the following were counted as positive responses as long as a letter representing the target sound was present: reversed order of sounds in the final cluster (e.g., a k-monitoring subject heard /zyŋp/ and responded “zuknp” instead of “zunkp”), incorrect nasal (e.g., a p-monitoring subject heard /frymp/ and responded “frunp”), incorrect initial sound (e.g., a t-monitoring subject heard /ʋyns/ and responded “bunts”), and incorrect vowel (e.g., a p-monitoring subject heard /ʃimp/ and responded “sjemp”). Any other response was treated as a nonresponse. These allowances were made because the perception of the initial consonant(s) and vowel are irrelevant to the current study and because subjects were not instructed in how to transcribe assimilated and unassimilated nasals. In the case of reversed segments in the final cluster (which were rare), it is unlikely that a subject responding “zuknp” to /zyŋp/ really means to show that he heard a /k/ before the nasal and not after it. A typing error is more likely, and it is clear that the subject did perceive a /k/ in the coda.

5. Perceptual results

We will present the results of the perception study as follows: first, we will give an overview of how often subjects perceive epenthetic stops in the stimuli, and of the differences in response rate across final cluster conditions (Section 5.1). We will also discuss how often subjects perceive epenthetic stops in stimuli with and without epenthetic bursts (Section 5.2). We will then discuss two subsets of the data which showed surprising results: the /n(t)k/ condition, in which epenthetic bursts are present but few subjects respond to them (Section 5.3), and items in which no epenthetic burst is present, some of which elicit many responses (Section 5.4). We will then turn to the question of whether the epenthetic stops listeners perceive are the ones predicted by the articulatory account of epenthesis, along with the question of how often listeners perceive epenthetic stops relative to intended stops (Section 5.5). In Section 5.6, we will present acoustic measures of the epenthetic stops and address the relationship between production and perception by comparing these acoustic measures to the perceptual response rate. Section 5.7 treats the effect of phonotactic legality on perception of epenthetic stops. In Section 5.8, we consider the processing of epenthetic stops through analysis of the reaction time data. Finally, in Section 5.9 we compare the phoneme monitoring and dictation tasks.

5.1. *Overall perceptual results by cluster condition*

The black bars in Fig. 5 show the average proportion of stimuli in which subjects responded to epenthetic stops in each final cluster condition, for the phoneme monitoring data (data shown by gray bars will be discussed in Section 5.9). This graph includes only data in which the stop the subject is monitoring for could occur epenthetically, hence the /mt, mk, ms/ conditions represent data from the p-monitoring subjects, the /np, nk, ns/ conditions represent the t-monitoring subjects, and the /ŋp, ŋt, ŋs/ conditions represent the k-monitoring subjects. There are large differences across final clusters in how often epenthetic stops are perceived. In the /ŋt, ŋs/ conditions listeners respond to epenthetic stops very often. The /np, nk/ conditions receive very few responses, and the other cluster conditions are intermediate.

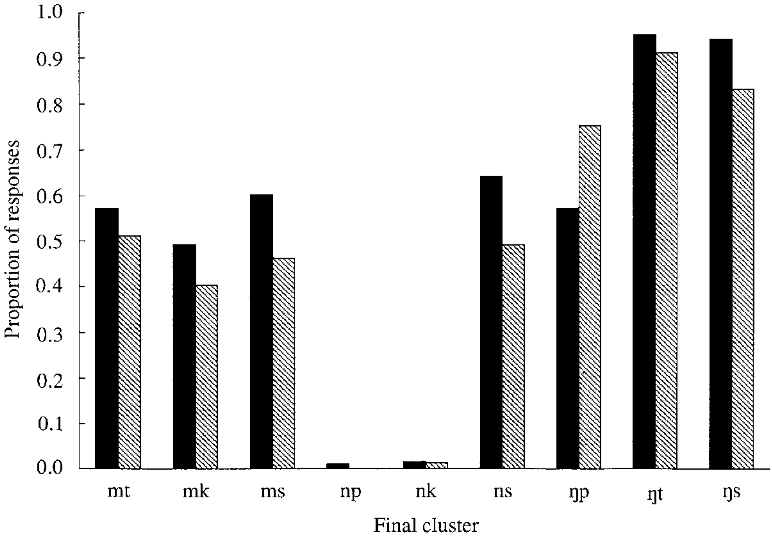

Figure 5. Average proportion of responses to epenthetic stops, for each cluster, for the phoneme monitoring and dictation data. ■ monitoring task; ▨ dictation task.

In some cases, these differences are not surprising based on the production data: the low response rate to /np/ items is probably explained by the low rate of production of epenthetic bursts in that condition. The high response rate to epenthetic [k] in the /ŋt, ŋs/ conditions may reflect not only the prevalence of epenthetic bursts in these stimuli, but also the fact that [k] has the longest resonating cavity anterior to the constriction, while [p] has none at all, leading to louder bursts for [k] than [p]. [k] also tends to have longer bursts than other stops (Stevens, 1998). In our data, these inherent differences are reflected in the epenthetic burst duration and peak power ratio measurements discussed in Section 5.6.1. The /ns/ items receive relatively few responses, even though all /ns/ items have epenthetic bursts. These epenthetic bursts may be relatively weak because /n/ and /s/ are homorganic: the tongue does not have to move far from the alveolar ridge to form the following /s/, resulting in very short epenthetic bursts (see Fig. 8). However, in some cases, the response rate is surprising when compared to the production data. The unexpectedly small number of responses to epenthetic [t] in the /nk/ cluster will be discussed in Section 5.3.

Averaged across all clusters, subjects responded 53% of the time in the phoneme monitoring task when the stop they were monitoring for could have occurred epenthetically. The response rate of 53% is less than the percentage of stimuli in which epenthetic bursts were produced (79%, see Fig. 4), so not all epenthetic bursts are perceived, but this does show that epenthetic stops are perceived quite often.

5.2. *The effect of presence of an epenthetic burst on perception of epenthesis*

Although the absence of an epenthetic burst in an item does not rule out the presence of an epenthetic stop, the presence of an epenthetic burst is likely to be a very strong perceptual cue. Therefore, we examined the proportion of stimuli with an epenthetic burst in which subjects perceived an epenthetic stop. These data, by final cluster, appear

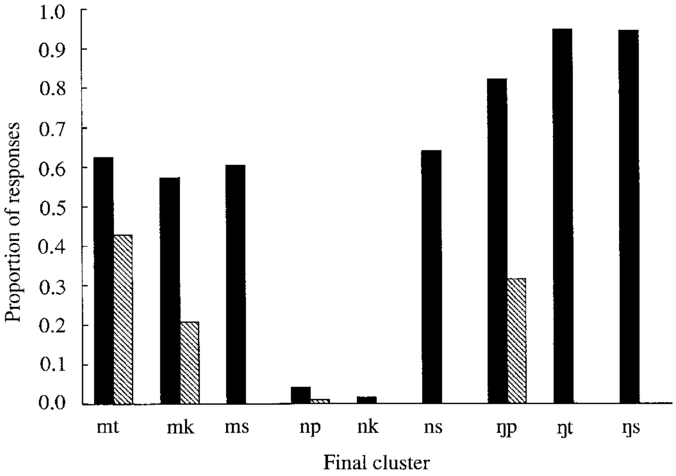

Figure 6. Average proportion of stimuli with and without an epenthetic burst in which listeners report an epenthetic stop. The clusters /mt, mk, np, ŋp/ are the only ones which have any items without an epenthetic burst. ■ with epen. burst; ▨ without epen. burst.

in Fig. 6. (As in Section 5.1, only data in which the stop the subject was monitoring for could occur epenthetically are included, e.g. /mt, mk, ms/ clusters for p-monitoring subjects, etc.) Viewed this way, it is clear that responses to epenthetic [k] (if a burst is present) are more frequent than responses to other epenthetic stops, and that responses to epenthetic [t] in /np, nk/ clusters are rare.

For the four final clusters in which there are stimuli lacking epenthetic bursts. Fig. 6 also shows the proportion of stimuli without epenthetic bursts in which subjects perceived epenthetic stops. It is apparent that in the /mt, mk, ŋp/ conditions, a burst is not necessary in order for subjects to perceive an epenthetic stop. The question of what cues listeners might use in these cases is discussed in Section 5.4.

Fig. 6 shows that listeners are more likely to perceive an epenthetic stop if the stimulus contains an epenthetic burst than if it does not, and it seems likely that the presence of an epenthetic burst would be a strong perceptual cue. However, since a burst is not necessary for the perception of an epenthetic stop, it is possible that listeners could simply guess that there is likely to be an epenthetic stop in the appropriate nasal-obstruent environments because epenthetic stops do often occur in those environments. If listeners are sensitive to the presence of an epenthetic burst, then items to which many subjects respond should be more likely to have epenthetic bursts than other items are. We tested this by examining all items to which more than three subjects responded (among those subjects who were monitoring for the stop which could occur epenthetically in that item). We calculated the percentage of such items which had a visible epenthetic burst in the spectrogram, and performed a chi-squared test to determine whether the proportion of items with an epenthetic burst was higher for items which received more than three responses than it was for all items. (That is, the expected frequencies for the chi-squared test were calculated from the proportion of items with epenthetic bursts among all items in the experiment.)

For the phoneme monitoring task, the items which received more than three responses are significantly more likely to have an epenthetic burst than items are overall

($\chi^2 = 4.55$, $p < 0.04$). That is, items with epenthetic bursts are concentrated among the items receiving more than three responses. (Three was chosen because a small number of subjects may sometimes respond mistakenly.) In order to confirm this finding, we performed the same analysis for the dictation task. (Throughout the paper, except where the effect of task is at issue, we analyze the dictation data to confirm the phoneme monitoring results.) For that task, the difference is not significant for items with more than three responses ($\chi^2 = 2.67$, $p = 0.102$), but it is for items with more than six responses ($\chi^2 = 4.01$, $p < 0.05$). This indicates that the phoneme monitoring results are somewhat more closely tied to acoustic cues than the dictation results are. The results confirm that listeners' likelihood of responding to items which have the environment for epenthesis is related to whether an epenthetic burst was produced in a given item.

5.3. *The* [*n*(*t*)*k*] *cluster*

Although in general items with epenthetic bursts receive more responses, some items have an epenthetic burst, but received almost no responses. All such items except one are in the /nk/ condition. Of the 14 items in the /nk/ condition, all had epenthetic bursts, but only four received any responses at all, and even those received responses from no more than two of the 24 subjects. Many of the epenthetic stops in this condition are quite strong: although the /nk/ condition has a relatively short silent period before the epenthetic burst, and exceptionally long nasal duration, it also has relatively long epenthetic bursts, and the total closure duration is not short. In many items, an epenthetic [t] is clearly perceptible to listeners who know the purpose of the experiment, when /nk/ items are presented in isolation. Furthermore, it should be emphasized that the speaker did not produce [ŋk] clusters in place of the [n(t)k] clusters. Listeners' failure to respond to these items may relate to the long nasal duration in this condition. It will be shown in Section 5.6.2 that nasal duration correlates (negatively) more strongly with number of responses than most other measures do. However, correlation does not show causation, and the correlations with many other acoustic measures are lowered by precisely the /nk/ items.

It is difficult to determine why so few listeners respond to the epenthetic stops in these items. It might seem that responses to epenthetic [t] in the /nk/ cluster are rare because the coda cluster /ntk/ would be phonotactically illegal (the topic of Section 5.7). However, this cannot be the reason. First, the clusters /mpk, ŋkp/ would also be phonotactically illegal, and yet there are far more responses to the epenthetic stops in these conditions. Secondly, in an additional manipulation not reported in this paper, we investigated the perception of epenthetic stops in bisyllabic forms (e.g., /frenkyːs/ instead of /frenk/). In this environment, the cluster /ntk/ would be phonotactically legal because it does not form a syllable coda. Even in this bisyllabic condition, however, responses to epenthetic [t] in /nk/ clusters were very rare.

There is an alternative explanation for subjects' failure to perceive epenthetic [t] in /nk/ clusters. Several other lines of research have found some indication that assimilation of /n/ to a following /k/ is perceptually different from assimilation between other places of articulation, at least for Dutch listeners. Otake, Yoneyama, Cutler & Lugt (1996) found that Dutch listeners, when instructed to monitor for "n" in a phoneme monitoring task, responded to [ŋ] in the environment of a following [k] nearly as often as they responded to dental [n̪] or alveolar [n]. Listeners did not respond to a bilabial nasal before a bilabial stop, though. Cutler & Otake (1998) found that when Dutch speakers are presented with a CVNCVC sequence and another bisyllabic sequence with a different consonant as the onset of the second syllable, and asked to combine the first syllable of one with the second syllable of the other, they usually do not assimilate the place of the nasal, although that creates a heterorganic NC cluster. However, when required to combine a first syllable ending in [ŋ] with a second syllable beginning with an alveolar stop (e.g., combining /tiŋkerk/ with /wiːdɛik/), subjects produced an assimilated form (i.e., /tindɛik/) quite often. This suggests that the Dutch subjects may have analyzed the stimulus /tiŋkerk/ as underlying /tinkerk/ with an assimilated nasal, as such a cluster would often be spelled.

The fact that these anomalous results occur only with alveolar–velar combinations, and not with alveolar–labial clusters, suggests that the difference has more to do with spelling conventions than with phonological rules of assimilation: /n/ assimilates before bilabials just as it does before velars in Dutch, but only in the velar case is the spelling ambiguous as to assimilation. A nasal homorganic with a following bilabial is spelled "m", but a nasal homorganic with a following velar is still spelled "n". Listeners' failure to respond to epenthetic [t] in the /nk/ condition may be related to orthography rather than to the acoustic cues for the epenthetic stop, but the problem applies only to this cluster.

5.4. *Items in which no epenthetic burst is produced*

In contrast to the situation just discussed, for some items there is no epenthetic burst visible in the spectrogram and yet relatively large numbers of listeners report hearing the stop for which an epenthesis environment is present. As discussed in Section 3, the absence of an epenthetic burst does not necessarily indicate the absence of an epenthetic stop, but this leaves the question of what other cues lead listeners to perceive an epenthetic stop in the absence of a burst. Of the 126 items in the perceptual experiment which had an environment for epenthesis, 27 items had no epenthetic burst visible in the waveform or spectrogram. Of these 27, 11 items received no responses, but nine received eight or more responses (up to a maximum of 14 out of 24 subjects responding). For example, 10 k-monitoring subjects responded to the item /fliŋp/, even though no epenthetic burst was present. These items are not randomly distributed among the cluster conditions: all such items to which three or more subjects responded have /m/ or /ŋ/ as the nasal, thus they had the environment of epenthetic [p] or [k] (see Fig. 6). All but one item with two or fewer responses had /n/ as the nasal, and thus had the environment for epenthetic [t].

Thus, among items with no epenthetic burst produced, epenthetic stops are reported if they should be [p, k] but not if they should be [t]. This might be due to the identity of the stops themselves: perhaps listeners are biased against interpreting epenthetic [t] as a real stop. However, it might be due to acoustic differences in the stimuli. Among items with no epenthetic burst, those with three or more responses have a significantly longer total closure duration for the intended stop (see Section 5.6.1) than items with fewer responses ($F(1,25) = 12.91$, $p < 0.01$). Closure duration does seem to be a cue for epenthetic stops, as discussed below. Thus, listeners might be responding to items without epenthetic bursts if the closure duration is long enough for two stops.

5.5. *Predicted epenthetic stops, other epenthetic stops, and intended stops*

To determine whether the epenthetic stop listeners perceive is the one predicted by the articulatory account of epenthesis, we analyzed the number of responses separately for subjects monitoring for each stop (i.e. /p, t, k/). For each group, we performed two planned comparisons. One compared all clusters with the target stop phonologically present (intended by the speaker) to all clusters with the environment for that stop to appear epenthetically (e.g., for p-monitoring subjects, a comparison of the /mp, np, ŋp/ conditions to the /mt, mk, ms/ conditions). The second compared all clusters with the environment for that stop to occur epenthetically to all clusters with the environment for some other stop to occur epenthetically (e.g., for p-monitoring subjects, a comparison of the /mt, mk, ms/ conditions to the /nk, ns, ŋt, ŋs/ conditions). Responses to the remaining two conditions (/nt, ŋk/ for p-monitors) were not included in either comparison.

The results of these comparisons (for the phoneme monitoring task) appear in Fig. 7. For each group (p-, t-, and k-monitoring subjects), subjects responded significantly more often when the stop for which they were monitoring was intended by the speaker than when it could occur epenthetically. The difference was significant both across subjects and across items (p-monitoring: t1=11.89, t2=12.79, t-monitoring: t1=32.34, t2=81.59, k-monitoring: t1=10.15, t2=2.65, all tests are significant at $p < 0.001$ and use separate variance estimates). They also responded significantly more often when the stop they were monitoring for was the one predicted to occur epenthetically than when some other epenthetic stop was expected (p-monitoring: t1 = 15.21, t2 = 16.08, t-monitoring: t1 = 7.06, t2 = 19.67, k-monitoring: t1 = 50.83, t2 = 27.94, $p < 0.001$ for all tests), and responses of a wrong epenthetic stop were very rare overall. In the dictation data as well, both comparisons for each stop showed a significant difference at $p < 0.001$, both across subjects and across items, in the same direction as the phoneme monitoring data (comparison of intended to epenthetic stops, p-monitoring: t1 = 13.70, t2 = 19.97, t-monitoring: t1 = 34.14, t2 = 64.27, k-monitoring: t1 = 5.99, t2 = 7.44; comparison of predicted epenthetic stop to other epenthetic stops, p-monitoring: t1 = 11.58, t2 = 16.70, t-monitoring: t1 = 6.53, t2 = 12.28, k-monitoring: t1 = 31.15, t2 = 38.51).

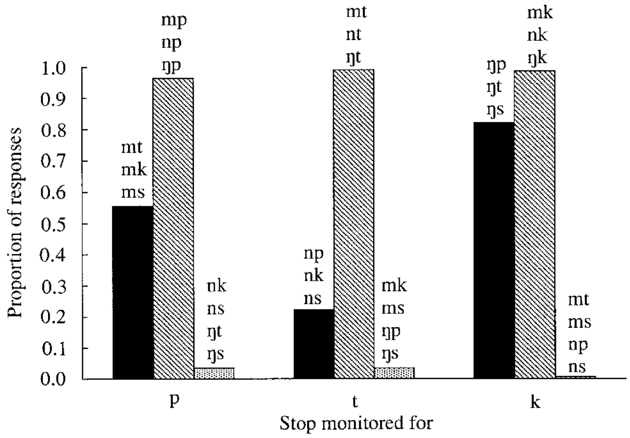

Figure 7. Average proportion of responses to /p, t, k/ in conditions where that stop is the intended obstruent, conditions where the environment for that stop to occur epenthetically exists, and conditions where the environment for some other stop to occur epenthetically exists. The clusters making up the condition are shown above the bars of the graph. Phoneme monitoring data. ■ this stop epen.; ▨ this stop intended; ▦ other stop epen.

These large and consistent differences show that listeners often perceive unintended epenthetic stops as tokens of the stop phonemes, but perceive them less consistently than stops which the speaker intended. Since not all tokens have epenthetic bursts, and some may not have an epenthetic stop in any sense, it is not surprising that listeners respond less frequently to epenthetic stops than to intended stops. However, since listeners do sometimes respond to epenthetic stops even in the absence of a burst, and there may be cues other than the burst to the presence of an epenthetic stop, it is useful to compare the proportion of responses to epenthetic and to intended stops. Furthermore, these results show that when listeners do perceive epenthetic stops, they perceive the one which is predicted by the articulatory account of how such variation is produced.

5.6. *Relationship of perception to acoustic characteristics*

5.6.1. *Acoustic measurements of epenthetic stops*

Several potential perceptual cues concern characteristics of the epenthetic stops themselves, and can only be measured for the stimuli in which epenthetic bursts were visible in the spectrogram. One important measure of the strength of an epenthetic stop is the duration of its burst. For items with an intended following /s/, the duration of the broadband burst before the onset of frication noise was measured. For items with a stop as the intended obstruent, measuring the duration of the epenthetic burst was difficult, because the epenthetic burst noise trails off into the closure for the intended stop, making it difficult to identify the end of the epenthetic burst. All items were measured with the spectrogram display set to consistent threshold and range levels. We defined the end of the epenthetic burst as the last visible noise at any frequency in the spectrogram with noise present at some frequency from the beginning of the epenthetic burst to that point.

Fig. 8 shows the average epenthetic burst duration for stimuli which had epenthetic stop bursts. Because some conditions (e.g., /np/) had very few epenthetic bursts, these data are not equally reliable for all conditions. However, there are some potential trends in the data. The one /np/ item with an epenthetic burst has an exceptionally short burst. Bursts before /s/ are, for the most part, shorter than bursts before intended stops, since frication begins shortly after the burst. We do not analyze these data statistically, nor the following acoustic measures, because measures which can only be done when an epenthetic burst is present represent drastically unequal numbers of items across conditions. Furthermore, all of the acoustic measures are primarily useful for correlational analyses with the perceptual data. Correlation analysis allows us to determine whether items with a longer burst elicit more responses, even if burst duration varies greatly within the items with a particular final cluster.

We also measured the duration of the silence from the cessation of voicing for the nasal to the onset of the epenthetic burst. This measure is equivalent to the closure duration for the epenthetic stop. An epenthetic burst well separated from the preceding nasal might be more readily perceptible as a stop than one very close to the nasal, since an epenthetic burst immediately following or even overlapping the voicing of the nasal would lack the closure portion of a stop. For this measurement, we defined the end of the nasal as the end of the strong voicing bar (the end of regular voicing), not the end of all traces of low-amplitude vibration. These data also appear in Fig. 8. There are no clear trends based on the place of articulation of the nasal (and hence epenthetic stop), but epenthetic bursts before /s/ seem to occur relatively soon after the nasal.

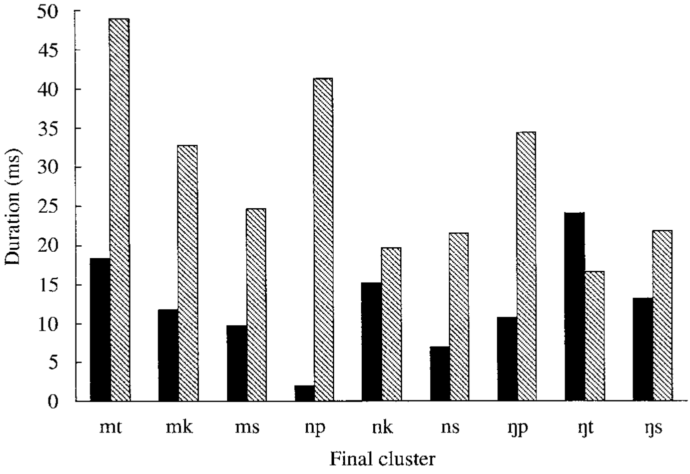

Figure 8. Average duration of epenthetic stop burst and average duration of silence from end of nasal voicing to beginning of epenthetic burst, by cluster, for stimulus items containing epenthetic bursts. ■ epenthetic burst; ▨ silence before epenthetic burst.

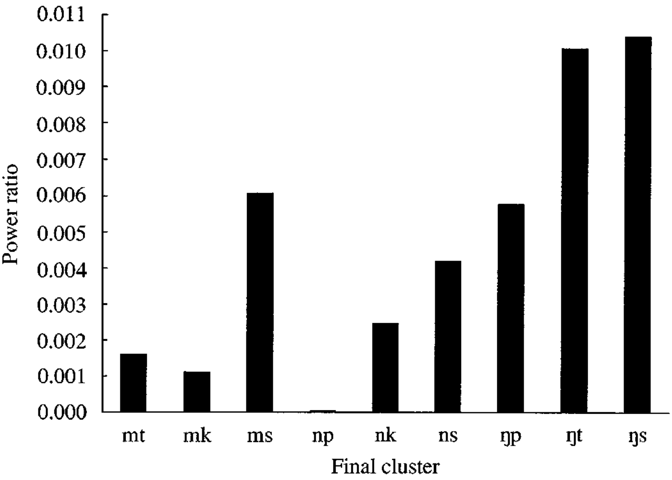

Figure 9. Average ratio of peak epenthetic burst power to peak vowel power, by cluster, for stimulus items containing epenthetic bursts.

We also evaluated amplitude of epenthetic bursts by locating the peak power during the epenthetic burst and the peak power of the preceding vowel, and calculating the ratio of the two power measurements. The results for this measure appear in Fig. 9. The strength of epenthetic bursts before /s/ is overestimated by this measure, because the frication of the immediately following /s/ raises the power value of the preceding burst, while before an intended stop, the closure period of the following intended stop prevents any influence of the following burst on the peak power value for the epenthetic one. Also, items in which the epenthetic burst occurs very close to the end of the nasal have unduly

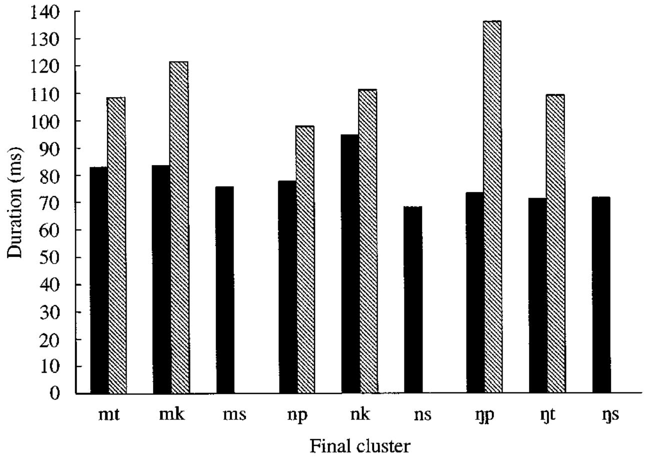

Figure 10. Average duration of nasal and average duration of entire closure from end of nasal voicing to beginning of intended stop burst, by cluster, for all stimulus items. Closure duration not included for clusters ending in /s/. ■ nasal; ▨ total stop closure.

elevated peak power because the power value is still influenced by the voicing of the nasal. One can see in Fig. 9 that epenthetic bursts before /s/ have very high power by this measure, but aside from this artefact, it also appears that epenthetic [k] bursts have higher power than others.

In addition to the acoustic characteristics of the epenthetic burst itself, there are several characteristics of the nasal-obstruent cluster which might serve as perceptual cues to an epenthetic stop. These characteristics can be measured for all stimuli, even those without epenthetic bursts. The duration of the preceding nasal, for example, might provide a cue: a short nasal might suggest a sequence of stops. We measured the duration of the nasal from the sudden discontinuity in the spectrogram between vowel and nasal to the end of the solid voice bar at the end of the nasal murmur. Both nasal duration data and the total closure duration data described below appear in Fig. 10. Nasals were considerably longer in the /nk/ cluster than in all other clusters, but there were no other clear patterns across the final clusters. In English, one might expect that glottalization during the nasal could provide another cue to the existence of an epenthetic stop. However, since Dutch does not have glottalization before final voiceless stops in general (Collins & Mees, 1981), glottalization is not relevant for our data, and in fact, we found no sign of glottalization during the nasal in any stimulus item.

Another potential cue could be the total duration from the end of the nasal murmur to the release of the intended stop. It is well known that in a spliced VC_1C_2V sequence without the burst of C_1, listeners perceive both stops if the closure duration is long enough, but perceive only C_2 if the closure duration is short (Fujimura, Macchi & Streeter, 1978; Kakehi, Kato & Kashino, 1996). Perhaps listeners are more likely to perceive an epenthetic stop if the silence before the burst of the intended stop is longer. We measured the duration of the total stop closure, from cessation of nasal voicing to the beginning of the intended stop burst (Fig. 10). For items with /s/ as the intended obstruent, we did not use this measure, because the /s/ does not contribute

TABLE II. Coefficients of correlation between number of responses and acoustic measures (r), for each task. All correlations are significant at $p < 0.01$

	Phoneme monitoring	Dictation	n
Epenthetic burst duration	0.41	0.37	126
Epenthetic silence duration	0.37	0.24	126
Burst peak power ratio	0.39	0.37	126
Nasal duration	– 0.47	– 0.43	126
Total closure duration (following /p, t, k/ only)	0.37	0.39	84

any closure duration.[4] There were no clear patterns across the final clusters for this measure.

5.6.2. *Correlation of perceptual results with acoustic measurements*

Although some of the acoustic measures do not show definitive patterns across the final clusters, they may still be useful perceptual cues, which simply do not vary systematically by cluster. In order to determine which acoustic cues are important in the perception of epenthetic stops, we calculated the correlation of the number of subjects responding to an item with each of the acoustic measures in Section 5.6.1. For these analyses, we used only the data where the stop for which the subject was monitoring could occur epenthetically in the item (e.g., the /mt, mk, ms/ conditions for p-monitoring subjects). We calculated these correlations separately for both tasks. For items in which no epenthetic burst was produced, all of the acoustic measures of the epenthetic stop itself (epenthetic burst duration, silence duration before epenthetic burst, power ratio) were set to zero for the purposes of computing correlations. The results are shown in Table II.

It is clear that subjects' likelihood of perceiving the epenthetically produced stop is related to the strength of each of the acoustic measures. More subjects report hearing the epenthetic stops when the epenthetic burst is long, when it is well separated from the voicing of the nasal, and when it has high power relative to the vowel. (However, the relationship of number of responses to power ratio is highly nonlinear, and the significant correlation is probably due to a few items with exceptionally high-power epenthetic bursts.) Furthermore, more listeners perceive epenthetic stops when the preceding nasal is short, and when the total closure duration, including both epenthetic and intended stop closure, is long. These relationships hold for both tasks, although most correlations are slightly weaker for the dictation task. This confirms that the phoneme monitoring task encourages subjects to concentrate on low-level acoustic cues, but that the dictation task does not sever the relationship to acoustics.

5.7. *Syllable structure constraint violations*

In conditions with a coronal as the intended obstruent, the final cluster would be phonotactically possible even if the epenthetic stop were fully (phonologically) present.

---

[4] A potential cue to epenthetic stops in the environment before /s/, however, might be rise time of the frication noise, if no epenthetic burst distinct from the frication is present. Rapid onset of frication noise might signal an affricate-like release, and hence an epenthetic stop. Since our /s/-final stimuli did have epenthetic bursts distinct from the frication noise, this measure was not useful for our materials.

In the other conditions, however, the cluster would be phonotactically impossible if the epenthetic stop were phonologically present. We compared these conditions in order to test the effect of language-specific syllable structure constraints on perception of epenthetic stops. In order to evaluate the syllable structure constraint effect, it is essential to correct for how strong the acoustic cues to the epenthetic stop are. To take one example, if syllable structure constraints have the hypothesized effect, listeners would be more likely to respond to epenthetic [p] in the /mt/ and /ms/ conditions (since /mpt, mps/ are phonotactically permissible coda clusters) than in the /mk/ condition (since /mpk/ is not a possible coda). Listeners would be equally likely to respond in the /mt/ and /ms/ conditions. However, epenthetic bursts are produced less often in the /mt/ condition than in the /mk/ condition, and are produced more often in the /ms/ condition than in either the /mt/ or /mk/ conditions. Differences in any of the other acoustic cues (burst duration, nasal duration, etc.) across clusters could also affect the results. If the differences in production are not accounted for, they might obscure any effect of syllable structure constraints.

The variability across clusters in production may not be fully adjusted for by using any one of the acoustic measures, or the presence or absence of an epenthetic burst, as a covariate. We therefore calculated one total measure of strength of production of epenthetic stops by combining all possible acoustic measures through a discriminant analysis (Noruŝis & SPSS, 1994). We used epenthetic burst duration, duration of silence before the epenthetic burst, duration of the nasal, and the presence or absence of an epenthetic burst as the input to the discriminant analysis. The other measures were excluded, since total closure duration could not be applied to items with /s/ as the intended obstruent, and power ratio is not linearly related to number of responses. We then used the value of each item on the discriminant as the covariate, in order to adjust for the presence or absence of an epenthetic burst and the strength of all of the acoustic measures simultaneously. Any effect which remains in this analysis cannot be due to production differences, unless it stems from acoustic characteristics we did not measure.

The comparison of phonotactically legal to illegal conditions is a comparison of the five cluster conditions which would form possible codas even with their epenthetic stops fully present (/mt, ms, ns, ŋt, ŋs/) to the four conditions in which the epenthetic stop, if phonologically present, would violate a constraint (/mk, np, nk, ŋp/). However, because of the difficulty with the /nk/ cluster (Section 5.3), and the strong evidence that the rarity of responses in this condition is not related to the syllable structure constraint violation, we excluded the /nk/ cluster from the analysis, leaving a comparison of five legal conditions to three illegal ones. We used pairwise planned comparisons, with the discriminant as covariate, analyzing the phoneme monitoring and dictation responses separately. Fig. 11 shows the average number of subjects responding in phoneme monitoring, by final cluster, adjusted for the covariate.

The clusters in which the presence of the epenthetic stop would not violate any syllable structure constraints received significantly more responses than the clusters in which considering the epenthetic stop as a phonemic stop would result in an impossible cluster ($F(1,116) = 74.50$, $p < 0.001$). The same was true of the dictation responses ($F(1,116) = 56.76$, $p < 0.001$). (If the /nk/ cluster is included in the analysis, the differences are also significant: $F(1,116) = 165.44$, $p < 0.001$ for phoneme monitoring, $F(1,116) = 14.18$, $p < 0.001$ for dictation.) The difference was greater for epenthetic /t/ and /k/ than for epenthetic /p/, but it was significant overall. To provide further confirmation that the phonotactic legality effect is not due to production differences, we

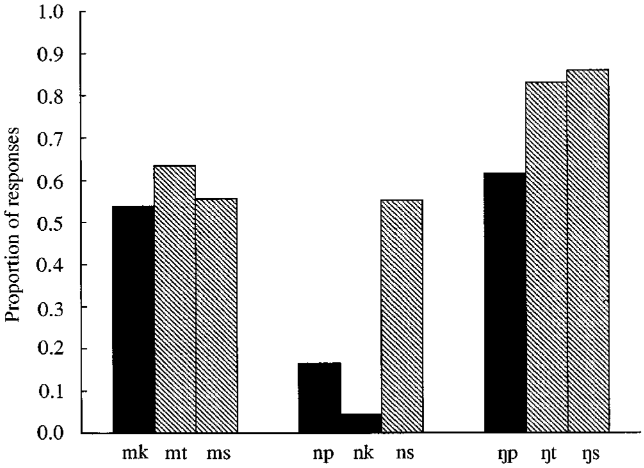

Figure 11. Average proportion of responses to each epenthetic stop for phonotactically legal and illegal final clusters, adjusted for the discriminant as the covariate. Phoneme monitoring data. ■ illegal; ▨ legal.

also performed the ANOVA with each acoustic measure as the covariate separately, and did this both for phoneme monitoring and dictation data, and with and without the /nk/ cluster. In all tests, the phonotactically legal conditions received significantly more responses than the illegal ones ($p < 0.001$ for each test). Thus, listeners are more likely to interpret an unintended epenthetic stop as an occurrence of the stop phoneme if doing so does not violate a syllable structure constraint in their language. This is true even when the influence of how many items contain epenthetic bursts, and how strong the acoustic cues for them are, is taken into account. Language-specific constraints affect how listeners interpret phonetic variability.

5.8. *Processing of epenthetic stops*

In the phoneme monitoring task, slower responses indicate greater processing load. Comparison of reaction times to epenthetic stops and to stops intended by the speaker can thus show whether epenthetic stops are more difficult to perceive. For this question, we analyzed only the data in which listeners did respond, and only data in which they responded to a stop which could be present in the stimulus either epenthetically or as the intended obstruent (e.g., responses from p-monitoring subjects to /mt, mk, ms, mp, np, ŋp/ items). We analyzed the responses for each stop monitored for separately, because so few t-monitoring subjects responded to any items in the /np, nk/ conditions that reaction times for these conditions could not be analyzed.

For each stop monitored for, we performed a planned comparison of items where the stop monitored for could occur epenthetically with items where the stop monitored for was the intended obstruent, and the nasal and intended stop were homorganic. Thus, for k-monitoring subjects, we compared the /ŋp, ŋt, ŋs/ conditions to the /ŋk/ condition. We did not include the heterorganic clusters with the stop as the intended obstruent, because Dutch listeners' reaction times to stops in heterorganic nasal-stop clusters are known to be slower than those to stops in assimilated clusters (Weber, in press). (Heterorganic nasal-stop clusters where the stop is not coronal are phonotactically impossible within the syllable in Dutch. Thus, /CVmk/ violates a phonotactic constraint even if no epenthetic stop is present, and this slows processing of the /k/.)

We performed all analyses both across subjects and across items. The average of a subject's reaction times to all items in the condition (across subjects), or the average of all subject's reaction times to a particular item (across items), was used when not all subjects responded to all items. The /np, nk/ conditions were excluded because so few subjects responded in these conditions. (Aside from the /np, nk/ clusters, all items received at least one response, and all subjects responded to at least one item in each condition.)

The average reaction times to epenthetic stops and to the same stop as an intended obstruent (in a homorganic cluster) are shown in Fig. 12. Reaction times to the stop as the intended obstruent in heterorganic clusters are shown for comparison although they were not included in the statistical tests. For each stop, both across subjects and across items, reaction times to epenthetic stops were significantly slower than reaction times to the same stop when intended by the speaker and in a homorganic cluster (p-monitoring: $t1 = 5.05$, $t2 = 8.28$; t-monitoring: $t1 = 5.73$, $t2 = 9.78$; k-monitoring: $t1 = 5.21$, $t2 = 9.36$; all tests significant at $p < 0.001$ using separate variance estimates). Reaction times to intended stops in heterorganic clusters are somewhat slower than to intended stops in homorganic clusters, but the difference between reaction times to epenthetic and intended stops is much greater.

Thus, listeners respond much more slowly to epenthetic stops than to stops which were intended by the speaker. Since this analysis uses only cases in which listeners did respond, these are cases in which listeners did interpret the epenthetic stop as a token of the target phoneme. However, they were slower to make this decision than they were with the stops which the speaker intended to produce. Thus epenthetic stops, even those which are perceived as occurrences of the stop phoneme, are more difficult for listeners to process than stops which are phonologically present for the speaker. Listeners probably

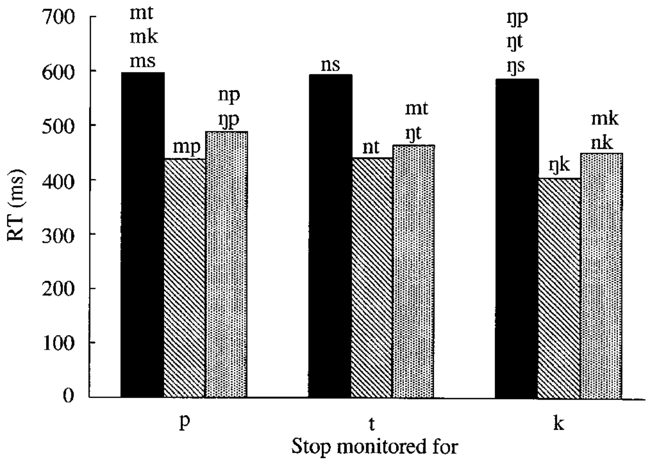

Figure 12. Average reaction times to epenthetic stops, the same stop when it is the intended obstruent and is in a homorganic cluster, and the same stop when it is the intended obstruent and is in a heterorganic cluster. ■ epenthetic; ▧ intended (homorganic); ▦ intended (heterorganic).

perceive the epenthetic stops to which they respond as being less good tokens of the target stop. The reaction time results show that epenthetic stops, when they are perceived, have a perceptual status intermediate between being normal segments (phonologically present) and not being present at all. Although a large proportion of subjects do perceive epenthetic stops as being tokens of the target phoneme in a large proportion of items (as was shown in Fig. 5), epenthetic stops are not processed in the same way as segments which are part of the intended segment string.

5.9. *The effect of task demands*

We turn now to the effect of different task demands on how epenthetic stops are perceived. The phoneme monitoring task puts subjects under time pressure, but in the dictation task, subjects knew that they had up to 16 s to complete their response, so they had ample time for processing the speech signal. The phoneme monitoring task also encourages subjects to pay more attention than in usual speech processing to low-level acoustic cues, since the only thing they have to concentrate on is whether they have heard the target phoneme or not. The dictation task, on the other hand, may encourage subjects to use their knowledge of orthographic patterns in their language, since they have to transcribe their responses. While orthography will not help them decide how to transcribe a given nonword item, it might guide them with regard to overall frequencies of particular clusters in the language. We predict that both the lack of time pressure and the reference to orthography would make subjects more likely to successfully recover the segment string intended by the speaker. Thus, we expect a lower number of responses to epenthetic stops in the dictation task. Furthermore, if subjects are more strongly guided by the statistical frequency of clusters in the language while performing the dictation task, the phonotactic legality effect might be stronger in the dictation task. Thus, the decrement in responses to epenthetic stops for the dictation task might be greater for conditions in which the epenthetic stop would violate a syllable structure constraint.

The number of responses to epenthetic stops, for each final cluster and each task, appears in Fig. 5. The overall average response rate to epenthetic stops was 53% for the phoneme monitoring task and 49% for the dictation task. We performed an analysis of variance, using only data in which subjects were monitoring for the stop which could occur epenthetically, with task as a within-subjects factor and final cluster (the nine final clusters where epenthetic stops are possible) as a between-subjects factor. In this test, the main effects of both task and final cluster were significant, but the interaction between the two was also significant (across both subjects and items, $F1\,(8, 207) = 5.32$, $F2\,(8, 117) = 13.27$, $p < 0.001$ for both).

We therefore analyzed the effect of task for each final cluster separately. For the clusters /ms, ns, ŋs/, the number of responses to epenthetic stops was significantly greater in phoneme monitoring than in dictation both across subjects and across items (/ms/: $F1\,(1, 23) = 6.72$, $p < 0.02$, $F2\,(1, 13) = 38.55$, $p < 0.001$; /ns/: $F1\,(1, 23) = 14.12$, $p < 0.01$, $F2\,(1, 13) = 32.08$, $p < 0.001$; /ŋs/: $F1\,(1, 23) = 4.80$, $p < 0.04$, $F2\,(1, 13) = 13.80$, $p < 0.01$). For the cluster /mk/, the effect reached significance only across items ($F1\,(1, 23) = 2.89$, $p < 0.11$, $F2\,(1, 13) = 8.90$, $p < 0.02$). Furthermore, the cluster /ŋp/ showed a reversal of the effect, with significantly fewer responses in phoneme monitoring ($F1\,(1, 23) = 16.12$, $p < 0.01$, $F2\,(1, 13) = 14.61$, $p < 0.01$). For the remaining clusters, there were more responses in phoneme monitoring than in dictation, but the effect did not reach significance across either subjects or items. Thus, the interaction between task and final cluster was significant because the effect is stronger for some clusters than others, and because the cluster /ŋp/ shows a reversal of the effect. For the /np, nk/ clusters, there may be a floor effect. In sum, the phoneme monitoring task did elicit more responses to epenthetic stops than the dictation task did, with the exception of the cluster /ŋp/.

Since eight of the nine clusters received more responses in phoneme monitoring than in dictation, it is safe to conclude that the effect does exist. When subjects have ample time for processing the speech signal, and are encouraged to make use of orthographic knowledge, they are more able to "factor out" the unintended phonetic variability of the epenthetic stop and recover the string of segments intended by the speaker. However, the difference between number of responses in the two tasks does not seem to be larger for clusters in which the epenthetic stop would violate a syllable structure constraint if it were interpreted as phonologically present. In fact, the three clusters which showed a reliable effect of task in the predicted direction, /ms, ns, ŋs/, are all clusters in which the epenthetic stop would violate no constraint. Thus, use of orthographic knowledge does not seem to strengthen listeners' tendency to report epenthetic stops less often if they would violate syllable structure constraints.

6. Conclusions

6.1. *The high rate of perception of epenthetic stops*

We have found that listeners perceive epenthetic stops approximately 50% of the time in nonwords which present the environment for epenthesis. This is a rather high rate of perception, considering that the speaker did not produce the epenthetic stops intentionally. This shows that epenthetic stops which are produced as a matter of variation in the timing of gestures are perceptible, and perceptible as tokens of the stop phoneme, a great deal of the time. This is true even when listeners have ample time for processing the signal, as in the dictation task. Historical change involving epenthetic stops is relatively rare and sporadic, though, as compared to sound changes such as assimilation which, when they apply, apply wherever their environment is met. For example, sound change has led to the insertion of a /p/ in the underlying representation of words like "empty" and "glimpse", but not in words with similar environments like "dreamt" and "teamster".[5] The overall high rate of listeners reporting epenthetic stops in our experiment suggests that the reason for the rarity of historical epenthesis lies in the conservative influence of orthography, or in the variability of epenthesis. Listeners hear a given word pronounced both with and without epenthesis, and with epenthetic stops of varying strength, so they may conclude that the epenthetic stop is not underlyingly part of the word. However, the rarity of resulting historical change is not due to the epenthetic stops being produced weakly or infrequently, as they are perceptible when previous knowledge about the item (orthographic and otherwise) is not available. Of course, the strength with which epenthetic stops are produced varies across languages and dialects (Fourakis & Port, 1986). However, even in languages where epenthetic stops are produced strongly and frequently, such as American English and Dutch, historical epenthesis is relatively rare.

[5] Of course, "dreamt" and "teamster" can often be produced with an epenthetic [p], and it is difficult to prove what the underlying representation of a given word is. However, we suspect that in careful pronunciation, [p] is far more common in "glimpse" than in "teamster", for example.

Our results also show that epenthetic stops are perceived as having the place of articulation which is predicted by the articulatory account of how they are produced. Although we have investigated some of the cues for whether an epenthetic stop is present or not, we have not investigated the cues for the place of the epenthetic stop. For example, we have not investigated whether the place information for the epenthetic stop is carried by the epenthetic burst or by the preceding nasal. However, wherever the place cues may be located in the signal, listeners almost never perceive an epenthetic stop with an incorrect place of articulation (a place not homorganic with the nasal).

6.2. *Perceptual cues and task effects*

We have investigated the contributions of several potential acoustic cues to the presence of epenthetic stops, by correlating the number of subjects responding to an item with the strength of that item on various acoustic measures. We found that acoustic characteristics both of the epenthetic stop itself and of the surrounding sounds correlate with how often listeners report hearing the epenthetic stop (although we did not test the perceptual contribution of each cue separately). The duration of the epenthetic burst and its degree of separation from the preceding nasal are cues, as well as the duration of the preceding nasal (negatively correlated), and for items with /p, t, k/ as the intended obstruent, the total stop closure duration. The power of the epenthetic burst may also play a role, but our results on this point are not clear, and a more accurate measure of power may be necessary.

Tests of these correlations for both phoneme monitoring and dictation data further indicate that the same perceptual cues play a role in both tasks. In the dictation task, the relationship of acoustic cues to perception is somewhat weaker than in the phoneme monitoring task. Listeners also perceive epenthetic stops as being tokens of the stop phoneme less often overall in the dictation task. Thus, they are more successful in recovering the string intended by the speaker when they are not under time pressure and when they are encouraged to use higher level knowledge such as knowledge of orthographic patterns in the language.

6.3. *Comparison of epenthetic and intended stops*

Although epenthetic stops were perceived rather often, they were perceived far less often than the stops intended by the speaker were. This is partly because the speaker may not have produced epenthetic stops in all items (and did not always produce epenthetic bursts), but even epenthetic stops which listeners responded to did not have the same perceptual status as intended stops. Listeners responded far more slowly to epenthetic stops than to the same stop when it was part of the intended segment string. Although they perceived the epenthetic stop as a token of the target phoneme, the epenthetic stops were more difficult to process. Fourakis & Port (1986) show that epenthetic stops are not phonetically identical to intended stops, even in an environment where the difference between presence and absence of a stop seems to be neutralized (/ns/ *vs.* /nts/, “prince” *vs.* “prints”). Even in this environment, speakers produce epenthetic stops with a shorter duration then they use for real stops. The difference in reaction times to epenthetic and intended stops shows that epenthetic stops are not only produced differently, they are also perceived differently.

6.4. *Perception of epenthetic* [*p*] *and* [*k*] *vs.* [*t*]

Overall, listeners are less likely to respond to epenthetic [t] than to epenthetic [p] or [k]. While this is partly due to differences in production, there is probably also a perceptual effect involved, as indicated by the data involving items with no epenthetic burst. If no epenthetic burst is produced in the environment for epenthetic [t], no epenthetic stop is perceived. However, if the environment for epenthetic [p] or [k] is present, listeners often perceive the signal as containing that stop even in the absence of an epenthetic burst. Furthermore, the one condition in which listeners routinely fail to perceive epenthetic stops which are produced with bursts is one with epenthetic [t], the /nk/ condition. Both the overall response rates and the special cases indicate that epenthetic [t] is less perceptible or less likely to be perceived as a real occurrence of the phoneme than epenthetic [p] and [k] are.

Given the widely acknowledged special phonological status of coronal consonants, one might infer that the low number of responses of epenthetic [t] is somehow related to this. Are listeners less likely to interpret epenthetic [t] as a real stop simply because it is coronal? Jun (1996) and Hume, Johnson, Seo & Tserdanelis (1999) suggest that phonological processes are less likely to change features which are perceptually salient, and relate perceptual salience to phonological markedness (although Hume *et al.* point out that other factors are also involved). Jun claims on the basis of arguments about perceptual cues, but no experimental results, that dorsal is the most perceptually salient place and coronal the least for unreleased stops. Thus coronals, the unmarked place, often assimilate to the place of articulation of a following consonant because they are less perceptually salient than other places. Perhaps epenthetic [t] is less likely to be perceived than other epenthetic stops for the same reason.

Jun (1996) only claims the perceptual salience ranking of dorsals, then labials, then coronals for unreleased stops. (The weakness of labial bursts, for example, was discussed in Section 5.1, and Hume *et al.* (1999) found only a small effect of place on salience for prevocalic stops.) In those items in our experiment which were produced without any epenthetic burst, any epenthetic stop would be equivalent to an unreleased stop.[6] Thus, for the items without epenthetic bursts, listeners' frequent responses for /p/ and /k/ but failure to respond for /t/ correlates well with Jun's (1996) perceptual salience ranking. If only cues for an unreleased stop are present, the dorsal and labial stops are perceived (despite the lack of an epenthetic burst), but the coronal stop is not. It would seem that the division by place among items without an epenthetic burst stems from differences in perceptual salience of unreleased stops at the various places of articulation, and is thus related to a perceptually based type of phonological markedness.

However, Lisker (1999, and older works cited therein) reports that the accuracy of perception of various unreleased stops is strongly affected by the quality of the preceding vowel. Furthermore, he reports well-established evidence that following most vowels, dorsal unreleased stops are the most likely to be misperceived, not the least. Thus, Jun's salience hierarchy is called into question. Our results showing that epenthetic /p, k/ are perceived even when no epenthetic burst is present, but epenthetic /t/ is not, bear

---

[6] The epenthetic stop could perhaps be released, but released with no burst, if there had not been time for oral air pressure to build up. However, the difference between an unreleased epenthetic stop and a silently released epenthetic stop cannot be great for the listener.

a similarity to the phonological data on place assimilation to CC clusters in Korean which Jun (1996) presents as an example, but his proposed perceptual salience ranking for unreleased stops cannot explain these data.

Since Hume *et al.* (1999) find that coronal place is only slightly less salient than other places when stops are released, low salience cannot be invoked for the extremely low rate of perception of epenthetic [t] when there is a burst. (Furthermore, place salience for stops released into vowels, which Hume *et al.* tested, cannot be readily transferred to released stops before another stop.) We propose that the low rate of responses to epenthetic [t] when there is an epenthetic burst derives from the combination of several factors which do not involve the special status of coronals. First, only one item in the /np/ condition has an epenthetic burst, and even that burst is weak. Second, although epenthetic bursts are widespread in the /ns/ condition, the homorganicity of /n/ and /s/ gives reason to expect epenthetic bursts to be weak in this environment, and the /ns/ stimuli do have shorter epenthetic bursts than most conditions (even shorter than the other nasal-/s/ conditions.) Thus, the low rate of perception of epenthetic [t] for /np, ns/ may be due solely to production factors. This leaves the /nk/ condition as the only case in which perception of epenthetic [t] is rare for reasons not related to production, and the most likely explanation here is related to orthography, rather than perception or phonological status. In sum, it appears as if listeners had a general bias against the perception of epenthetic [t] and toward the perception of epenthetic [p, k], but this is probably not related to a perceptual salience-based version of phonological markedness.

6.5. *Effects of language-specific phonological constraints on parsing of variability*

Our results show that listeners are more likely to perceive an epenthetic stop as real if doing so does not violate a syllable structure constraint. This perceptual effect is robust even when the differences in production among the conditions are accounted for. Thus, how Dutch listeners parse the variability of epenthetic stops is affected by language-specific syllable structure constraints.

There is extensive evidence that listeners' parsing of speech is affected by language-specific phonological constraints. For example, Dupoux, Kakehi, Hirose, Pallier & Mehler (submitted) show that perception of consonants with and without a vowel [u] separating them is different for Japanese and French listeners. Because Japanese allows very few consonant clusters, and vowels are epenthesized to break up consonant clusters when foreign words are borrowed, Japanese listeners parse even a signal from which the vowel [u] has been deleted as having that vowel present. French listeners can distinguish the signals with the [u] from those without the [u], however. Otake *et al.* (1996) show that Japanese listeners consider any moraic nasal, regardless of place, to be the same target in a phoneme monitoring task, whereas Dutch listeners do not. The moraic nasal is not specified for place in Japanese and always assimilates to the place of the following segment. Even when such phonological constraints are not absolute, they may influence listeners' parsing of the signal (Cutler, 1997). Listeners apply knowledge of what is typical in the language in parsing speech.

These and many other studies show that phonological patterns of a language, whether unviolated constraints or statistical tendencies, affect how listeners parse speech. However, most of this work addresses how listeners recognize words for which there is only one possible pronunciation in a particular environment. In the case of epenthetic stops, as with many types of deletion, lenition, and assimilation that occur in connected speech, a given word can appear in more than one form even in a particular environment. This study shows that language-specific phonological constraints affect not only how listeners parse strings which must appear in a particular form in their language, but also how listeners interpret the inherent phonetic variability of connected speech.

We are grateful to Anne Cutler, James McQueen, Roel Smits, Terry Nearey, and John Ohala for discussion of this material, and to Ignatius Mattingly and an anonymous reviewer for their helpful suggestions. We would also like to thank Jules Ellis for advice on statistical analysis, the two speakers for the experiment for their effort, and Niels Janssen for running the experiment. Any errors or misinterpretations are our own.

References

Ali, L., Daniloff, R. & Hammarberg, R. (1979) Intrusive stops in nasal-fricative clusters: an aerodynamic and acoustic investigation, *Phonetica*, **36,** 85–97.

Anderson, S. R. (1976) Nasal consonants and the internal structure of segments, *Language*, **52,** 326–344.

Baayen, H., Piepenbrock, R. & Rijn, H. van. (1993) *The CELEX Lexical database* (*CD-ROM*). Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Barnitz, J. G. (1974) Bloom-p-field, Chom-p-sky, and phonetic epen-t-thesis, *Studies in the Linguistic Sciences*, **4**(2), 1–13.

Blankenship, B. (1992) What TIMIT can tell us about epenthesis, *UCLA Working Papers in Phonetics*, **81,** 17–25.

Booij, G. (1995) *The phonology of Dutch*. Oxford: Clarendon Press.

Clements, G. N. (1987) Phonological feature representation and the description of intrusive stops. In *Proceedings of the parasession on autosegmental and metrical phonology*, Chicago Linguistic Society, pp. 29–51. Chicago: Chicago Linguistic Society.

Collins, B. & Mees, J. (1981) *The sounds of English and Dutch*. The Hague: Leiden University Press.

Connine, C. M. & Titone, D. (1997) Phoneme monitoring. In *A guide to spoken word recognition paradigms* (F. Grosjean & U. H. Frauenfelder, editors), pp. 635–645. Hove: Psychology Press.

Cutler, A. (1997) The comparative perspective on spoken-language processing, *Speech Communication*, **21,** 3–15.

Cutler, A. (1998) The recognition of spoken words with variable representations. In *Proceedings of the workshop sound patterns of spontaneous speech*, La Baume-les-Aix, 24–26 September, pp. 83–92.

Cutler, A., Dahan, D. & Donselaar, W. van. (1997) Prosody in the comprehension of spoken language: a literature review, *Language and Speech*, **40**(2), 141–201.

Cutler, A. & Otake, T. (1998) Assimilation of place in Japanese and Dutch. In *Proceedings of the international conference on spoken language processing*, Sydney, Australia, December.

Donselaar, W. van, Kuijpers, C. & Cutler, A. (1999) Facilitatory effects of vowel epenthesis on word processing in Dutch, *Journal of Memory and Language*, **41,** 59–77.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C. & Mehler, J. (1999) Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology*: *Human Perception and Performance*, **25**(6), 1568–1578.

Fourakis, M. & Port, R. (1986) Stop epenthesis in English, *Journal of Phonetics*, **14,** 197–221.

Frauenfelder, U. H. & Segui, J. (1989) Phoneme monitoring and lexical processing: evidence for associative context effects, *Memory & Cognition*, **17,** 134–140.

Fujimura, O., Macchi, M. J. & Streeter, L. A. (1978) Perception of stop consonants with conflicting transitional cues: a cross-linguistic study, *Language and Speech*, **21,** 337–346.

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.

Hume, E., Johnson, K., Seo, M. & Tserdanelis, G. (1999) A cross-linguistic study of stop place perception. In *Proceedings of the international congress of phonetic sciences*, San Francisco, August, pp. 2069–2072.

Jun, J. (1996) Place assimilation as the result of conflicting perceptual and articulatory constraints. In *The proceedings of the 14th West Coast conference on formal linguistics* (J. Camacho, L. Choueiri & M. Watanabe, editors), pp. 221–237. Stanford: CSLI.

Kakehi, K., Kato, K. & Kashino, M. (1996). Phoneme/syllable perception and the temporal structure of speech. In *Phonological structure and language processing* (T. Otake & A. Cutler, editors), pp. 126–143. Berlin: Mouton de Gruyter.

Keyser, S. J. & Stevens, K. N. (1994) Feature geometry and the vocal tract, *Phonology*, **11,** 207–236.

Kuijpers, C. & Donselaar, W. van. (Forthcoming) Phonological variation and phoneme identification in Dutch.

Lisker, L. (1999) Perceiving final voiceless stops without release: effects of preceding monophthongs versus nonmonophthongs, *Phonetica*, **56,** 44–55.

Marslen-Wilson, W. D., Nix, A. & Gaskell, M. G. (1998) Phonological variation in lexical access: abstractness, inference and English place assimilation, *Language and Cognitive Processes*, **10,** 285–308.

Murray, R. W. (1967) On epenthesis, *Folia Linguistica*, **23,** 293–316.

Norušis, M. J. & SPSS, Inc. (1994) *SPSS professional statistics 6.1.* Chicago: SPSS Inc.

Ohala, J. (1974) Experimental historical phonology. In *Historical linguistics* (J. M. Anderson & C. Jones, editors), pp. 353–389. Amsterdam: North-Holland Publishing Company.

Ohala, J. J. (1995) A probable case of clicks influencing the sound patterns of some European languages, *Phonetica*, **52,** 160–170.

Ohala, J. J. (1997) Emergent stops: diachronic and phonetic data. In *Proceedings of Seoul international conference on linguistics*, pp. 84–91. Seoul: Linguistic Society of Korea.

Otake, T., Yoneyama, K., Cutler, A. & van der Lugt, A. (1996) The representation of Japanese moraic nasals, *Journal of the Acoustical Society of America*, **100,** 3831–3842.

Picard, M. (1987) On the general properties of consonant epenthesis, *Canadian Journal of Linguistics*, **32**(2), 133–142.

Picard, M. (1989) Consonant epenthesis as a regular sound change, *Diachronica*, **6**(2), 223–236.

Piggott, G. L. & Singh, R. (1985) The phonology of epenthetic segments, *Canadian Journal of Linguistics*, **30,** 415–451.

Stevens, K. N. (1998) *Acoustic phonetics*. Cambridge: MIT Press.

Weber, A. (In press) Help or hindrance: how violation of different assimilation rules affects spoken-language processing, *Language and Speech.*

Wetzels, W. L. (1985) The historical phonology of intrusive stops: a non-linear description, *Canadian Journal of Linguistics*, **30,** 285–333.

Appendix A: Target materials

Targets with intended /p/:

flemp	flenp	fleŋp
flimp	flinp	fliŋp
flomp	flonp	floŋp
flymp	flynp	flyŋp
fremp	frenp	freŋp
frymp	frynp	fryŋp
xremp	xrenp	xreŋp
lymp	lynp	lyŋp
sxemp	sxenp	sxeŋp
sxymp	sxynp	sxyŋp
ʃamp	ʃanp	ʃaŋp
ʃimp	ʃinp	ʃiŋp
ʋymp	ʋynp	ʋyŋp
zymp	zynp	zyŋp

Targets with intended /t/:

flemt	flent	fleŋt
flimt	flint	fliŋt
flomt	flont	floŋt
flymt	flynt	flyŋt
framt	frant	fraŋt
frimt	frint	friŋt
xremt	xrent	xreŋt
lymt	lynt	lyŋt
sximt	sxint	sxiŋt
ʃamt	ʃant	ʃaŋt
ʃemt	ʃent	ʃeŋt
ʃomt	ʃont	ʃoŋt
ʋymt	ʋynt	ʋyŋt
zymt	zynt	zyŋt

Targets with intended /k/:

flemk	flenk	fleŋk
flomk	flonk	floŋk
flymk	flynk	flyŋk
frymk	frynk	fryŋk
xremk	xrenk	xreŋk
lemk	lenk	leŋk
lymk	lynk	lyŋk
sxamk	sxank	sxaŋk
ʃemk	ʃenk	ʃeŋk
ʃimk	ʃink	ʃiŋk
ʃomk	ʃonk	ʃoŋk
ʋymk	ʋynk	ʋyŋk
zemk	zenk	zeŋk
zymk	zynk	zyŋk

Targets with intended /s/:

fams	fans	faŋs
flyms	flyns	flyŋs
frems	frens	freŋs
frims	frins	friŋs
fryms	fryns	fryŋs
lyms	lyns	lyŋs
sxems	sxens	sxeŋs
ʃems	ʃens	ʃeŋs
vlems	vlens	vleŋs
vlims	vlins	vliŋs
vloms	vlons	vloŋs
ʋrems	ʋrens	ʋreŋs
ʋyms	ʋyns	ʋyŋs
zyms	zyns	zyŋs