To appear in K. Flack and S. Kawahara eds. 2005. *University of Massachusetts Occasional Papers in Linguistics 31: Papers in Phonetics/Laboratory Phonology.*

The Perception of Epenthetic Stops in English:
The Effects of Cluster Type and Silent Interval Duration[*]

Takahito Shinya

University of Massachusetts Amherst

1. Introduction

Speech abounds with different kinds of variability. Generally, there are two types: phonological and phonetic variability. A speech sound may be phonologically altered in one way or another (assimilation or dissimilation), dropped (deletion) or inserted (epenthesis) in various circumstances. Such phonological variability is always predictable since underlying forms are recoverable from their contexts. Phonetic variability, on the other hand, involves variation that is not predictable. For example, English word-final nasals may, but not always, assimilate to the place of the following initial consonant (Gaskell and Marslen-Wilson 1996). Dutch has an optional epenthetic schwa as in [tylp]~[tyləp] (Donselaar, Kuijpers and Culter 1999). In this paper we address the latter type of variability.

This paper deals with epenthetic stops that are inserted between nasals or laterals and voiceless obstruents in English. Some examples are given in (1a), contrasting with underlying stops in similar or the same environments (1b). Epenthesized stops are indicated by superscripted symbols.

(1) Epenthetic vs underlying stops in English
a. epenthetic b. underlying
[pɹɪn^{t}s] prince [pɹɪnts] prints

---

[*] I thank John Kingston, Lyn Frazier, and Tom Roeper for their advice on this study. Also thanks to the editors of this volume Kathryn Flack and Shigeto Kawahara for their comments on an earlier draft. All remaining errors and shortcomings are, of course, my own.

[fɔl^{t}s]	false	[fɔlts]	faults
[hæm^{p}stə˞]	hamster	[dʌmpstə˞]	dumpster
[dɹɛm^{p}t]	dreamt	[tɛmpt]	tempt, temped
[jʌŋ^{k}stə]	youngster	[pɹæŋkstə˞]	prankster

The phenomenon of stop epenthesis is well-known in English; epenthesized stops apparently neutralize the underlying distinction between a sonorant followed by an obstruent (e.g. /ns/) and a sonorant followed by a stop followed by an obstruent (e.g. /nts/); both of these forms have the identical surface forms (e.g. [nts]). Therefore, words such as *prince* vs. *prints* and *false* vs. *faults* may not be distinguished.

This study focuses on perceptual aspects of epenthetic stops in English. We ask the following three questions:

(2) a. Is it phonological constraint or cluster frequency (or both) that is at play in the perception of epenthetic stops?
b. Under what circumstances do epenthetic stops (or more generally segments) are perceived?
c. In identifying a stop, how are the acoustic properties of the adjacent segments used?

We see each of these questions in turn below.

The primary question we ask is whether or not phonological constraints influence the perception of epenthetic stops. One suggestion made by Warner and Weber (2001), who experimentally examined the perception of epenthetic stops in Dutch, is that phonotactics are an influential factor. They found that stops are less likely to be perceived as underlying between nasals that are followed by obstruents when doing so violates a phonotactic constraint of the language (e.g. more /t/ percepts in /ns/ than in /np/, where /nts/ is legal but */ntp/ is not). In a related study (Warner and Weber 2002), they present statistical data on cluster frequency and suggest that frequency may also affect the perception of epenthetic stops, such that they may be perceived more in frequent clusters than in infrequent ones. Thus, the specific question that this paper addresses is: is it phonological knowledge or frequency that is at play in the perception of epenthetic stops?

A related question in this paper is under what environment segments are perceived. In other words, we ask what acoustic properties would serve as the cues for a segment percept. Some studies, e.g. Ali, Daniloff and Hammarberg (1979) and Warner and Weber (2001), point out that in the perception of naturally produced nasal-stop/fricative clusters, listeners often hear a stop even if there is no acoustic burst between the two consonants. This suggests that a burst is not a necessary condition for perceiving a stop. Some other acoustic properties should play a role. One such property would be silent gaps between sonorants and obstruents. In the experiment reported in this paper, we examine how a silent interval of varying duration affects the perception of stops between different sonorants and fricatives.

Another question addressed in this paper is, when one hears a stop where there is only a silent interval, how would one use the acoustic properties of the adjacent segments to identify this stop? We mentioned above Clements' (1987) claim that an epenthetic stop between a sonorant and a fricative shares its [place] feature with the preceding sonorant and its [voice] feature with the following fricative. However, bursts and formant transitions which cue epenthetic stops are often present in natural speech, which hinders us from investigating listeners' use of the acoustic properties of the adjacent segments themselves. Below I report a perception experiment in which only a silent interval of varying duration between sonorants and voiceless fricatives induces stop percepts. Thus, the acoustic property that cues the stops was only the silent interval, with no burst or formant transition contained in the stimuli. Therefore, the only information that can be used to identify the stop is the acoustic properties of the adjacent sonorant and fricative.

We report a perceptual experiment that examines the effect of cluster type and silent interval duration. We test several cluster types each of which consists of a sonorant and a fricative. A silent interval of varying duration is inserted between the sonorant and the fricative. We show that although phonological constraints and cluster frequency both do not explain the obtained data quite satisfactorily, the phonology account is superior to the frequency account for explaining one cluster type. We also show that silent interval plays an important role in the perception of epenthetic stops, which suggests that a stop burst is not a necessary condition to perceive an epenthetic stop. Furthermore, we see that the featural composition of an epenthetic stop does not always follow what Clements (1987) claims. Specifically, we see a case in which a perceived epenthetic stop shares its place feature with the following obstruent and not with the preceding sonorant.

In the rest of the paper, We first review the previous studies on epenthetic stops (§2) and then describe two different general approaches that have been proposed in the literature to account for the phonotactic effect found by Warner and Weber (2001): one based on phonological knowledge on syllable structure constraints and one based on cluster frequency (§3). The report of the experiment follows in §4. §5 concludes the paper.

2. Previous studies on epenthetic stops

Epenthetic stops in English have a number of characteristics. Clements (1987) argues that an epenthetic stop always shares the [voice] feature with the following obstruent and the [place] feature with the previous sonorant. Epenthetic stops normally arise when the obstruent is voiceless, but, though much less commonly, they also appear when it is voiced (e.g. /nz/).(Fourakis and Port 1986).

Epenthetic stops show another interesting property. Jones (1966) and Fourakis and Port (1986) claim that stop epenthesis in English is dialect-specific. Jones (1966), based on his impressionistic observations, points out that the occurrence of an epenthetic stop between a sonorant and a fricative is not characteristic of British English. Fourakis and Port (1986) experimentally showed that epenthetic stops occur in American English but never in South African English.

While stop epenthesis has been shown to be dialect-specific, there is also an opposing idea that the phenomenon is governed by articulatory constraints and is universal to all languages. Ohala (1974) suggests that in a nasal-obstruent sequence a stop is produced unintentionally in the articulatory transition from the nasal to the following obstruent, due to mistiming of the closure of the velic port and the oral release of the nasal. Specifically, if the raising of the velum for the velic closure for the following obstruent occurs before (instead of simultaneously with) the oral release at the offset of the nasal, the oral cavity is sealed, and a burst will arise when the oral closure for the nasal is released. Ohala (1974) also suggests that the occurrence of stops in lateral-fricative clusters is explained in terms of universal constraints on articulation. Based on palatographic data, he points out that the contact areas for /l/ and /s/ are to a certain degree complementary. He says that "in moving from an [l] to an [s], contact and release of contact must be made simultaneously in these complementary areas" and that "to the extent that they are not simultaneous, complete contact all around the alveolar ridge may result and thus complete stoppage of the air, that is, a [t] will result" (Ohala 1974: p. 359).

Ali, Daniloff, and Hammarberg (1979) is the first study that systematically examined nasal-fricative clusters in English, using oral and nasal air flow and oral air pressure data. They found that what happens first in producing an epenthetic stop is cessation of voicing during the nasal, followed by cessation of the nasal airflow. Because voicing and nasal airflow cease, the oral air pressure increases and reaches its maximum toward the offset of the nasal, while the oral closure is maintained. When the oral closure for the nasal is released, a burst appears due to the abrupt emission of the highly pressured air from the oral cavity.

Ali et al. (1979) also provide preliminary perception data obtained by auditory judgments from 3 listeners. They report a finding that some /ns/ clusters elicited epenthetic stop perception even without a burst or silent gap being present. They also note that the presence of a syllable or morphological boundary between the nasal and the fricative does not elicit perception of epenthetic stops. This is consistent with the results obtained by Warner and Weber (2001), which we review below.

Fourakis and Port (1986) did a production experiment on the English /ns/, /nz/, /ls/ and /lz/ clusters in monosyllabic word-final position. They analyzed data on American English (midwestern) and South African English, and claimed that stop epenthesis is a dialect-specific phenomenon that is observed only in American English. They found that, in American English, when the fricative was voiceless there was always an interval of silence lasting more than 10 ms between the sonorant and fricative, which they judged as the occurrence of an epenthetic stop. When the fricative was voiced, epenthetic stops were still observed, but far less often than when the fricative was voiceless. No such silence was seen for South African English in either of the conditions. Based on these findings, Fourakis and Port argue against the claim that epenthetic stops are caused by universal constraints on articulation and aerodynamics.

Fourakis and Port's (1986) claim that stop epenthesis is dialect-specific appears to be a bit too strong, because they base their claim on a purely acoustic parameter: the presence of silent intervals of more than a certain duration. Though there is a systematic difference between American English and South African English in articulating sonorant-obstruent clusters, we need perceptual evidence that a 10 ms silent interval in

/ns/ and /ls/ is always identified as a stop in order to make their claim valid.

Fourakis and Port (1986) also found a systematic temporal difference between clusters with epenthetic stops and clusters with underlying stops. They show that when the stop is underlying as in /nts/ or /ndz/, the oral closure duration is longer and the vowel and nasal are both shorter than when the stop is epenthetic. When the sonorant is /l/ as in *false* and *faults* the vocalic portion /ɔl/ is shorter when it is followed by the underlying /t/ than by the epenthetic [t]. This finding is important because it raises an interesting question of whether listeners use these acoustic differences as cues for distinguishing between epenthetic and underlying stops.

However, the results of the subsequent studies on closure duration of epenthetic stops are mixed. In her study of the English /ns/ cluster using the TIMIT database (see Keating, Blankenship, Byrd, Flemming, and Todaka 1992 for the description of the database), Blankenship (1992) showed that epenthesis occurs in about one-fourth of the /ns/ strings of American English. Though she found that epenthetic stops were shorter than underlying stops, the difference was not statistically significant.

Yoo and Blankenship (2003) recently followed up these results with an independent production experiment. They examined the difference in stop closure duration between the /ns/ and /nts/ clusters in American English appearing in different positions in a syllable and in different stress environments. They confirmed Blankenship's (1992) finding that no durational difference was found between underlying and epenthetic stops, with one exception. They found a statistically significant difference in the environment where the /ns/ cluster was separated by a syllable boundary and the following vowel was stressed, as in *co**n**s**ent*** vs. *blu**nt** **s**aying.* The closure durations in the /ns/ and /nts/ clusters did not differ significantly in word-final position (e.g. *inte**ns**e* vs. *inte**nts***), contra Fourakis and Port's (1986) result. Also, it turned out that stress does not affect the closure duration, since Yoo and Blankenship (2003) did not find statistically significant effects of stress.

Yoo and Blankenship (2003) also reanalyzed the occurrence of the /ns/ cluster using the TIMIT database with respect to its position in a word and stress, and obtained a result that is different from Blankenship's (1992). They found that the closure duration was significantly shorter for the /ns/ cluster than for the /nts/ cluster in any position and in any stress condition.[1] The issue still remains to be settled.

Warner and Weber (2001) experimentally investigated the perceptual characteristics of the epenthetic stops that appear between nasals /m, n, ŋ/ and voiceless stops /p, t, k/ or the fricative /s/ in Dutch. They recorded two speakers who naturally produced monosyllabic nonwords containing all of the clusters obtained by combining the nasals and the obstruents in word-final position. The speakers were explicitly instructed to produce nasals without assimilating to the places of the following obstruents. Warner and Weber first analyzed the production data acoustically. The data show considerably different patterns with respect to the presence of burst depending on the

---

[1] Yoo and Blankenship (2003) do not explain why this discrepancy was found. The reason is not at all clear.

cluster type.[2] Specifically, very few bursts were observed in the clusters such as /np/ and /ŋp/, while relatively high rates of epenthetic burst occurrences were seen in the other clusters. Warner and Weber offer an articulatory speculation. They suggest that in a cluster with the nasal posterior to the intended stop like /np/ or /ŋp/, the oral closure for the stop may often be made before the release of the oral closure for the nasal, masking the latter articulatory gesture. Therefore, no burst results.

Warner and Weber (2001) presented their listeners with the recorded materials whose acoustic characteristics have just been described above in a phoneme monitoring task (see Connie and Titone, 1996 for a review of the method). The listeners were asked to monitor the kind of epenthetic stop that is expected in the particular cluster. For example, they were required to monitor /p/ if the stimulus was /zymt/. Warner and Weber found that listeners perceived epenthetic stops about 50% of the time. They also showed that listeners tend to miss an epenthetic stop more often if it would form a phonotactically illegal cluster with the flanking consonants than if it would not. In Dutch, heterorganic two-consonant clusters are illegal, except when the second one is coronal. Thus, in combining the nasals /m, n, ŋ/ and the voiceless stops /p, t, k, s/, only /mt, ms, ns, ŋt, ŋs/[3] are legal while the others /mk, np, nk, ŋp/ are illegal. In three-consonant clusters, only clusters with two homorganic consonants followed by a coronal are legal. Thus, the same legal and illegal sets of clusters seen in two-consonant clusters can be extended to three-consonant clusters: /mpt, mps, nts, ŋkt, ŋkt/ are legal whereas the others /mpk, ntp, ntk, ŋkp/ are illegal. Warner and Weber's listeners perceived more epenthetic stops when they heard the legal clusters such as /mt/ and /ŋs/ than the illegal ones such as /mk/ and /nk/, because /mpt/ and /ŋks/ are legal in Dutch but /mpk/ and /ntk/ are not. Furthermore, they found that listeners were less likely to respond to epenthetic [t] than to epenthetic [p] or [k]. Interestingly, in the environments where an epenthetic [p] or [k] appears, listeners often heard a stop even when a burst was not present, but in those where an epenthetic [t] appears, they did not hear a stop unless there was a burst in the signal. Warner and Weber first relate this effect to the widely acknowledged special status of coronals. However, based on the experimental result in Hume, Johnson, Seo, and Tserdanelis (1999) that coronal place is only slightly less salient than other places when stops are released, they conclude that the low rate of perception of epenthetic /t/ in their experiment derives from the combination of several factors which do not involve the special status of coronals.

---

[2] Warner and Weber (2001) adopted the criterion of whether an acoustic burst appears or not, rather than whether there is a silent interval of a certain duration or not, in determining the presence of an epenthetic stop. They suggest that "it is not always possible to determine from acoustic information that a silent gap is a reflection of an epenthetic stop" and "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" (p. 61). In this case a silent gap is produced by an articulatory mechanism that is different from the one that causes a gap in clusters like /mt/, where a gap is produced by the cessation of voicing while the oral closure of the nasal is being made.

[3] Clusters /mp, nt, ŋk/ are omitted. Though they are legal in Dutch, no epenthetic stop is expected in any of them since the two consonants are homorganic, which means that articulatory contact holds throughout the clusters.

They attribute the effect to factors such as orthographic and production influences.

In their subsequent study, Warner and Weber (2002) investigated the effect of a syllable boundary on the perception of epenthetic stops in Dutch. They replicated Ali et al.'s (1979) result that listeners hear fewer epenthetic stops when the nasal and the following obstruent are separated by a syllable boundary. Furthermore, Warner and Weber (2002) conducted a statistical survey on the frequency of the nasal-stop/fricative clusters in Dutch and English, using the CELEX database (Baayen, Piepenbrock, and van Rijin 1993). They found that, both within a syllable and between syllables, clusters in which epenthetic stops potentially arise are far more frequent in Dutch than in English, and suggest that epenthetic stops are predicted to be observed more in Dutch than in English.

We have seen above that one of the major findings of Warner and Weber's (2001) study is the phonotactic effect. They note that "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." (p.78). In this view, the phonotactic effect is accounted for by assuming phonological knowledge in the form of abstract rules or constraints that prohibit particular sequences of phonemes (Moreton 2002, Moreton and Amano 1999). For example, the illegality of the cluster */ŋkp/ in Dutch, as opposed to /ŋkt/, can be explained in such a way that there is a constraint that bans segments that have noncoronal places of articulation in the *syllable appendix*, the subsyllabic unit that does not belong to the rhyme but is directly adjoined to the syllable node (Booij 1995).

However, I argue that Warner and Weber's (2001) claim is invalid. In Dutch, it is always the case, at least with the clusters that Warner and Weber tested, that when a two consonant cluster is illegal (e.g. */mk/) the corresponding three consonant cluster is also illegal (e.g. */mpk/). Similarly, when a two consonant cluster is legal (e.g. /mt/), then the corresponding three consonant cluster is legal too (e.g. /mpt/). Thus no perceptual bias is expected toward one of the two clusters if both are illegal (or legal). The results obtained by Warner and Weber (2001) must be accounted for by some other principles. The phonotactic effect would be observed with respect to epenthetic stops when only one of the clusters is illegal. In that case, listeners would be biased toward the legal clusters.

Even in such cases, however, an alternative account based on cluster frequency is possible. There is large body of work in the literature suggesting that frequency affects the perception of language-specific phonotactic patterns (Hay, Pierrehumbert, and Beckman 2004, Massaro and Cohen 1983, Pitt and McQueen 1998, McClelland and Elman 1986, Vitevich, Luce, Charles-Luce, and Kemmerer 1997, Vitevich and Luce 1999, among others). The specific claims differ among the authors, but, generally speaking, in a frequency-based account phonotactics are attributed to the listener's different sensitivity to different segment sequences; this sensitivity is established by his/her linguistic experience. Thus, if a putative phonotactic effect was observed in a pair of consonant clusters such that the perceptual bias was toward the legal cluster, one could say that such effect was obtained because the frequency of the other cluster is zero, not because it is phonotactically illegal.

In order to investigate this claim, what we need is a pair of clusters where only one is phonotactically legal and their frequencies are not compatible with the phonotactic pattern. Phonotactics and frequency are usually correlated because only legal clusters occur. However, the perceptual experiment reported in §4 will show that there is at least one case which a phonology-based account can account for but a frequency-based account cannot. Specifically, we will see that in listening to /lf/ cluster listeners invariably reported epenthetic /p/ between /l/ and /f/ rather than epenthetic /t/. We argue that a phonology-based account can explain this fact while a frequency-based account cannot, since the frequencies of /lpf/ and /ltf/ are both close to zero.

3. Phonology-based and frequency-based accounts on "phonotactic effect"

3.1 Predictions of phonology-based account

Before reporting the experiment, let us look at the predictions that phonology-based and frequency-based accounts would make about its results. As described below in detail, we compare the following six continua[4] made from word-final clusters in (3). We put silent intervals of varying durations between the sonorants and the fricatives to induce listeners to hear epenthetic stops. Silent intervals are indicated by underscores.

(3) a. /dɛlf/ – /dɛl_f/
b. /dɛlθ/ – /dɛl_θ/
c. /dɛls/ – /dɛl_s/
d. /dɛmf/ – /dɛm_f/
e. /dɛmθ/ – /dɛm_θ/
f. /dɛms/ – /dɛm_s/

Assuming Clements' (1987) claim, the epenthetic stops heard between the sonorants (/l/ or /m/) and the fricatives (/f/, /s/ or /θ/) are predicted to be voiceless and homorganic to the preceding sonorants. Thus, listeners are will hear /t/ after /l/ and /p/ after /m/. We now have clusters like the ones in (4):

(4) a. /dɛlf/ – /dɛ**ltf**/
b. /dɛlθ/ – /dɛ**ltθ**/
c. /dɛls/ – /dɛlts/
d. /dɛmf/ – /dɛ**mpf**/
e. /dɛmθ/ – /dɛ**mpθ**/
f. /dɛ**ms**/ – /dɛmps/

The bolded pairs of clusters (/ltf/, /ltθ/, /mpf/, /mpθ/ and /ms/) are unattested in English (Cruttenden 2001). However, in some generative phonological theories of English syllable structure, /ltf/ is considered ungrammatical while the other four clusters are grammatical. The rationale is as follows. First, it is often claimed that English syllable codas have two segmental slots (Fudge 1968, Kiparsky 1981, Clements and Keyser 1983, Giegerich 1992, Kenstowicz 1994). When there are more than two consonants after the syllable nucleus, all of the consonants after the second one fall into the *syllable appendix*. Given that constraint, the two consonants in the cluster /ms/ both fall into the coda with

---

[4] Hereafter I use /lf/, /lθ/, /ls/, /mf/, /mθ/ and /ms/ to refer to each of the continua.

no appendix consonant. There is no apparent restriction that bans the /ms/ cluster: a voiced labial followed by a voiceless coronal (e.g. *dreamt*) is legal. The absence of this cluster is probably an accidental gap. Therefore, this cluster should be equivalent to the other attested ones in (4). There is also a widely acknowledged constraint that restricts the consonants in the syllable appendix to coronal obstruents (Blevins 1995). One way of formally representing this would be to posit a constraint of the form given in (5):

(5) Syllable appendix constraint[5]

$$\begin{array}{c} *\ \ ]_{\sigma}C \\ | \\ \text{[noncoronal]} \end{array}$$

The clusters /ltθ/ and /mpθ/ satisfy this constraint because /θ/ is a coronal consonant in the appendix (note that /lt/ and /mp/ are in the coda, not in the appendix), hence they are legal English clusters. The other two clusters /ltf/ and /mpf/ violate this constraint. However, as will be described below, only /ltf/ is illegal. The key idea is the *linking constraint*, which requires association lines to be interpreted as exhaustive (Hayes 1986, Itô 1986, 1989, also see Goldsmith 1990 and Itô and Mester 1993 for similar ideas). Under this idea the appendix constraint in (5) is violated if a noncoronal feature is linked exhaustively to the appendix consonant(s). Given this idea, /ltf/ violates the exhaustivity condition but /mpf/ does not, as illustrated in (6a) and (6b):

(6) a. $$\begin{array}{c} \text{mp}]_{\sigma}\text{f} \\ |\diagup \\ \text{[labial]} \end{array}$$ b. $$\begin{array}{c} *\ \text{lt}\ ]_{\sigma}\ \ \text{f} \\ |\quad\ \ | \\ \text{[coronal][labial]} \end{array}$$

In /mpf/, the noncoronal [labial] feature is associated to the /f/ *as well as* the /p/ consonant in the coda, whereas in /ltf/ the labial feature is linked exhaustively to the appendix consonant /f/, which violates the constraint in (5).

Since the /ltf/ cluster is phonologically illicit, we predict that listeners will perceive more /lf/ than /ltf/, ignoring the silent interval inserted between /l/ and /f/. Except for this cluster, the phonology-based account does not predict any perceptual differences among the other clusters. They are in principle all equivalent to one another.

The phonology-based account makes another prediction that is based on the markedness of labials. That is, the /m/-initial clusters will induce more epenthetic stops than the /l/-initial clusters. It is widely known that coronals show phonologically unmarked behaviors compared to labials and velars (Paradis and Prunet, 1991). Moreover, Hume, Johnson, Seo, and Tserdanelis (1999) and Jun (1995, 1996) show that this asymmetry between coronals on the one hand and labials and velars on the other is also seen in speech perception such that labials and velars are perceptually more salient than

---

[5] This constraint does not intend to explain everything about the syllable appendix consonants in English. For example, more than two consonants may be contained in the appendix, as in /glɪmpst/ ‘glimpsed’.

coronals.[6] Given the assumption that /p/ is perceived in the /m/-initial clusters and /t/ in the /l/-initial clusters, it is expected that epenthetic stops are more likely to be perceived in the /m/-initial clusters than in the /l/-initial ones, because the labiality of /m/ would give listeners a stronger cue for /p/ than the coronality of /l/ would.

Based on the discussions above, we come to two predictions. One is that /lf/ will induce less epenthetic stops than the other clusters, which are predicted not to differ with one another in terms of the likelihood of perceiving epenthetic stops. The other is that /mf/, /mθ/ and /ms/ will induce more epenthetic stops than /lθ/, /ls/ and /lf/. These predictions are illustrated in (7) below:

(7) Prediction made by phonology-based account
/mf/ ≈ /mθ/ ≈ /ms/ > /lθ/ ≈ /ls/ > /lf/
(ordered from high to low likelihood of perceiving epenthetic stops)

3.2 Predictions of frequency-based account

Frequency has to do with how many times one encounters a particular lexical item or a strings of segments. In listening to ambiguous stimuli listeners are biased toward hearing frequent words or strings of segments rather than infrequent ones, because frequent items have greater strength in memory (Bybee 1998).

When considering frequency we need to distinguish two types: token frequency and type frequency. Token frequency refers to the frequency of individual items. For our current purposes, the items are the word-final consonant clusters in (3). When calculating in token frequency, each occurrence of a cluster in a corpus counts, whether it occurs in the same word or not. Type frequency is the number of items that appear in different words. All occurrences of a certain cluster in a given word count as one occurrence, regardless of how many instances of that word are observed.

Both types of frequency information are relevant in speech perception. Token frequency is important in studies that investigate the effects of so called “word frequency”. It is claimed that frequent words induces greater lexical activation and more resistance to change (Bybee 1998). Type frequency is claimed to be important in accounting for the “lexical neighborhood effect”, where the lexical neighborhood of an existing word or a nonsense form is defined by the set of words which differ from it by a single phoneme (Luce, Pisoni, and Goldinger 1990). Luce and his colleagues demonstrate that the lexical neighborhood density of a word has a strong influence on word perception such that it takes more time to recognize a word with high lexical neighborhood density than one with low lexical neighborhood density, because there are more “competitors” for the word with high lexical neighborhood density.[7]

Since both token and type frequencies have been shown to be relevant in speech

---

[6] As mentioned above, the effect that Hume et al. (1999) found was small. However, it was statistically reliable, which means that the effect is not negligible.

[7] Token frequency matters in calculating neighborhood density too, as neighbors are weighted by their token frequencies.

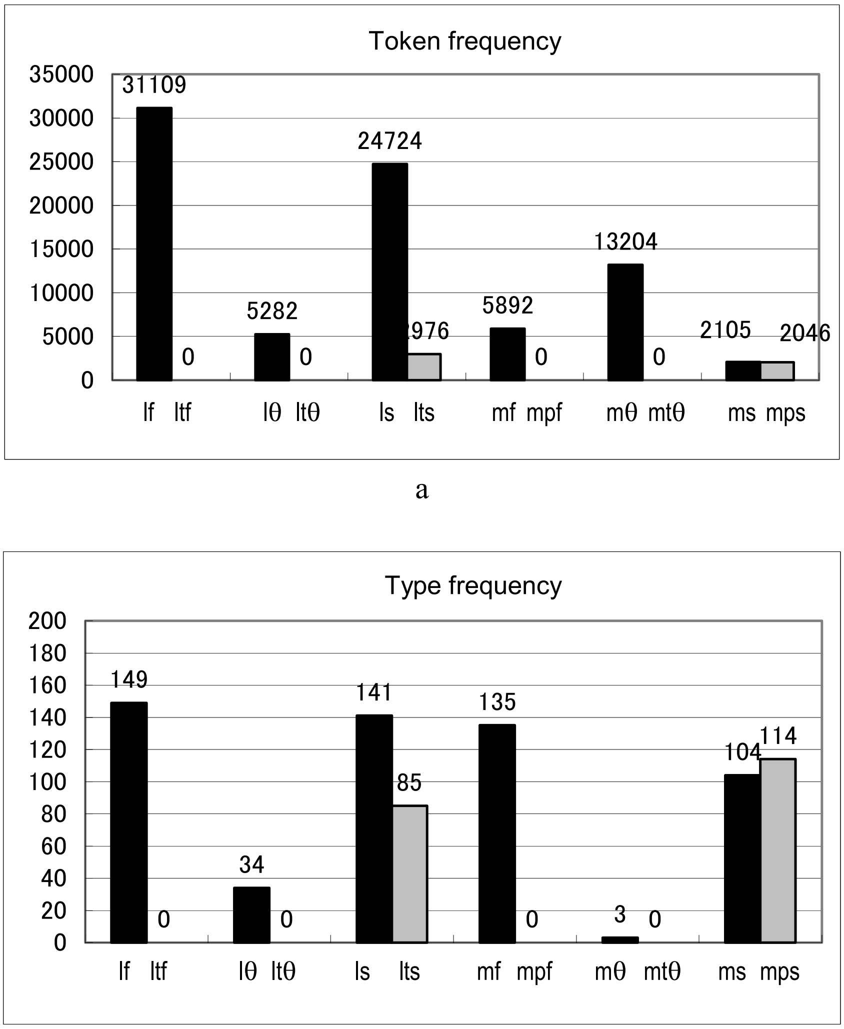


Figure 1 Frequencies of the sonorant-fricative clusters and the corresponding three-consonant clusters where epenthetic stops are between the two consonants. The clusters words in the CELEX database of English (Baayen et al. 1993).

perception, we should consider both kinds of frequency information here. The two panels in Figure 1 show the token and type frequencies of the clusters in (4), based on the CELEX database (Baayen et al. 1993). The data are the combined frequencies of the clusters occurring within a syllable (e.g. /fɔls/ *false*) and across syllables (/maɪl.ston/ *milestone*). This is justified by the fact that epenthetic stops occur in both environments, though, as pointed out above, they occur less often across syllables (Warner and Weber 2002).

The magnitude of the frequency difference between the paired clusters with and without epenthesis would tell us how strongly listeners would be biased toward one of the clusters. That is, it is predicted that listeners would be biased toward the more frequent cluster in each pair, and that this bias would be greater when the frequency difference between the clusters is larger. To begin with, in both types of frequency data, three-consonant clusters never occur in the cluster pairs /lf/-/ltf/, /lθ/-/ltθ/, /mf/-/mpf/ and /mθ/-/mpθ/ while the two-consonant clusters in the same pairs occur at least occasionally. It is predicted from this fact that listeners will be biased toward the two-consonant clusters in these pairs rather than the corresponding three-consonant clusters, and, as a consequence, they will not hear many epenthetic stops in these environments. On the other hand, both two- and three-consonant clusters occur with frequency in the pairs in /ls/-/lts/ and /ms/-/mps/[8], which suggests that epenthetic stops are more likely to be perceived in those contexts than in those in which one of the clusters never occurs.

In the token frequency data, the /lf/-/ltf/ pair is predicted to show the strongest bias toward /lf/, since the difference between the frequency of /lf/ and that of /ltf/ is the largest among the six cluster pairs. Next strongest is the /ls/-/lts/ pair, and then the /mθ/-/mpθ/ pair. In these pairs, the bias would also be toward to the two consonant clusters. The /lθ/-/ltθ/ and /mf/-/mpf/ cluster pairs would be more or less the same, predicting weaker bias toward /lθ/ and /mf/, respectively. Finally, there would be no bias in the /ms/-/mps/ pair because their frequencies are quite close. Therefore, based on the token frequency data, the order that shows the strength of bias is /ms/ > /mf/ ≈ /lθ/ > /mθ/ > /ls/ > /lf/. For ease of understanding the clusters are arranged from the highest to the lowest likelihood of perceiving epenthetic stops.

The type frequency data show somewhat different patterns, especially for the /m/-initial clusters. The frequency of /mf/ is relatively high and that of /mθ/ is relatively low, which makes the frequency difference of the /mf/-/mpf/ pair large and that of the /mθ/-/mpθ/ pair small with respect to the other cluster pairs. The frequencies of /ms/ and /mps/ are both fairly high, and there would be no bias for either of the two clusters.[9] When we order the likelihood of perceiving epenthetic stops among the six cluster pairs on the basis of the type frequency data, they would be arranged /ms/ ≈ /mθ/ > /lθ/ ≈ /ls/ > /mf/ ≈ /lf/. Some bias against hearing epenthetic stops is predicted in all of the cluster pairs except /mθ/-/mpθ/ and /ms/-/mps/.

The two relative orderings obtained from the token and type frequency data are summarized in (8):

---

[8] That the /ms/ cluster shows certain frequency comes from the fact that the data contain clusters that occur across a syllable boundary. This cluster is impossible within a syllable but possible across syllables as in /tʃɔ*ms*kɪ/ *Chomsky*.

[9] The type frequency of the /mps/ cluster is 114, which is greater than the type frequency of the /ms/ cluster by 10. I assume that this difference is negligible, and is comparable to the difference seen between /mθ/ and /mpθ/.

(8) Predictions made by frequency-based (token and type) account

Token: /ms/ > /mf/ ≈ /lθ/ > /mθ/ > /ls/ > /lf/

Type: /ms/ ≈ /mθ/ > /lθ/ ≈ /ls/ > /mf/ ≈ /lf/

(ordered from high to low likelihood of perceiving epenthetic stops)

Compare these orders with the one obtained from the phonological account in (7), which is repeated below:

(9) Prediction made by phonology-based account

/mf/ ≈ /mθ/ ≈ /ms/ > /lθ/ ≈ /ls/ > /lf/

(ordered from high to low likelihood of perceiving epenthetic stops)

We conclude this section with two remarks. First, in both accounts the /lf/ cluster is predicted to induce the fewest epenthetic stops. In both the token and type frequency data, this cluster is very frequent and the corresponding three-consonant cluster /ltf/ has zero frequency. Listeners are thus expected to often ignore the silent interval inserted as the cue for a stop. In the phonology-based account the same prediction is made, but the rationale is different: the /ltf/ cluster is ungrammatical while the corresponding /lf/ cluster is grammatical.

The second remark is that in both frequency-based accounts, more epenthetic stops are predicted to be perceived in the /m/-initial clusters than the /l/-initial clusters (though the relative order between /mf/ and /mθ/ is different between the two frequency-based accounts). Note that, as we saw in the previous section, the phonology-based account also makes this prediction. Roughly speaking, the differences between the two-consonant clusters and the corresponding three-consonant clusters are smaller for the /m/-initial clusters than the /l/-initial ones, which suggests that more epenthetic stops are expected to be perceived in the /m/-initial clusters.

4. Experiment

4.1 Method

4.1.1 Participants

Seventeen listeners were recruited from an undergraduate introductory linguistics course at the University of Massachusetts Amherst for extra credit. All were native speakers of American English and were all from the New England area. No hearing or speaking disorders were reported. Data collected from three listeners were not used for analysis because they used only two of the three possible responses.

4.1.2 Stimuli

Six continua of CVC_1C_2 monosyllable nonwords were created by splicing CVC_1 and C_2 together with silent intervals of varying duration between them. The duration of the silent intervals varied in 10 steps from 0 ms to 100 ms, with an increment of 10 ms at each step. The CVC_1 was /dɛl/ or /dɛm/ and the C_2 was /f/, /s/ or /θ/. All possible combinations of the two types CVC_1 and three C_2 made the six experimental continua, i.e., /dɛlf/ – /dɛl_f/, /dɛlθ/ – /dɛl_θ/, /dɛls/ – /dɛl_s/, /dɛmf/ – /dɛm_f/, /dɛmθ/ – /dɛm_θ/ and /dɛms/ – /dɛm_s/ (underscores indicate silent intervals).

The stimuli were all based on tokens that were naturally produced by a male native speaker of American English. The recording took place in a sound-attenuated room. The materials were recorded onto a CD with 44.1 kHz sampling rate and 16 bit quantization level. The recorded materials were three tokens of six nonwords with no underlying stops between the sonorant consonants and the fricatives, /dɛlf/, /dɛlθ/, /dɛls/, /dɛmf/, /dɛmθ/ and /dɛms/ ('no-stop' set), and three tokens of six more nonwords where the /l/ or /m/ in each nonword was followed by a underlying homorganic voiceless stop, /dɛltf/, /dɛlts/, /dɛltθ/, /dɛmpf/, /dɛmps/ and /dɛmpθ/ ('stop' set). They were recorded in isolation without any carrier sentence, and digitized at 16 kHz/16 bits.

The CVC_1 /dɛl/ and /dɛm/ stimuli were taken from the no-stop set. For each of the C_2 /f/, /s/ and /θ/, a token that showed robust and uniform frication on spectrograms were chosen from the entire tokens in the no-stop and the stop sets. Their durations were 264ms for /ɛl/,[10] 218 ms for /ɛm/ (128 ms for /ɛ/ and 90 ms for /m/), 225 ms for /f/, 236 ms for /θ/ and 285 ms for /s/. They all had gradually falling F_0. For the /dɛl/ the F_0 value declines from 139 Hz at the initial pulse of the vowel to 112 Hz at the last pulse of the sonorant. For the /dɛm/ F_0 started at 144 Hz and ended at 129 Hz.

The reason for the decision to take /dɛl/ and /dɛm/ from the no-stop set rather than the stop set was that they inherently cautioned no acoustic indications of following stops. The acoustic analysis of the recorded materials showed that the /dɛl/ and /dɛm/ in the stop set were shorter than the ones in the no-stop set, confirming Fourakis and Port's (1986) claim. Thus, the short duration of the vocalic part would serve as a cue for the presence of a stop after it. We do not know precisely at this point how this acoustic cue could interact with the silence duration manipulated in the current experiment.[11] But it is possible that listeners would be biased toward hearing more stops if /dɛl/ or /dɛm/ were taken from the stop set, due to their short durations.[12]

A 20 ms cosine amplitude tapering was applied to the end of /dɛl/ and /dɛm/ in order to avoid abrupt offset of the sonorants, which might also help cue a stop. Also, to avoid abrupt onsets of the three fricatives which might also give listeners a cue for a stop, a 30 ms cosine amplitude tapering was applied at the beginning of the fricative.

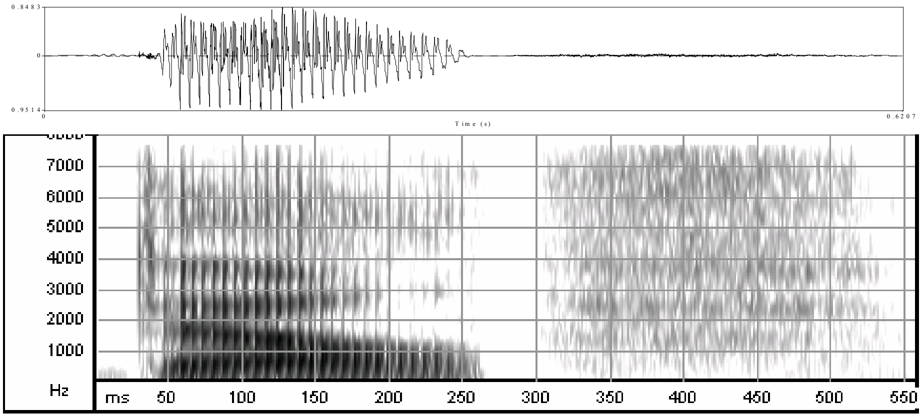
As an example of the stimuli used in the experiment, the waveforms and the spectrograms of the stimuli /dɛl_f/, /dɛl_s/ and /dɛm_θ/ with 30 ms of silence are shown

---

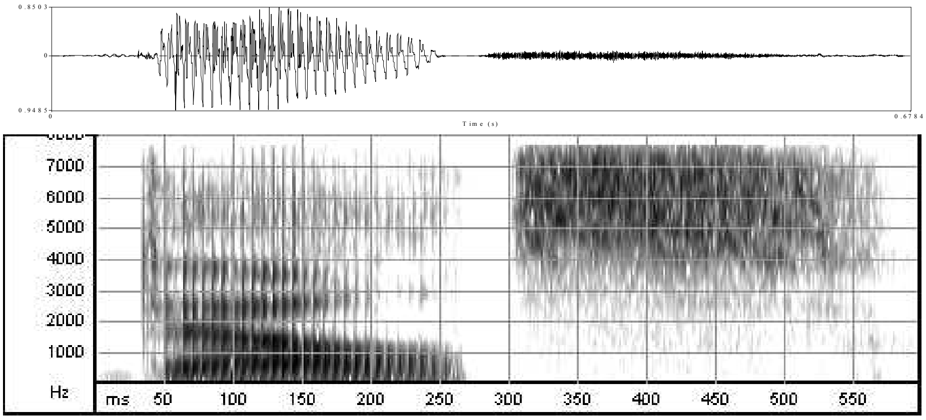
[10] /dɛl/ could not be segmented into /dɛ/ and /l/ due to the smooth transition from /ɛ/ and /l/.

[11] Some kind of trading relation would be expected between the duration of the vocalic part and silent interval (cf. Fitch et al.1980).

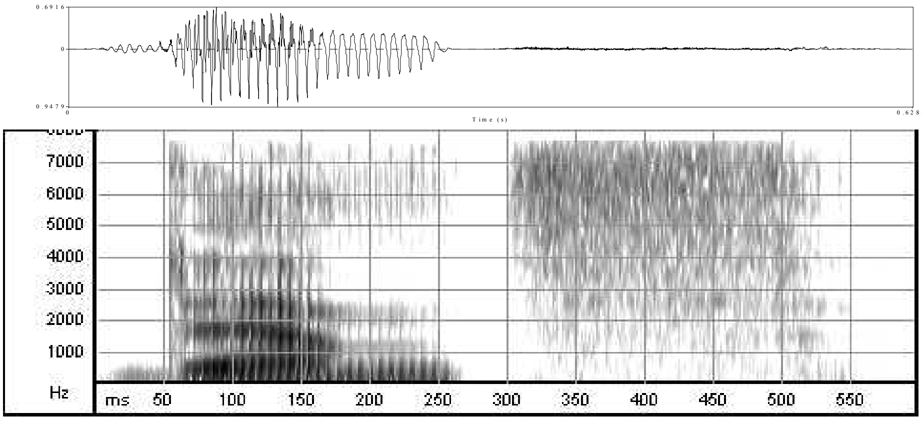
[12] This effect was found in a pilot experiment in which the ambiguous durations were used.



(a) /dɛl_f/ (30 ms silent interval)



(b) /dɛl_s/ (30 ms silent interval)



(c) /dɛm_θ/ (30 ms silent interval)

Figure 2 Waveforms and spectrogram of (a) /dɛl_f/, (b) /dɛl_s/ and (c) /dɛm_θ/ with 30 ms of silence between the sonorants (/l/ and /m/) and the fricatives (/f/, /s/, and /θ/).

in Figure 2. In all of the spectrograms, no burst-like noise is observed in the area where silence was inserted (roughly between 270 ms and 310 ms). Also, there is no abrupt ending of the sonorant in /dɛl/ and /dɛm/, so there should be no possible acoustic cues for a stop other than the introduced silent gaps.

4.1.2 Procedure

The stimuli were transferred from the CD into an MS-DOS Windows PC, which controlled the stimulus presentation. The constructed stimuli were presented to listeners at 20 kHz sampling rate and 16 bit quantization level, one word at a time, through headphones at own comfortable volume levels. A phoneme monitoring task was adopted. The listeners were given three choices of response: /p/, /t/ or 'n(one)'. They were asked to judge whether they heard a /p/, /t/ or 'none' in a word and press corresponding buttons placed in front of them as quickly as possible. Where in the word they should focus on was not told. They had 2 seconds to give a response for each stimulus. The interstimulus interval was 2.5 seconds. The number of obtained responses per cluster type was 5 for the two end point stimuli, 10 for the two stimuli that were next to the endpoints, and 15 for the remaining stimuli. (2 endpoint stimuli × 5 responses + 2 second last endpoint stimuli × 10 responses + 7 remaining stimuli × 15 responses = 135). In each block of trials, 1 response for the 2 endpoint stimuli, 2 responses for the two second-last endpoint stimuli, and 3 responses for the remaining stimuli were obtained for each cluster type. Thus, in each block of trials, 162 stimuli were presented to the listeners ([(2×1) + (2×2) + (7×3)] × 6 = 162). It took approximately 8 minutes to complete a block of trials. Five blocks of trials were run for each listener and the whole session took about 75 minutes. The total number of stimuli each listener listened was 810 (162 stimuli × 5 blocks of trials). Before the test blocks, the listeners went through 24 practice trials, which consisted of two repetitions of the 12 endpoint stimuli of each of the six continua (2 endpoint stimuli for each continuum × 6 continua).

4.2 Results

The response rates for the six different cluster contexts were calculated for each listener across all of the continua. The graphs in Figure 3 represent listeners' mean response rates to each continuum in the six different cluster types. The x-axis represents the 10 stimulus steps from 0 (0 ms silent interval) to 10 (100 ms silent interval) and the y-axis represents response rates. The graph in Figure 3a plots the mean /p/ response rates, the one in Figure 3b the mean /t/ response rates, and the one in Figure 3c the mean 'none' response rates. Figure 3d plots the mean rates obtained by /p/+/t/ response rates, which was obtained by "1-'none' response". Figures 4c and 4d represent the same data in a different way. It is intuitively easier to think how many epenthetic stops are perceived (4d), rather than how many epenthetic stops were ignored (4c). Thus, when we consider the response rates for "none" in the next section, we discuss Figure 3d rather than Figure 3c.

4.2.1 "Stop" vs "no stop" responses

Let us start with the data represented in Figure 3d. The most noticeable pattern is that in most of the cluster types, the listeners reported many stop percepts even when silent interval duration was short. It is striking that many stops are perceived, particularly in /lf/, /mf/ and /ms/, even at the stimulus s0, which contains no silent interval. The /lf/ cluster shows the highest stop response rates across the silent intervals. The mean response rate for /lf/ at stimulus 0 (= no silent interval) was .74. The /ls/ context shows the lowest stop response rates with most of the silent interval durations. The other contexts (/lθ/, /mf/, /mθ/ and /ms/) fall somewhere between the two extreme clusters /lf/ and /ls/. Among the four cluster contexts, /mf/ and /ms/ showed somewhat higher response rates than /lθ/ and /mθ/, especially at shorter silent interval durations. As an overall tendency, we have found the three /m/ contexts generally elicited more stops than the /l/ contexts, except for /lf/.

Recall that the frequency data in Figure 1 make the prediction that a strong perceptual bias against hearing epenthetic stops would be observed in most of the cluster pairs, specifically at least in /lf/, /lθ/, /ls/, and /ms/. The results obtained here do not support this prediction. The /lf/ and /ms/ clusters show very high rates of stop responses at shorter silent intervals. The stop response rates for /lθ/ are also high, though not as high as for /lf/ and /ms/.

Across the silent interval steps, we can see the general tendency that as silent interval increases the stop response rates also increases. However, the magnitude of the change, i.e., the slope, varies depending on the context. The /ls/ context shows the sharpest rise across the silent interval durations, /lθ/ and /mθ/ are intermediate, and the remaining /lf/, /mf/ and /ms/ contexts show much more constant response rates across the silent intervals. We also found an asymmetry between the /l/ and /m/ series of contexts: the listeners show more drastic change in the /l/ contexts except /lf/, while they show more or less steady response rates in the /m/ contexts.

A repeated measures ANOVA was performed on the data presented in Figure 3d, with cluster type and silent interval duration as the independent variables. As our interest is in overall pattern and not in every difference between two adjacent steps, the 11 steps of the continua were reduced to three: (i) from s0 to s2, (ii) from s3 to s6, and (iii) from s7 to s10. The data for each of the three new silent interval steps were pooled and the mean was calculated for each listener. There was a significant effect of context ($F(5,65)=7.946$, $p<.004$) and silent interval ($F(2,26)=46.449$, $p<.0001$), and the interaction ($F(10,130)=5.920$, $p<.001$). Post-hoc pair-wise t-tests for contexts, with α-level corrected by the Bonferroni procedure,[13] revealed that the rates for /lf/ was significantly higher than those for /ls/ ($t(13)=7.347$, $p<.0001$) and those for /lθ/ ($t(13)=3.906$, $p<.002$), respectively, and /mf/ and /ms/ were significantly higher than /ls/, respectively (/mf/-/ls/: $t(13)=-5.054$, p<.0003; /ms/-/ls/: $t(13)=-3.471$, $p<.005$). None of

---

[13] Four paired samples t-tests + 3 planned contrasts (described below) were carried out. Therefore, $\alpha=.05/7=.007$.

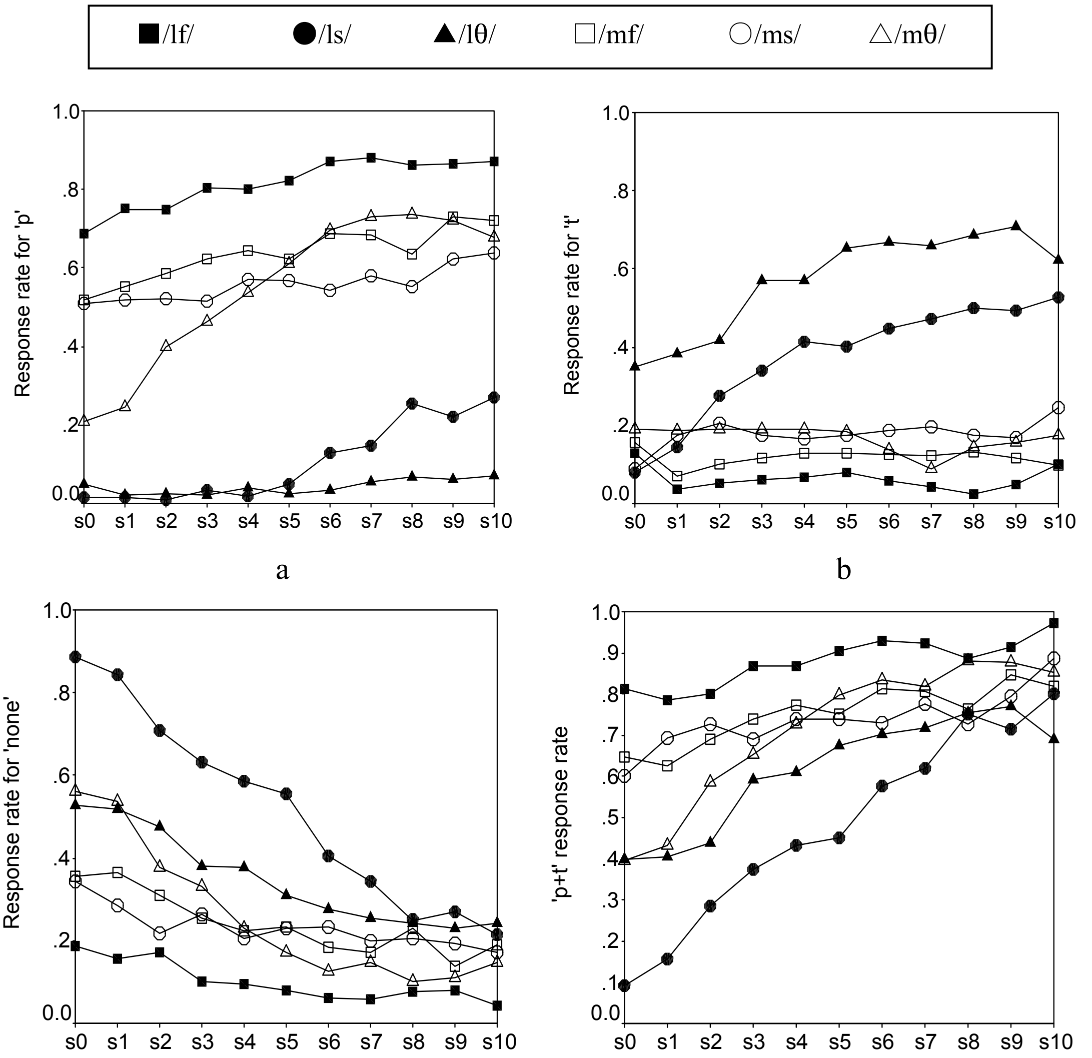


Figure 3 Mean response rates to each of the stimuli in the six different cluster contexts: (a) represents the response rates for /p/, (b) for /t/, (c) for /none/, and (d) for /p/ + /t/. Note that (c) and (d) are essentially equivalent since (d) represents the /p/+/t/ responses for ease of understanding the graph (c). On the y axis "0" indicates 0% response rate and "1" 100%. The x axis represents the stimuli, from s0 (no silent interval) to s10 (100 ms silent interval).

the other response rates differed significantly from one another.

The reasons for the significant interaction effect in the ANOVA is the steepness difference between the response functions of the /lf/, /ms/ and /mf/ contexts, which all show more or less flat functions, and that of the /ls/ context which is much steeper. Planned contrasts showed that the rate differences between /lf/, /ms/ and /mf/ on the one hand and /ls/ on the other were significantly larger at the shortest interval than at the longest interval ($F(1,13)=24.326$, $p<.0003$). Another planned contrast comparing the differences between /lf/, /ms/, and /mf/ as one group and /lθ/ and /mθ/ as the other group was only marginally significant ($F(1,13)=7.381$, $p<.019$). Yet another planned contrast of the response rate differences between /lθ/ and /mθ/ on the one hand and /ls/ as the other was also only marginally significant ($F(1,13)=4.684$, $p<.06$). These results indicate that the /p/+/t/ response function is the gentlest in /lf/, /ms/ and /mf/, intermediate in /lθ/ and /mθ/, and the steepest in /ls/.

Based on the analyses of the /p/+/t/ response rates, we have found the following orderings for the context and the interaction effects shown in (10a) and (10b). The most notable patterns are: (i) high stop response rate for /lf/, (ii) the /m/ contexts otherwise elicited more stop responses than the /l/ contexts, and (iii) a sharper change in response rate across the silent intervals in the /l/ contexts than in the /m/ contexts.

(10) a. Obtained context effect on /p/+/t/ response rate
/lf/ ≈ /mf/ ≈ /ms/ > /lθ/ ≈ /ls/
/mθ/
(From high to low /p/+/t/ response rates)

b. Obtained interaction effect on /p/+/t/ response rate
/ls/ > /lθ/ ≈ /mθ/ > /lf/ ≈ /mf/ ≈ /ms/
(From steepest to gentlest functions)

In (9a) the /mθ/ is put just below the ">" symbol, which indicates that there is no statistically significant difference between /mθ/ and the higher three contexts and the lower two contexts, respectively.

4.2.2 /p/ responses

Next, let us look at the breakdown of the 'p' response rates. They are shown in Figure 3a. One can first notice that, generally speaking, the listeners hear many /p/s in the /m/-initial contexts with the obvious exception of the /lf/ cluster case. In other words, the listeners are *p*-biased in the /m/ clusters. This result generally supports Clements' (1987) claim that an epenthetic stop shares its [place] feature with the preceding sonorant. Note that we are not in a good position to say something about the [voice] feature because listeners did not have response options of /b/ and /d/. The magnitude of the *p*-bias is relatively constant across the silent intervals, especially for /lf/, /mf/ and /ms/. That is, in the /m/ contexts the listeners kept hearing a considerable number of /p/s regardless of silent interval duration. The obvious exception is /lf/. Surprisingly, the listeners were highly *p*-biased (instead of *t*-biased) throughout the silent intervals, contradicting Clements' claim, which suggests that it is necessary to consider some other principle that governs the determination of the nature of epenthetic stops. Moreover, the degree of the *p*-bias for

/lf/ was larger than the three /m/ context cases. What is especially striking is the listeners heard more /p/s in /lf/ than in /mf/, in which both the sonorant and fricative are labials while only one of the two is so in /lf/. The /mθ/ response rates for /p/ in Figure 3a showed a rise from the shorter to the longer silent intervals (especially from s0 to s7), indicating that in this context the listeners tend to ignore more silent intervals as a possible cue for a stop when they were relatively short.

The /p/ response rates were submitted to a repeated-measure ANOVA with cluster type (6 levels) and silent interval (3 levels, just as in the analysis of the /p/+/t/ responses above) as the independent variables. Both of the main effects and the interaction were significant ($F(5,65)=55.834$, $p<.0001$ for cluster type, $F(2,26)=18.223$, $p<.0001$ for silent interval, $F(10,130)=4.888$, $p<.008$ for the interaction;). Planned contrasts on /p/ response rate revealed that there are four distinct response patterns: (i) high flat response function (/lf/), (ii) intermediate flat function (/mf/ and /ms/), (iii) intermediate positive function (/mθ/), and (iv) low flat or slightly positive function (/ls/ and /lθ/). Specifically, the mean response rates for /lf/ were significantly greater than those averaged between /mf/ and /ms/ ($F(1,13)=55.834$, $p<.0001$) and those for /mθ/ ($F(1,13)=26.325$, $p<.0002$), respectively. There was no significant main effect on cluster type between the mean response rates for /mθ/ and those averaged between /mf/ and /ms/, but the interaction between cluster type and silent interval was significant ($F(1,13)=5.67$, $p<.026$). This indicates that the response function of /mθ/ is significantly steeper than /mf/ and /ms/; this is caused by the low response rates in /mθ/ at shorter silent intervals. The response rates for /mθ/ cluster and those averaged between /mf/ and /ms/, respectively, significantly differed from those averaged over /ls/ and /lθ/.

Summarizing, the responses are *p*-biased in clusters /lf/, /mf/, /ms/ and /mθ/. The p-bias is the highest in the /lf/ cluster, and /mf/, /ms/ and /mθ/ are lower. The responses for /mθ/ cluster show a sharp change depending on silent interval durations whereas those for the other clusters are more or less constant across the silent intervals.

4.2.3 /t/ responses

The mean /t/ response rates are plotted in Figure 3b. One conspicuous pattern is that the listeners heard more /t/s in the cluster environments /lθ/ and /ls/ than in the other clusters, indicating that they are more *t*-biased in /lθ/ and /ls/ than in the other clusters. These results also support Clements' (1987) claim about the featural composition of epenthetic stops. The clusters other than /lθ/ and /ls/ show consistently low /t/ response rates across the silent intervals.

Across the entire range of silent intervals, the response rates are greater for the /lθ/ cluster than the /ls/ one. The examination of the 'no-stop' responses in Figure 3c tells us that the listeners ignore silent intervals more in /ls/ than in /lθ/.

The /lθ/ and /ls/ contexts also show a similar pattern to /mθ/ in the /p/ responses in that the magnitude of the *t*-bias increases as silent interval duration increases. When the silent interval was relatively short the listeners tend ignore it but they respond to it quite reliably when the silent interval was long. Here again, we found an asymmetry between the /l/ and the /m/ contexts. That is, setting aside /mθ/ and also /lf/, the /l/-initial clusters showed a steeper response functions than the /m/-initial clusters.

A two-way repeated measure ANOVA with cluster type (6 levels) and silent interval (3 levels) as independent variables was significant for both of the main effects and interaction ($F(5,65)=12.454$, $p<.0001$ for cluster type, $F(2,26)=11.337$, $p<.0014$ for silent interval, $F(10,130)=8.275$, $p<.0002$ for interaction). Planned contrasts comparing the response rates for /lf/, those for /lθ/, and those averaged over the rest of the clusters /lf/, /mf/, /ms/ and /mθ/ revealed that /ls/ and /lθ/ were significantly different from the averaged response rates, respectively (/ls/ vs averaged rates: $F(1,13)=8.910$, $p<.011$; /lθ/ vs averaged rates: $F(1,13) = 52.158$, $p<.0001$). However, the difference between /ls/ and /lθ/ was marginal ($F(1,13) = 4.059$, $p<.065$).

4.2.4 Comparing /p/ and /t/ responses

In our discussion so far, we have found that the responses for the /lf/, /mf/, /ms/ and /mθ/ clusters are *p*-biased and those for /ls/ and /lθ/ are *t*-biased. Here, one can ask the question of whether there is any difference in the degree of bias between *p*-biased and *t*-biased cluster types. In fact, the relevant data are already provided in Figure 3d, where response rates for /p/+/t/ are represented.

The graph shows that the response rates are the lowest in /ls/ and somewhat lower in /lθ/ than in the three /m/-initial clusters and /lf/. Since we know that the /ls/ and /lθ/ responses are *t*-biased and the responses for the /m/-initial clusters and /lf/ are instead *p*-biased, it can be concluded that the degrees of the *p*-bias observed in /lf/, /mf/, /ms/ and /mθ/ are greater than those of the *t*-bias in /ls/ and /lθ/.

4.3 Discussion

4.3.1 Evaluation of the phonology-based account and frequency-based accounts

The patterns obtained in the experiment are repeated in Table 1, and are contrasted with the predictions made by the phonology- and the frequency-based accounts discussed in §3. In Table 1, a comparison of the predicted cluster orders in (a), (b) and (c) on the one hand, with the obtained results in (d) on the other reveals that the ability to account for the obtained patterns is quite poor for both frequency-based and phonology-based accounts. The result that is most damaging to the frequency-based accounts is that some clusters for which some bias against hearing epenthetic stops was predicted turned out to induce many stops. For example, in both types of the frequency-based accounts, clusters such as /lf/, /mθ/ and /ms/ are predicted to induce stop percepts less than the other clusters, since the two consonant clusters are much more frequent than the corresponding three consonant clusters /ltf/, /mpθ/ and /mts/. What we saw in the experiment is high

Prediction	Phonology-based		a.	/mf/ ≈ /mθ/ ≈ /ms/ ≈ /lθ/ ≈ /ls/ > /lf/
	Frequency- based	Token	b.	/ms/ > /mf/ ≈ /lθ/ > /mθ/ > /ls/ > /lf/
		Type	c.	/ms/ ≈ /mθ/ > /lθ/ ≈ /ls/ > /mf/ ≈ /lf/
Result	‘Some stop’ response		d.	/lf/ ≈ /mf/ ≈ /ms/ > /lθ/ ≈ /ls/ /mθ/

Table 1 Summary of the predictions made by the phonology-based and frequency-based theories and the results obtained in the experiment.

response rates for epenthetic stops in all of these contexts.

We obtained high response rates for /p/ instead of /t/ in the /lf/ cluster. However, even when we take this into consideration, the frequency accounts’ ability to explain the obtained results is still poor. The /lpf/ cluster is only possible across a syllable boundary as in /hɛlp.fʊl/ *helpful*, and its frequency is far lower than that of /lf/ (/lpf/: 510 by token frequency; 3 by type frequency). Thus, a strong bias toward /lf/ is still expected (we discuss the /lf/ cluster further in the next section).

Apparently, the phonology-based account does not do any better than the frequency-based accounts. However, it does a bit better job on the /lf/ cluster, as is discussed in the next section. The prediction of the phonology-based account on clusters other than /lf/ is that no perceptual bias is expected among those clusters, which is obviously not supported by our experimental results. We found that the way a silent interval is perceived as a stop varies incredibly among those clusters. What we need is to find some other factor(s) that governs the perception of the epenthetic stops.

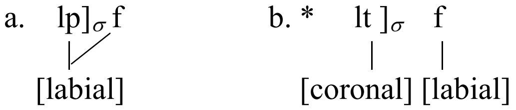
4.3.2 The /lf/ effect

In both phonology-based and frequency-based accounts, /lf/ was predicted to induce the fewest epenthetic stops, which is completely opposite the obtained results.

I argue that this result is problematic for the frequency-based accounts but is not anomalous in the phonology-based one. When we discussed the predictions that the two different types of account make in §3, we assumed that listeners would hear /t/s in the /lf/ cluster. However, as the results show, it was /p/ not /t/ that was perceived in this environment. The question we should ask then is whether the two different types of account are capable of distinguishing /lpf/ from /ltf/ such that /p/ is perceived more than /t/ in this context.

The phonology account can account for this asymmetry between /p/ and /t/, taking advantage of the fact that the cluster /lpf/ does not violate the syllable appendix constraint illustrated in (6) but /ltf/ does (Hayes 1986, Ito 1986, 1989). The consonant in the syllable appendix is noncoronal in /lpf/, but it shares its [labial] feature with the preceding /p/, whereby it circumvents the violation of the constraint (just as in /mpf/ in (6a)), as shown in the representation in (11a). The /ltf/ cluster, however, violates the constraint, as in (11b):

(11)



Because of this violation, /ltf/ was disfavored in comparison with the /lpf/ cluster. The phonology-based account therefore is compatible with the result obtained for the /lf/ context.

The frequency account, whether based on token or type frequency, cannot distinguish /lpf/ from /ltf/. We can see that /ltf/ has zero frequency both by token (Figure 1a) and type (Figure 1b). In the CELEX database, the frequency of the /lpf/ cluster is not zero but is still extremely low, since the cluster can only occur across a syllable boundary (refer to the frequency values for /lpf/ mentioned above). The only prediction the frequency-based accounts make regarding the number of epenthetic /p/s and /t/s between /l/ and /f/ is that there will be no perceptual preference (or there might be very weak preference for /p/ over /t/ since /lpf/ shows frequency while /ltf/ never occurs). Clearly, this cannot explain the strong *p*-bias observed in the /lf/ cluster.

4.3.3 Asymmetries between /l/ and /m/ contexts

One of the predictions made by the frequency-based accounts was that more epenthetic stops would be perceived in the /m/-initial clusters than in the /l/-initial ones. This prediction is generally supported by our experimental results. Aside from the special /lf/ context discussed above, we found the general tendency to be /m/ contexts > /l/ contexts: the /mf/, /ms/ and /mθ/ contexts elicited more stops than /lθ/ and /ls/ (ordering (d) in Table 1).

A phonology-based theory can also give an account of the overall "/m/ > /l/" effect by invoking the widely acknowledged idea of markedness of labials (or unmarkedness of coronals) (Paradis and Prunet, 1991). Jun (1995, 1996) and Hume et al. (1999) show that labials (and velars) are more marked than coronals, not only in formal phonological treatments but also in speech perception. They claim that labials (and velars) are perceptually more salient than coronals and thus more perceptible. Based on this claim, we can give an account of the obtained result that more labial epenthetic stops were perceived in the /m/ contexts than were epenthetic coronal stops in the /l/ contexts. The preceding /m/ is perceptually more salient, which facilitates listeners' attributing the acoustic cue for the labial sonorant to the place specification of the stop perceived in the following silent interval. Thus, the silent interval is identified as a labial stop after /m/ more often than it is identified as a coronal stop after /l/.

Since both frequency and phonology could account for the /m/ > /l/ effect, it is difficult to determine which is the right one from the data we have at hand with respect to this asymmetry. However, there is a piece of evidence for the phonological account. Warner and Weber (2001) report a similar result in Dutch, and conclude that the general perceptual bias against epenthetic /t/s and toward epenthetic /p/s and /k/s are probably not related to perceptual salience based on the phonological markedness of labials and velars. They show that their listeners are less likely to respond to /t/ than to /p/ or /k/, but they argue that that is because the epenthetic bursts in the environment of /t/ were weaker than those in the environment of epenthetic /p/ or /k/. They suggest that this may be due solely to production factors for the reasons that only one item in their materials (/np/) had an epenthetic burst but very weak, that even in an item such as /ns/ in which the epenthetic burst is widespread it is expected that epenthetic burst is weak due to the homorganicity of /n/ and /s/, and that the /ns/ stimuli itself had shorter bursts than most conditions.

This interpretation of Warner and Weber's (2001) is called into question here. The stimuli in our experiment did not contain bursts of any kind and aimed to induce stop percepts only by manipulation of the duration of the silent interval. If Warner and Weber's interpretation was correct in that only production factors were involved in the asymmetry between coronals and labials (and velars), we would obtain the result in which the listeners' bias is neutral between the two place features. This was clearly not the case in our experiment. We found that the asymmetry is still present even when there are no bursts to cue epenthetic stops.

An equally likely account that is based on the featural difference between nasals and laterals is also possible. According to this account, a stop percept is more likely after /m/ than /l/ because the former is specified [−continuant] while the latter is specified [+continuant].[14] Listeners may have attributed the [−continuant] to the stop they perceived at the silent interval, and as a consequence more stops were perceived after /m/ than /l/.

To tease the two possibilities apart, we could use pairs of sonorant-obstruent clusters such as /ns/-/ms/ and /ns/-/ls/. The /n/ in /ns/ differs from the /m/ in /ms/ and from the /l/ in /ls/ in its place features and its manner feature, respectively. The account based on the featural contrast between [−continuant] and [+continuant] predicts equal likelihood of perceiving /t/s and /p/s for the /ns/-/ms/ cluster pair, since both /n/ and /m/ are [−continuant]. If the account based on the markedness of labial is correct, on the other hand, more epenthetic /p/s are predicted in /ms/ than epenthetic /t/s in /ns/, because /n/ is coronal while /m/ is labial. As for the /ns/-/ls/ cluster, the featural contrast account predicts more /t/ responses in /ns/ than in /ls/ while the place markedness account predicts equal likelihood of epenthetic /t/s for both clusters. I leave this issue for future research.

4.3.4 When are segments perceived?

We should ask under what environments stops, or more generally segments, are perceived. That is, why were so many epenthetic stops perceived at shorter silent intervals, or even when there was no silence, especially in /lf/ and the /m/ clusters?

The answer to the question is not obvious, especially in the environment of no silent interval. If there was no acoustic cue for a stop in the stimuli, the stop percept may be a perceptual illusion caused by some top-down process. An example of perceptual illusion is nicely demonstrated by Dupoux, Kakehi, Hirose, Pallier and Mehler (1999). They showed that Japanese listeners could not reliably judge which of a pair of nonwords such as *ebuzo* and *ebzo* they heard, while French listeners had no difficulty in

---

[14] I thank John Kingston for this suggestion.

discriminating them. They conclude that the phonotactic knowledge of the Japanese listeners corrected deviant forms in a way that they conform to the syllable structure of the language. However, it appears difficult to consider that English phonotactics is responsible for the stop percepts with no silent interval. For one thing, we saw that our phonological account permits most of the clusters used in the experiment to be grammatical. Moreover, although two consonant clusters such as /lf/ and /mf/ are perfect in English, the listeners identified very few such clusters; instead, they perceived /lpf/ and /mpf/, respectively.

The account based on the difference in feature specification for [±continuant] mentioned above could account for the many stop percepts in the /m/ clusters when there is no silent gap: listeners know that /m/ is specified for [–continuant] and /l/ for [+continuant]. The noncontinuancy of the /m/ may generate a stop percept after it. /l/ would not allow such a possibility due to its [+continuant] specification.

The high response rates for epenthetic stops in the cluster /lf/ with no silent interval, at least for now, are a mystery. However, an account is possible for the overall high stop response rates if we seek an explanation based on articulatory constraint. The /lf/ cluster consists of a lateral whose place of articulation is posterior to the following fricative. When producing this cluster, the constriction of the /f/ may precede the release of the /l/, due to an anticipatory coarticulation. When this happens, the release of the /l/ may be articulatorily masked by the labial constriction of the /f/, and therefore, no epenthetic stop would be audible. We could consider that the high response rates for /p/ in the /lf/ cluster may be a consequence of the listeners' perceptual compensation for this articulatory masking. It may be that they responded to the very short silent gaps more because they somehow knew this articulatory constraint and compensated for the masking effect of the /lf/ cluster.

4.3.5 Use of acoustic information of adjacent segments for epenthetic stops

Our experimental results generally support Clements' (1987) claim that an epenthetic stop appearing between a sonorant and an obstruent shares its place feature with the preceding sonorant. With respect to the voice feature, we cannot conclude that it is shared with the following fricative because in our experiment the listeners were not given voiced stops as response options. However, one issue we need to consider is whether they used the silent interval or the voicelessness of the fricative to identify the voice feature of the epenthetic stop. Since there is acoustically nothing during the silent interval, there is no voice bar either. There is a possibility that the listeners might have judged the absence of the voice bar in the silent portion to indicate the presence of a voiceless consonant. If this was the case, Clements' (1987) claim would be damaged, because an acoustic property which does not belong to either of the adjacent segments is used to identify the [voice] of the epenthetic stop. Clements, on the other hand, assumes that the [voice] feature does not come from the silence: it spreads from the following fricative.

Our finding that /p/ responses were dominantly perceived in the /lf/ cluster is also problematic to Clements' claim (1987) in that the epenthetic stop does not share its place feature with the preceding sonorant. Clearly, some other account is necessary. My proposal involves the markedness difference between coronals and noncoronals. First, it is reasonable to assume that when the place of articulation of a segment is to be identified where there is no robust cue, listeners use the place information of the adjacent segments. The idea is that the acoustic information regarding the place feature contained in the adjacent segments "compete" with one another to "win" the perceptual place of articulation of the silent gap. The place feature of the stop perceived in the gap is determined in accordance with the asymmetry between coronals and noncoronals. In the case of /lf/, listeners use the coronality of the /l/ and the labiality of the /f/. Since the labiality is perceptually more robust than the coronality, listeners would attribute it to the place of articulation of the stop corresponding to the silent interval. When they encounter a silent interval between /m/ and /f/, a /p/ would be perceived because [labial] is the only available feature. In /ms/ or /mθ/, one of the consonants is labial and the other is coronal, so the labial would win, and hence /p/ is the percept. Since both consonants are coronal in clusters /ls/ and /lθ/, only epenthetic /t/ is possible.

5. Conclusion

In this paper, we have reported an experiment investigating the perception of epenthetic stops in English, with respect to how it is affected by cluster type and duration of a silent interval. We addressed three issues: (i) whether the perception of epenthetic stops is influenced by phonotactic constraints or by cluster frequency, (ii) under what circumstances listeners hear a stop, or more generally, a segment between a sonorant and a fricative, and (iii) when listeners hear an epenthetic stop, what is its featural composition and how are the acoustic properties of the adjacent segments used to identify it.

We first considered Warner and Weber's (2001) phonotactic effect seen in Dutch, which is that listeners tend to respond less to an epenthetic stop if it would form a phonotactically illegal cluster with the adjacent consonants than if it would not. However, I argued that the phonotactic effect in perception of epenthetic stops that Warner and Weber claim is invalid in that no response bias is expected between illegal two consonant clusters and the corresponding illegal three consonant clusters. Our experiment contained stimuli that predict response biases toward one of the clusters. Two different kinds of predictions were presented based on two distinct accounts – one based on frequency and one on phonological knowledge, and those predictions were tested in the experiment.

We have found that the obtained results lend little support to either the phonology-based or the frequency-based accounts. The most striking result is that an extremely high response rate for /p/ was observed in the /lf/ cluster. However, it has been shown that the phonology-based account is superior to the frequency-based account at least in one respect: it can distinguish the /ltf/ cluster from the /lpf/ cluster such that /lpf/ cluster is grammatical but the cluster /ltf/ is not, while the frequency-based account cannot distinguish these two clusters.

The experiment has shown that a silent interval plays an important role in the perception of epenthetic stops such that a long interval induces more stop percepts. The most striking result, however, is that many epenthetic stop responses were observed even when no silent interval existed. We also found that there is an asymmetry between the /m/ and /l/ clusters: more epenthetic stops are likely after /m/ than /l/. We considered two possibilities to account for this discrepancy: one based on the markedness of labial place feature and the other based on the feature [±continuant]. Which account is correct remains to be determined.

We have seen that Clements' (1987) claim that an epenthetic stop shares its place feature with the preceding sonorant and its voice feature with the following obstruent is generally supported. However, we found one case where the place feature of an epenthetic stop is shared with not the preceding sonorant but with the following obstruent, i.e., /lpf/. We proposed an account based on the phonological markedness of noncoronals, whereby the acoustic properties of the adjacent segments compete with one another to determine the place of articulation for the stop perceived between them. In the case of /lpf/, /p/ is the perceived epenthetic stop because the labiality of /f/ is perceptually more salient than the coronality of /l/.

References

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

Baayen, H., Piepenbrock, R., and van Rijin, H. (1993) *The CELEX Lexical Database.* Linguistic Data Consortium, University of Pennsylvania, Philadelphia.

Bybee, J. L. (1998) "The emergent lexicon," *Proceedings of the Chicago Linguistic Society: The Panels,* vol. 2, Chicago Linguistic Society 34: 421–436.

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

Blevins, J. (1995) "The syllable in phonological theory," In J. Goldsmith (ed.) *The Handbook of Phonological Theory*. Oxford: Blackwell, 206-244.

Booij, G. (1995) *The Phonology of Dutch*. Oxford: Oxford University Press.

Clements, G. N. (1987) "Phonological feature representation and the description of intrusive stops," *Proceedings of the Parasession on Autosegmental and Metrical Phonology*, Chicago Linguistics Society 23, 29-51.

Clements, G. N. and Keyser, S. J. (1983) *CV Phonology: A Generative Theory of the Syllable.* Cambridge, MA: MIT Press.

Cruttenden, A. (2001) *Gimson's Pronunciation of English.* Sixth edition, Edward Arnold.

Donslaar, W. van, Kuijpers, C. and Culter, 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., and Mehler, J. (1999) "Epenthetic vowels in Japanese: A perceptual illusion?," *Journal of Experimental Psychology: Human Perception and Performance* 25, 1568-1578.

Fitch, H. L., Halwes, T., Erikson, D. and Liberman, A. (1980) "Perceptual equivalence of two acoustic cues for stop-consonant manner," *Perception and Psychophysics* 27, 343-350.

Fourakis, M. and Port, R. (1986) “Stop epenthesis in English,” *Journal of Phonetics* 14, 197-221.

Fudge, E., C. (1968) “Syllables,” *Journal of Linguistics* 4, 253-286.

Gaskell, M. G. and Marslen-Wilson, W. D. (1996) “Phonological variation and inference in lexical access,” *Journal of Experimental Psychology: Human Perception and Performances* 22, 144-158.

Giegerich, H. (1992) *English Phonology: An introduction.* Cambridge: Cambridge University Press.

Goldsmith, J. (1990) *Autosegmental and metrical phonology*. Oxford: Blackwell

Hay J., Pierrehumbert, J. and Beckman, M. (2004) “Speech perception, well-formedness and the statistics of the lexicon,” In J. Local, R. Ogden, R. Temple (eds.) *Phonetic Interpretation: Papers in Laboratory Phonology VI*, Cambridge: Cambridge University Press, 58-74.

Hayes, B. (1986) “Inalterability in CV phonology,” *Language* 62, 321–351.

Hume, E., Johnson, K., Seo, M. and Tserdanelis, G. (1999) “A cross-linguistics study of stop place perception,” In *Proceedings of the International Congress of Phonetic Sciences*, San Francisco, 2069-2072.

Itô, J. (1986) *Syllable Theory in Prosodic Phonology*. Ph.D. Dissertation, University of Massachusetts, Amherst.

Itô, J. (1989) “A prosodic theory of epenthesis,” *Natural Language and Linguistic Theory* 7, 217-259.

Itô, J. and Mester, A. (1993) “Licensed segments and safe paths," *Canadian Journal of Linguistics* 38, 197-213.

Jones, D. (1966) *The Pronunciation of English*. Cambridge: Cambridge University Press.

Jun. J. (1995) *Constraint Based Analysis of Place Assimilation Typology*. Ph.D. Dissertation, UCLA.

Jun, J. (1996) “Place assimilation as the results of conflicting perceptual and articulatory constraints,” *Proceedings of the West Coast Conference on Formal Linguistics* 14, 221-237.

Luce, P. Pisoni, D., and Goldinger, P. (1990) “Similarity neighborhood of spoken words,” In M. Altman (ed.) *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives.* Cambridge, Mass: MIT Press, 122-147.

Keating, P., Blankenship, B., Byrd, D., Flemming, E., and Todaka, Y. (1992) “Phonetic analyses of the TIMIT corpus of American English,” In J. J. Ohala, T. M. Nearey, B. L. Derwing, M. M. Hodge and G. E. Wiebe (eds.) *Proceedings of the International Conference on Spoken Language Processing*, 823-826.

Kenstowicz, M. (1994) *Phonology in Generative Grammar.* Cambridge, Mass: Blackwell.

Kiparsky, P. (1981) “Remarks on the metrical structure of the syllable,” In W. Dresser, O. Pfeiffer, and J. Rennison (eds.) *Phonologia 1980*. Innsbruck: Innsbrucker Beitrage zur Sprachwissenschaft, 245-256.

Massaro, D. W. and Cohen, M. (1983) “Phonological context in speech perception,” *Perception and Psychophysics* 34, 338-348.

Moreton, E. (2002) *Phonological grammar in speech perception.* Ph.D. dissertation, University of Massachusetts, Amherst.

Moreton, E. and Amano, S. (1999) “Phonotactics in the perception of Japanese vowel length: Evidence for long-distance dependencies,” *Proceedings of the 6th European Conference on Speech Communication and Technology*, Budapest.

Ohala, J. J. (1974) “Experimental historical phonology,” In J. M. Anderson and C. Ones (eds.) *Historical Linguistics II: Theory and Description in Phonology*, North-Holland Publishing Co., 353-389.

Paradis, C. and Prunet, J-F. (eds.) (1991) *Phonetics and Phonology: The Special Status of Coronals*. San Diego: Academic Press.

Pitt, M. (1998) “Phonological processes and the perception of phonotactically illegal consonant clusters,” *Perception and Psychophysics* 60, 941-951.

Pitt, M. and McQueen, J. (1998) “Is compensation for coarticulation mediated by the lexicon?,” *Journal of Memory and Language* 39, 347-370.

Vitevich, M. S., Luce, P. A. Charles-Luce, J. and Kemmerer, D. (1997) “Phonotactics and syllable implications for the processing of spoken nonsense words,” *Language and Speech* 40, 47-62.

Vitevich, M. S. and Luce, P. A. (1999) “Probabilistic phonotactics and neighborhood activation in spoken word recognition,” *Journal of Memory and Language* 40, 374-408.

Warner, N. and Weber, A. (2001) “Perception of epenthetic stops,” *Journal of Phonetics* 29, 53-87.

Warner, N. and Weber, A. (2002) “Stop epenthesis at syllable boundaries,” *Proceedings of the International Conference on Spoken Language Processing*, 1121-1124.

Yoo, I. W. and Blankenship, B. (2003) “Duration of epenthetic /t/ in polysyllabic American English words,” *Journal of International Phonetic Association* 33, 153-164.

Department of Linguistics
South College
University of Massachusetts
Amherst, MA 01003

tshinya@linguist.umass.edu