The Effect of Consonant Clusters on Vowel Duration in English[*]

Paul de Lacy
University of Massachusetts, Amherst

14 May 1998

Abstract. This study examines the effect of word-final consonant clusters on vowel duration in English. The duration of the high front lax vowel [ɪ] was measured with respect to 64 different word-final consonant sequences of one, two, and three segments in length. A set of 960 [tɪC(C)(C)] words were produced by three speakers and analysed. In accordance with previous studies, it was found that both voicing and continuancy of a single word-final consonant affects vowel duration. In comparison, for Sonorant+C clusters the effect of the C's voicing was evident but its continuancy had no effect. All other consonants blocked the effect of continuancy as well as voicing. The number of consonants was found to have a small but systematic effect on vowel duration, with sequences of two and three consonants causing shortening of 10-20ms except in sequences beginning with voiceless stops.

1. Introduction

Many studies have shown that the voicing and continuancy of a following consonant has a significant effect on vowel duration in English (Peterson & Lehiste 1960, Chen 1970, Mack 1982, Crystal & House 1988b,d, Laefur 1992, van Santen 1992). However, examination of the effects of bi-consonantal clusters on vowel duration have shown that liquids and nasals are *transparent* with regard to voicing: it is the voicing of the C in Liquid+C and Nasal+C clusters affects vowel duration, the voicing of the sonorant has no effect (Chen 1970, Crystal and House 1988d, van Santen 1992). However, both liquids and nasals are *opaque* for continuancy: the continuancy of a C in Liquid+C and Nasal+C clusters has no influence on the preceding vowel's duration.

The present study re-examines the effect of Sonorant+C clusters on vowel duration. In addition, the influence of voicing, continuancy, and number of segments on duration in other CC and CCC clusters is investigated. The aim is to achieve a more precise description of the transparency and opacity of various segment types as well as to determine whether the number of segments in a syllable-final consonant cluster affects vowel duration.

The details of the experiment undertaken are presented in section 2. The results are presented in section 3. The implications of these results are considered in section 4.

2. Experiment

2.1 Subjects

Three adult speakers of American English (one male – NR – and two females – NH and ML) participated in this study. The speakers were born and raised in a variety of places –

[*] I would like to thank John Kingston for his guidance and advice and Norvin Richards, Nancy Hall, and Meredith Landman for undergoing the trauma of phonetic experimentation. All errors are, of course, my own.

Alabama, Maryland, and California respectively. Despite this wide geographical distribution, there was no significant disparity in the results, suggesting that dialectical differences in vowel duration are slight between these dialects. The speakers were all in their twenties.

2.2 Test Materials

A set of 64 mono-syllabic words was used. Each word had the form [tɪC(C)(C)]. The onset consonant was kept constant so as to eliminate the possibility of it affecting duration (Fischer-Jørgensen 1964, Crystal & House 1998c, van Santen 1992:530-1). Only one vowel was used so as to reasonably limit the study. A short vowel was chosen (as opposed to a long vowel) so as to admit the widest variety of coda consonant clusters. [ɪ] was selected since it provided the most easily visible vowel-to-liquid transitions in spectrograms.

The consonant clusters consisted mainly of the alveolar consonants [t d s z n l] and the alveo-palatals [ʧ ʤ]. Only the lateral liquid was used in this study (though see §3.3.1). These will be referred to by the following abbreviations (based on manner of articulation): S = stop, F = fricative, A= Affricate, N = Nasal, L = liquid. The following consonant sequences were investigated:

(i) 1-member: S, F, A, N, L
(ii) 2-member: (i) LN, LA, LF, LS
(ii) NA, NF, NS
(iii) AS
(iv) FF, FS
(v) SS, SF
(iii) 3-member: (i) LNS, LNF, LAS, LFF, LFS, LSF, LSS
(ii) NAS, NFF, NFS, NSF
(iii) FFF, FSF, FSS
(iv) SFS, SFF

Affricates were treated as single consonants, not as a SF sequence. There is phonological justification for this, but in terms of vowel duration the phonetic evidence is ambiguous: the results show that vowel duration before an A, an SF cluster, and an S alone are all very similar.

In the cases where clusters consisting entirely of alveolar consonants were impossible to find, interdentals or labials were used instead (e.g. in FF clusters [fs], FFF clusters [fθs], NFF clusters [mfs], and SS clusters [pt]).

Given the restricted word template, a number of non-occurring words were used (e.g. [tɪnʧ], [tɪlʤ]). Of the 66 forms, 30 were actually occurring words. There was no evidence that using non-occuring forms unduly affected the outcome of the experiment. A number of measurements were made of actually occurring forms (e.g. *pinch* for [tɪnʧ], *bilge* for [tɪlʤ]); the vowel durations in the occuring forms were found to be virtually indistinguishable from the non-occuring forms. Given this, it was decided to use the same onset consonant for each word so as to eliminate any possible effect that different onsets might have.

As final evidence that the unattested forms had no effect on the results, the results themselves show a high degree of consistency with other studies (see below), and the relevant vowel length distinctions are internally consistent – there is no disparity between attested and unattested forms (i.e. vowel duration in attested *tint* is similar to that in unattested [tɪntʃ], as expected from previous studies). Consonant clusters were limited to only those that are possible in English (see §3.3.1).

2.3 Procedure

Five wordlists were constructed with the 66 words in random order (a total of 330 words). The speakers were instructed to read over the wordlists before being recorded so as to familiarise themselves with the unattested words. With each unattested word an actually occurring word was provided so as to give the reader a reference.

Speakers were instructed to read the words at their normal speed. Each word was read in the frame sentence *I saw* ___ with focus (phrase-stress) on the final word. This frame sentence was used so as to keep prosodic factors (e.g. stress, intonation) constant. The test words were pronounced phrase-finally and with focus. This environment was chosen since it has been demonstrated from other studies that the influence of voicing and continuancy on vowel duration is most distinct in this context (Klatt 1973, Crystal & House 1988a, van Santen 1992). The speakers were told to repeat any mispronounced word.

2.4 Equipment and Measurement Methods

Recordings were made in a sound-attenuated room in the Phonetics Laboratory at the University of Massachusetts, Amherst. The recordings were made with a Beyerdynamic M300N(C) microphone connected to a Symetrix SX202 microphone preamplifier, which was in turn connected to an NAD 3225PE amplifier and a Nakamichi MR-2 tape deck. All words were then digitised at 16000Hz, 16-bit, monoaural mode using the computer program *xwaves* running on a Sun Operating system, via an Ariel Proport 656 analog to digital converter. Spectrograms were made of each test word. Durational measurements were taken directly from the display, which was accurate up to a hundredth of a millisecond.

Following Peterson & Lehiste (1960), the vowel was measured from the stop-burst (i.e. VOT was included as part of the vowel) (*cf* Chen 1970, Mack 1982:174). The end of the onset consonant's stop burst (i.e. beginning of VOT) was clearly evident for all words.

2.4.1 Post-vocalic Consonants

There was no difficulty in discerning the beginning of post-vocalic stops, affricates, and fricatives from broad-band spectrograms. The end of the vowel before stops, affricates, and voiceless fricatives was marked by the sudden disappearance of energy at F1 and F2. The beginning of both voiced and voiceless fricatives was obvious from the appearance of high frequency noise.

Peterson & Lehiste (1960:695) note that segmentation of a vowel and a following nasal is sometimes obscured by nasalisation of the vowel. This did not prove to be a complicating factor in the majority of words. Only seven had to be discarded as having indiscernible transitions. The vowel-to-nasal transition was typified by a sharp decrease in intensity of formants at all levels.

Vowel to liquid transitions presented some difficulties, as they have done in other studies (see esp. Peterson & Lehiste 1960:698). The V to L transition was occasionally so smooth as to preclude accurate segmentation. A total of 20 words were discarded because of this problem. The principle criterion used to segment the other cases was a rise in F2. In a number of cases, narrow band spectrograms were needed to be certain of segmentation.

2.4.2 Rate of Speech

Measurements were taken of both vowel duration and total word length. The measurements of the entire word were used to give an approximate indication of rate of speech of each of the informants. All the informants had similar rates of speech, with the mean length of each word at 452 ms (NR=446ms, ML=459ms, NH=451ms). Accordingly, the results were not normalised.

The informants were asked to read at their normal rate of speech. It might be expected that deletion or overlap would occur in some longer sequences of consonants (e.g. [nts]→[ns]). However, this rarely occurred (probably due to the fact that word was pronounced phrase-finally with focus, with the accompanying longer duration precluding deletion/overlap).

3. Results

3.1 Single Consonant

The effects of a single consonant on vowel duration were similar to those found in other studies. The following table gives mean vowel durations of the three speakers in the context of each alveolar consonant (measurements are in milliseconds). The findings from Peterson & Lehiste (1960) and van Santen (1992) are included for comparison:[1]

	Present Study	P&L	van S.
t	160	147	177
d	212	206	259
s	200	199	195
z	275	262	293
ʧ	158	145	177
ʤ	218	191	–
n	216	216	252
l	213	218	213

[1] The data from Peterson & Lehiste is taken from their table II on p.702, in the 'Duration of short syllable nucleus' column. In their study, the vowel [ɪ] is shorter by 16ms from the average short vowel, so strictly speaking the results should be decreased by that amount. The values for van Santen (1992) are taken from his graph on p.527 (Fig.4). Given the size of the graph, these are approximate values only.

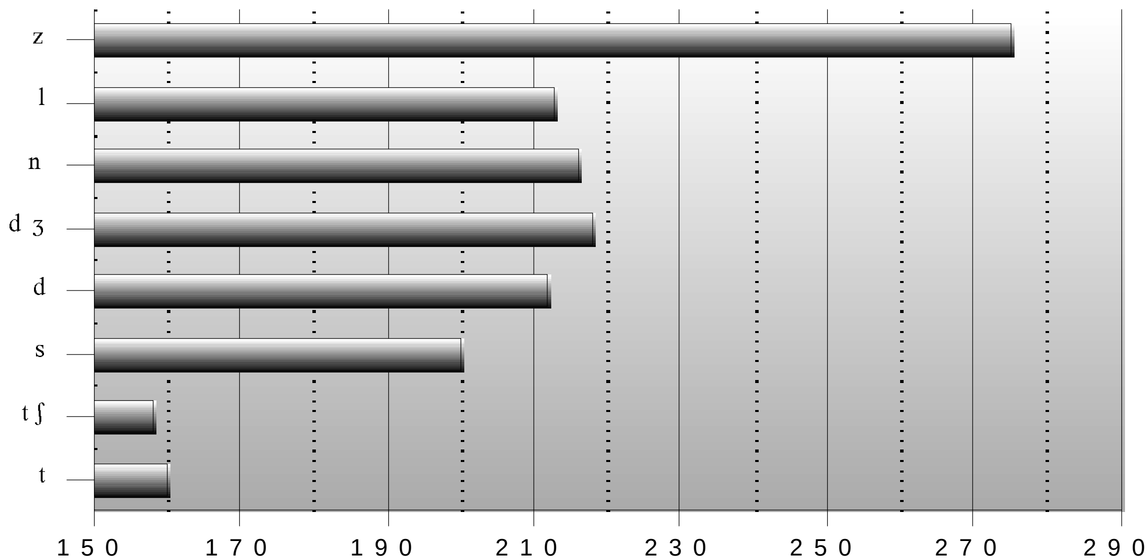

Figure 1. Vowel Duration as a function of following C

In the present study, [ɪ]'s duration before a voiceless consonant is approximately 73% of that before a voiced consonant. This is close to Peterson & Lehiste's (1960) findings (74%), but somewhat greater than those found by van Santen (67%) and others (Chen 1970 – 61%, Mack 1982 – 53%). The reason for this disparity is probably due to the fact that the figures given for van Santen, Chen, and Mack are for *all* vowels, whereas for P&L it is only for short vowels, and for the present study only for [ɪ]. From P&L's data, it is evident that there is a greater difference in *long* vowel duration in different environments than there is for short vowels. For example, a short vowel before a voiceless consonant is 71% of the its length before a voiced consonant whereas for a long vowel it is 66%. This accounts for the difference in findings from the present study and others.

From Figure 1 it is evident that both continuancy and voicing affect vowel duration.[2] The mean duration before voiceless consonants is 172ms, whereas before voiced consonants it is 226ms. Similarly, the mean duration of the non-continuants [t d ʧ ʤ n l] is 196ms, whereas for the continuants [s z] it is 237ms. The following summarises the magnitude of the effect on vowel duration by each consonant:

(1) t, ʧ < s < d, ʤ, n, l < z

'*x*<*y*' stands for 'vowel duration is shorter before *x* than *y*'. This ranking correlates with P&L's and van Santen's results relatively well:

(2) P&L: t, ʧ < s, d, ʤ, < n, l < z
van Santen: t, ʧ < s, l < d, n < z

---

[2] By 'continuancy' I refer to the feature [+continuant]. This does not include [l] (see the discussion in §4.2).

The main point of divergence is the ranking of [l] – P&L have it equal with [n], as does the present study, whereas van Santen has it ranked lower than [n], at the level of [s]. This variability is perhaps indicative of the difficulty of segmentation of [l], and the use of slightly different criteria.

The main points of commonality are in terms of voice and contiuancy: no voiceless consonant outranks a voiced one, and no non-continuant outranks a continuant with the same value for voice.

The vowel durations reported here are slightly longer than those reported by P&L (Present results, mean = 206.5ms, P&L mean = 198ms). This is probably due to the differing environments in which the test words were produced. The mean duration for van Santen's results is 223ms (with 220ms interpolated for [ʤ]). van Santen's results are longer since his are an amalgamation of durational measurements for both long and short vowels.

In summary, the results for vowel duration before single consonants accord with previous studies: vowel duration is affected by both voicing and continuancy of a following consonant.

3.2 Bi-Consonantal Clusters

3.2.1 Liquid + C

l t	l d	l s	l z	l ʧ	l ʤ	l n
161	201	157	206	161	192	218

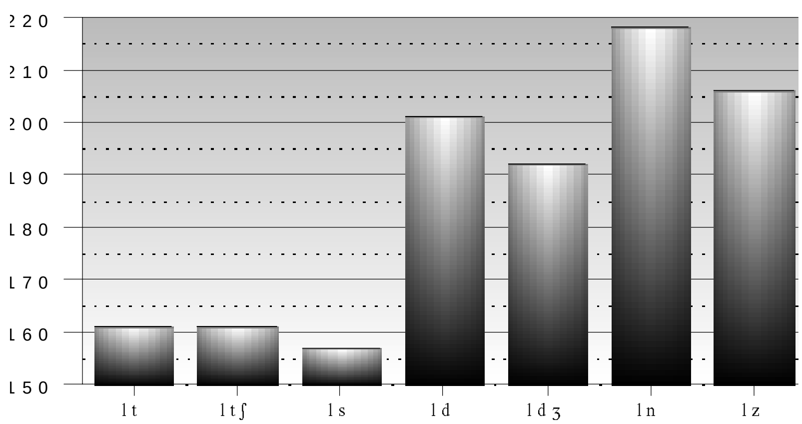

Figure 2. Vowel Duration as a function of a following LC

The table and figure above show that [l] is 'transparent' in the sense that the voicing of the following consonant can influence preceding vowel duration: a vowel is on average 159.7ms long before LC^{-VOICE} and 204.3ms before LC^{+VOICE}. However, there is no effect of continuancy: the vowel is 157ms before [ls], and 161ms before [lt] and [lʧ]; it is 206ms before [lz], and 203ms before [l{d,ʤ}]. So, there are two groups of sounds in this environment: the sounds [lt, ls, lʧ] all cause duration of approximately 160ms, while [ld, lz, lʤ, ln] cause vowel duration of approximately 205ms. This is clear from the graph, where the data appears in two distinct groups differentiated by voicing alone. Of particular note are the L+continuant clusters [ls] and [lz], which are significantly shorter than their single member counterparts [s] and [z] (cf Fig.1).

A secondary effect can be seen in the shortening of the vowel before voiced obstruents: the vowel is 201ms before [ld] compared to 213ms before [d] alone, and 192ms before [lʤ] compared to 218ms before [ʤ] alone. In comparison, duration before [ln] is much the same as before [n]. This shortening indicates that the *number* of coda consonants has some effect on duration. However, this effect is slight insofar as it is not perceptually significant (Lehiste 1970:13 identifies 10-40ms as a ‘just noticeable’ difference). Why shortening only occurs before sequences of voiced obstruents is discussed in §4.

3.2.2 NC

n t	n d	n s	n z	n ʧ	n ʤ
160	194	170	201	165	191

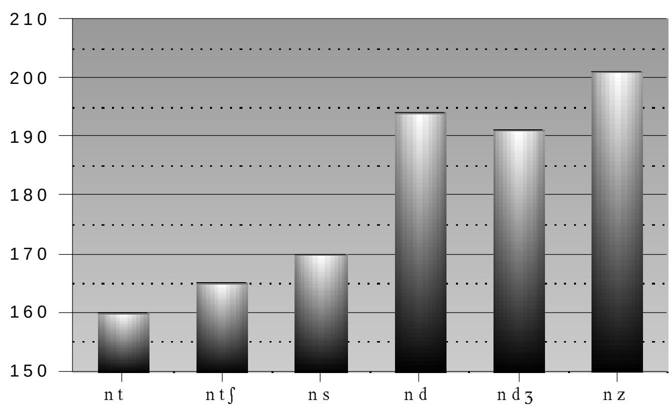

Figure 3. Vowel Duration as a function of a following NC cluster

These data show that [n] is also opaque with respect to continuancy, but transparent for voicing. There are again two groups of sounds: [nt, ns, nʧ] ranging from 160-170ms, and [nd, nz, nʤ] ranging from 194-201ms.

As with LC sequences, the duration of the vowel is ~160ms before voiceless consonants, and 200ms before voiced consonants. As with LC clusters, there is shortening due to the number of consonants: 194ms before [nd] compared with 212ms before [d] alone, 191ms before [nʤ] compared with 218ms before [ʤ] alone. Again, this shortening only applies to voiced clusters; the duration before [nt] and [nʧ] is almost identical to that before [t] and [ʧ] alone.

3.2.3 FC

s t	s	z d	z	f s	f	v z	v
188	200	247	275	182	192	204	246

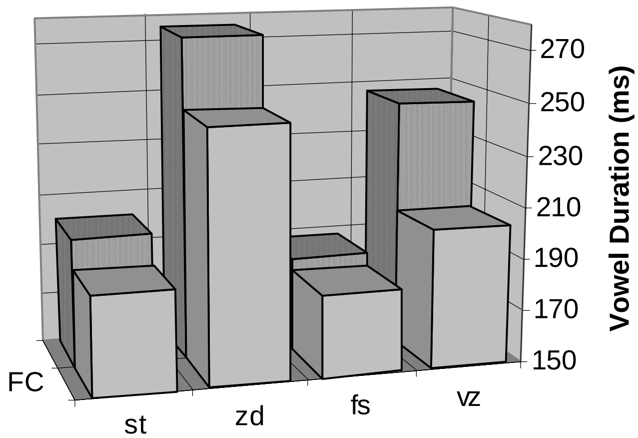

Figure 4. Vowel duration as a function of a following FC. F represents the duration of the vowel before the F alone.

All FS clusters are 10-30ms shorter than their F counterparts, as shown in the table above.

The FF clusters consist of a *labial* fricative with an alveolar stop ([fs], [vz]) since there are no clusters of two alveolar fricatives (e.g. [sz], [zs]) in English, neither are there clusters of [alveolar fricative+fricative]. Due to the introduction of the labial fricatives the single clusters consisting of [f] and [v] alone were measured. These were found to be shorter than their alveolar counterparts (for [f] 10ms shorter, and for [v] 29ms shorter). There was slight shortening of the vowel before [fs] in comparison with [f] (10ms). There was also shortening before the voiced cluster: before [vz] the vowel is 43ms shorter than before [v]. As with LC and NC clusters, the number of consonants affects duration before voiced clusters, but does not perturb it before voiceless ones.

Given the data from LC and NC clusters, it is possible that any initial consonant in a cluster is 'transparent' – i.e. that it is always the second consonant that affects duration. FC clusters show that this is not so – an FC cluster has a similar effect on duration as its F counterpart. This is shown by the FS clusters – if it is the S that determines duration, the vowel should be much shorter (i.e. for [st] at ~160ms instead of 188ms, and for [zd] at ~205ms instead of 247ms).

3.2.4 AC

ʧ t	ʧ	ʤ d	ʤ
153	158	191	218

Two Affricate+C clusters were examined: [ʧt] and [ʤd]. Before a [ʧt] cluster the mean vowel duration was 153ms; before a [ʤd] cluster it was 191ms. These compare with vowel duration before [ʧ] (158ms) and before [ʤ] (218ms). It is evident that the addition of an extra consonant has a shortening effect (of 29ms) in the voiced cluster, similar to the results for LC, NC, FC clusters.

3.2.5 SC

pt	p	bd	b	ts	t	dz	d
149	156	196	210	149	160	198	212

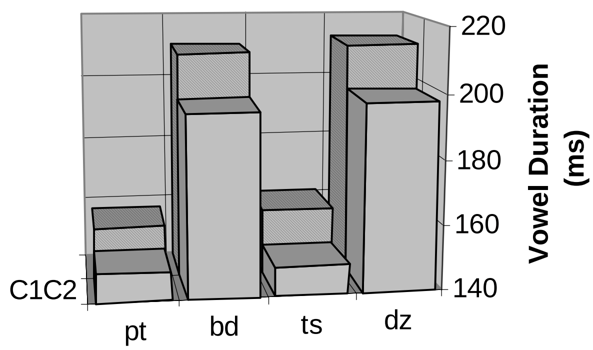

Figure 5. Vowel Duration as a function of a following SC

The results for SC clusters are analogous to those for FC clusters: duration before SC is as before the corresponding S, with shortening in voiced clusters, and slighter shortening in voiceless clusters. It is evident that stops are opaque for continuancy – duration before [ts] is analogous to [t], not [s], and likewise for [dz] and [z].

3.3 CCC Clusters

3.3.1 LCC

LCC	**Duration**	LC_1 duration	Difference (LC-LCC)
l p t	152	161	9
l b d	190	201	11
l t s	167	161	-6
l d z	192	201	9
l s t	156	157	1
l v d	200	206	6
l f s	160	157	-3
l v z	179	206	27
l ʧ t	147	161	14

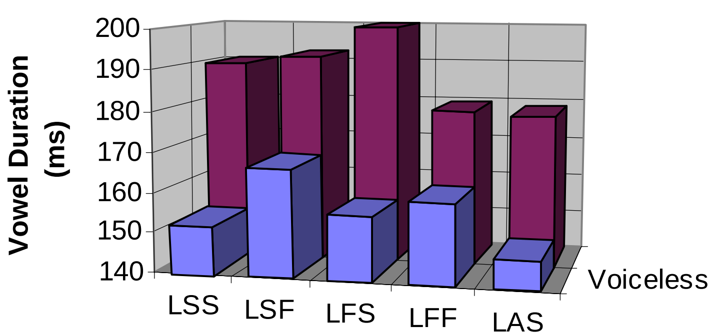

Figure 6. Vowel Duration as a function of a following LCC.

L+obstruent+C sequences induce virtually the same durational differences as their L+obstruent counterparts. For example, [lpt] causes 152ms of duration, and [lt] causes 160ms. For most clusters, there is a shortening of 9-14ms in comparison to their LC counterparts. This difference is at least not perceptually significant (Lehiste 1970:13).

However, the values for [lnC] sequences are a different matter. It has been established that [l] and [n] are transparent with respect to voicing. This means that there are two possible descriptions:

(3) (i) [l] and [n] are transparent.
(ii) C_1 in a cluster $C_1C_2(C_3)$ is transparent if C_1 is [l] or [n].

These hypotheses make different predictions with respect to an [lnC] sequence. The first predicts that there will be a significant durational difference before [lnC^{-VOICE}] and [lnC^{+VOICE}] clusters. The second predicts that [lnC] sequences will have the same effects as [ln] sequences. The results are as follows:

LN+C	Duration	L+C	N+C
l n t	161	161	160
l n d	200	201	194
l n z	198	206	201

Vowel duration before [ln] is 218ms. It is obvious that [lnC] sequences are like [lC] and [nC] clusters, and unlike [ln] since they show variation in duration depending on the voicing of the final consonant. While this supports the first hypothesis, it does not absolutely decide the case since there are no attested [lnt] sequences in English. This places the present results under some suspicion.

However, to decide the case the liquid [ɹ] can be used instead since there are attested [ɹnt] sequences in English. Accordingly, recordings of an American English speaker were made under the same circumstances and with the same equipment as described in §2. The V to [ɹ] transition was difficult to locate in many of the examples, the entire duration of the [Vɹ] sequence was used instead. Assuming that the duration of [ɹ] holds relatively constant, the differences in total duration reflect differences in vowel length. Words ending in [ɹt], [ɹd], and [ɹz] clusters were also measured so as to provide a comparison to the RNC sequences:

Word	Cluster	Vowel Duration
burr	[ɹ]	357
burt	[ɹt]	255
burred	[ɹd]	375
burrs	[ɹz]	368
burnt	[ɹnt]	241
burned	[ɹnd]	315
burns	[ɹnz]	321

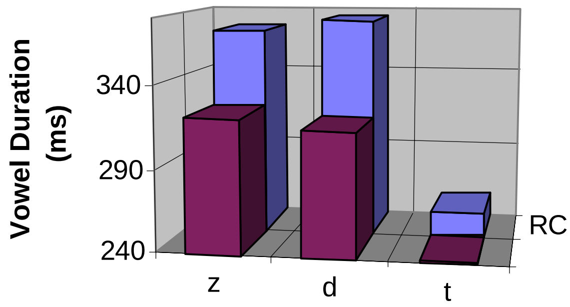

Figure 7. Vowel Duration as a function of a following [rNC], with [rC] durations for comparison.

The data accords with the hypothesis that [l] and [n] are transparent to voicing in any position, not just in C_1.

In summary, L+obstruent+C sequences cause the same durational differences as in L+obstruent sequences, with minor shortening (about 10ms). [lnC] sequences suggested that [l] and [n] are entirely transparent to voicing; this was confirmed by a further study of [rNC] sequences.

3.3.2 NCC

n t s	n d z	n s t	n z d	n θ s	n ʧ t	n ʤ d
156	176	162	179	159	153	178

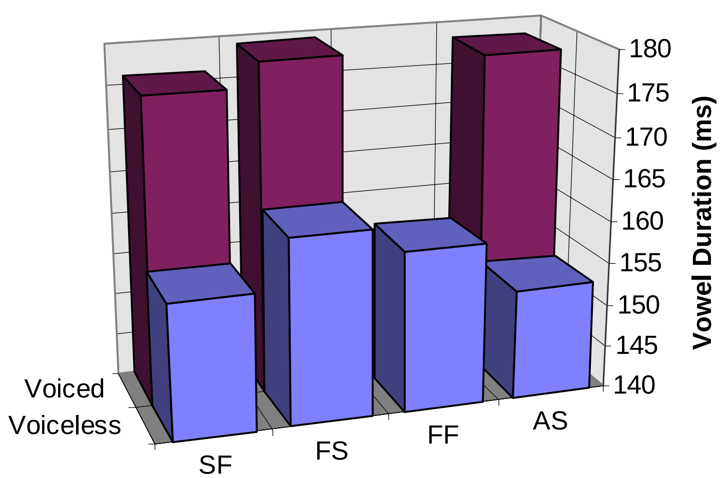

Figure 8. Vowel Duration as a function of a following NCC cluster

The results again show that [n] is transparent to voicing differences, but not to continuancy. Obstruents are similarly not transparent to continuancy (it is impossible to tell with respect to voicing since obstruent sequences must agree in voicing).

N+voiceless C+C sequences have identical durational patterns to their corresponding N+voiceless C sequences. For example, [nts] induces 156ms, while [nt] induces 160ms. There is shortening of 10-20ms with N+voiced C+C sequences in comparison to their two-member counterparts:

NCC		NC
n z d	179	201
n ʤ d	178	191

In summary, NC_1C_2 sequences induce similar durational affects as their NC_1 counterparts, with some shortening in N+voiced C+C sequences.

3.3.3 FCC

F C_1 C_2	Duration	F C_1	F	C_1
s k t	151	188	200	160
s t s	160	188	200	160
z d z	227	247	275	210
f θ s	153	182	187	–
v ð z	185	204	246	–

LC_1C_2 and NC_1C_2 sequences have durations similar to their LC_1 and NC_1 counterparts. The FC_1C_2 results differ in that durations before FC_1C_2 are not the same as before FC_1. In fact, there is a tendency for durations before voiceless FCC clusters to be ~155ms and for voiced FC_1C_2 clusters to behave like their FC_1 counterparts.

3.3.4 SCC

p s t	t s	d z d	d z	t θ s	t s	d ð z	d z
155	149	190	196	145	149	192	196

Before an SC_1C_2 cluster, [ɪ] has almost the same duration (within 6ms) as before the corresponding SC_1 sequence.

3.3.5 Summary: CCC clusters

In general, $C_1C_2C_3$ clusters induce the same vowel duration as before their C_1C_2 counterparts. However, voiceless FC_1C_2 clusters do not fit this pattern, with durations acting as before C_1, not as before FC_1. Notably, differences in the voicing of C in LNC clusters show that L and N are transparent to voicing.

4. Discussion

The results show that a number of factors that affect vowel duration:

(4) (i) Voicing of C in an [(S)C…] cluster, where S is a sequence of sonorants.
(ii) Continancy of C_1 in a $[C_1(C_2)(C_3)]$ cluster.
(iii) Number of voiced consonants.

Both voicing and continuancy have a (perceptually) significant effect on vowel duration. The number of consonants also has a shortening effect of 10-20ms, though this only applies before voiced clusters and voiceless FC(C) clusters. The next section will formulate a descriptive model of vowel duration. The final section will offer some tentative explanations as to why durational differences exist in the first place.

4.1 Descriptive Model of Vowel Duration

There is no *a priori* reason to choose between formulating an additive or subtractive descriptive model of vowel duration. Klatt (1976) opts for the latter. Neither option will be taken here. Instead, the notion of 'target' will be employed.

When continuancy is eliminated as a factor in C, LC, and NC clusters, vowel durations tend to group around a constant value before voiced and voiceless consonants. Before voiced consonants, that value is ~215ms; before voiceless consonants it is ~160ms. The value before voiced consonants (i.e. 215ms) will be assumed to be the more basic; it will be termed the *target*. Contextual variation in vowel duration can be described in terms of this target:

(1) Continuancy:
In a [C_1(C)] sequence, add 50ms to the target iff C_1 is [+continuant].

(2) Voicing:
In a [C(C)(C)] sequence, if any of the C's are [-voice] subtract 60ms.

(3) Number of Segments:
In any entirely voiced CC(C) sequence (i.e. where all members are voiced), subtract 10-20ms.

This produces the duration variations shown in the results with a small margin of error (210ms). For example, [z] is [+continuant] so by (1) 50ms is added to the target, giving a duration of 265ms (actual duration = 275ms); no other rules apply to it. For [d], no rules apply at all. Since [t] is voiceless, rule (2) applies giving a value of 155ms. Two rules apply to [s]: 50ms is added to the target since [s] is [+continuant], giving a total of 265ms. However, [s] is also [-voice] so 60ms are subtracted, giving a total of 205ms.

The fact that [l] and [n] are transparent to voicing is captured in rule 2 by stipulating that if *any* C in a cluster is [-voice] the target is to be shortened. This means that the rule applies in LC(C) and NC(C) clusters where C is [-voice]. It also applies in CC(C) obstruent clusters, but this is not undesirable since in such clusters all consonants agree in [voice].

That [l] and [n] are opaque to continuancy is expressed in rule (1): only the first consonant can induce lengthening if it is [+continuant]. Rule (1) also captures the fact that FC_1C_2 clusters do not induce the same duration effects as FC_1 clusters: Rule 1 does not apply to tri-consonantal clusters, but only to mono- and bi-consonantal sequences. This means that only the voicing rule applies to FCC clusters, so that they cluster around the target values, as is the case.

The final rule relates to the number of segments in the consonant cluster. If the cluster has more than one member, vowel duration is decreased by 10-20ms. This is generally matched by the data. For example, LC^{+VOICE} sequences have a mean duration of

204ms, 11ms less than the target value. NC^{+VOICE} sequences are also on average 20ms shorter than the target value. For FC sequences, [zd] is lengthened by rule (1) to 265ms, then shortened by rule (3) to 245ms (actual value = 247ms). This rule also predicts the duration of AC, SC, LCC, NCC, and SCC clusters well.

However, some problems remain. Shortening also seems to occur in FC^{-VOICE} sequences – they are on average 20ms shorter than F. In addition, there is no evidence of shortening before the cluster [ln]. These issues will be addressed in the following section.

The model presented above is not complete since it does not take other factors into account (e.g. stress, position in word, number of syllables in word, etc.) and is only based on data from the vowel [ɪ]. It should also be expressed as a percent change model so as to take into account rate of speech; again, the limitations of this study preclude precise development of such a model (*cf* Klatt 1976). Even so, the descriptive model makes it clear that variations in vowel duration are systematic, and can be described with reference to only a few factors. However, this is only a *descriptive* model – I do not claim that it reflects actual cognitive process. The next section will address this issue.

4.2 Explaining Vowel Duration

Before attempting to explain the nature of the cognitive process involved in determination of vowel duration, it must be established that cognition does actually have something to do with the phenomenon. The alternative is that vowel duration is a by-product of physiological factors.

This latter possibility is eliminated by the cross-linguistic variation in vowel production observed by Chen (1970), Mack (1982), and Laefur (1992). These researchers found that shortening of vowels before voiceless consonants was more pronounced in English than in other languages (esp. French). If only physiological factors determined vowel duration, there should be no difference between English and other languages.

Vowel shortening is, then, determined at least partially by cognitive factors. Crucial to the following explanation is the notions of underspecification – the lack of a predictable phonological feature on a segment. Following Keating (1988), I assume that some segments may remain underspecified in the phonetic component of the grammar. In the case at hand, sonorants have no specification for [voice] since [voice] is non-contrastive in sonorants. However, obstruents are specified for [voice] since it serves as a phonologically distinctive feature (e.g. /t/ vs /d/, /s/ vs /z/). The feature [continuant] is specified in both sonorants and obstruents. Only fricatives are [+continuant]; all other sounds including [l] and [n] are [-continuant] (Chomsky & Halle 1968:318).[3]

Unlike the descriptive model, the duration before [z] (i.e. 275ms) will be taken as the target vowel duration. With underspecification, variation due to voicing and continuancy can be described in the following manner:

(5) (1) Subtract 65ms before [-continuant]
(2) Subtract 60ms before [-voice].

---

[3] Chomsky & Halle note that the status of [l] with respect to [continuant] is variable from a phonological point of view. [l] does require *some* obstruction of the oral cavity (or at least contact with the mid-saggital region), as compared to 'true' continuants.

The features [-continuant] and [-voice] are taken to be articulatory targets. A vowel is shortened if the articulatory target [-continuant] and/or [-voice] *immediately* follows it. A target *immediately* follows a vowel if there is no intervening contradictory target – i.e. [+continuant] or [+voice] respectively. Since some segments are underspecified for these features, those segments may be 'transparent' with respect to voicing.

An [ɪlnt] sequence will serve as an example: there are two segments intervening between the [ɪ] and the [t], yet [t]'s [-voice] feature shortens the vowel. This is due to the fact that the [t]'s [-voice] feature is immediately adjacent to the vowel – there is no intervening [voice] feature since [l] and [n] are underspecified.

In comparison, every consonant is specified for [continuant]. So, there will be no segmental 'transparency' as there is for [l] and [n]. For example, in the sequence [ɪst] the [-continuant] feature of the [t] does not induce shortening of the vowel since there is an intervening [+continuant] target belonging to [s].

While this offers a potential account of the transparency of [l] and [n] to [-voice], it still does not explain *why* there should be shortening in the first place. In addition, it must be explained why shortening happens before *voiceless* consonants but not before *voiced* ones, and before *non-continuants* but not before *continuants*.

This can be explained if variation in vowel duration is seen as being caused by 'articulatory overanticipation'. In other words, the gesture for an articulatory target (i.e. a feature) is made 'too soon' (i.e. the tongue is raised too soon for [-continuant], the glottis is is aducted? too soon for [-voice]). This early realisation of the target effectively shortens the vowel. This assumes that there are specified units of timing (probably syllabically based, e.g. Hubbard 1994).[4] If there were no such 'temporal reference points' the notion of overanticipation would be meaningless.

There are three potential problems with this account:

1. Why are <u>vowels</u> shortened in VLNC/VLC/VNC sequences (C is [-voice])? Why isn't just the L or N shortened?

A possible solution to this is that the L and N *cannot* be shortened – their inherent duration is already at its minimum. This has a 'domino' effect on duration, forcing the vowel to be shorter. This solution carries with it some notion of 'target duration of the syllable', where the syllable can take a specific duration, and its segments must fit inside that duration (e.g. Hubbard 1994). For example, suppose that a [VNC] sequence is assigned a duration of 6*x*, where *x* is a unit of duration. Usually, each segment would span 2*x*. However, overanticipation of the C makes it effectively span 3*x* (or at least begin one *x* before it 'should'). This leaves 3*x* for both the V and the N. The N cannot be compressed below 2*x*, so the V is left with 1*x*, effectively shortening it.

2. Why aren't [+voice] or [+continuant] realised 'too soon', shortening the V?

[4] For example, a syllable is assigned a certain temporal duration. This duration is not a constant value in English, but is rather a function of the total number of consonants in the syllable. See below for further discussion.

The vowel is already voiced, so there is no sense in which the target for [+voice] can be realised 'too soon' with respect to the vowel. A similar explanation can be given for [+continuant] – the jaw is already opened for the vowel, and [+continuant] continues this openness. [+continuant] is more problematic than [+voice] though, since [+continuant] does have a distinct realisation to the [+approximant] feature of the vowel.

3. Why doesn't [-continuant] encroach on [+continuant] consonants, shortening them in [st] and [zd] sequences?

This is where underspecification is crucial. [-continuant] cannot be realised if a contradictory specification is in force. In [st], [s] has a [+continuant] feature. So, for the [-continuant] feature of the [t] to be realised 'over' the [s] it would conflict with the target ([+continuant]) specified by the [s]. So, [-continuant] (and [-voice]) can only encroach on another sound if that sound is unspecified for that feature. Since vowels are unspecified for both these features, as are vowels, then they will be affected by articulatory overanticipation.

This raises an issue with regard to the form of underspecification: if vowels are not specified as [+voice], then why are they voiced at all? To answer this, I must suppose that redundancy rules operate at the phonetic level. So, if something is [+sonorant], it will be realised as voiced by stipulation of a phonetic rule.

4.2.1 Number of Consonants

It must also be explained why it is that the *number* of consonants has an effect on vowel duration. In general this effect is small (10-20ms), but it is systematic. The variation can be explained if it is assumed that each *syllable* is assigned some specific duration. Addition of segments, then, causes shortening since there is less duration left for the vowel.

However, this shortening does not take place indefinitely. If a vowel reaches its minimal duration, it cannot be shortened any further (Klatt 1973). For [ɪ] that minimum duration seems to be between 150 and 160ms. This explains why shortening does not affect a vowel before a sequence of voiced stops: it is already at its minimum duration, so cannot be shortened any further. This explanation predicts that sequences of voiced stops should induce shortening, as they do. However, it also predicts that sequences of voiceless fricatives should cause vowel shortening. This is due to the fact thtat vowel duration before a voiceless fricative is not at its minimum. This prediction is borne out in the results, with duration before FC^{-VOICE} clusters ~12ms shorter than their corresponding F counterparts.

This solution also predicts that vowel duration should become increasingly shorter with longer and longer sequences of consonants. So, duration before a CCC sequence should be shorter than before a CC sequence. In fact, this is the case: the mean duration before CC^{+VOICE} cluster is 203ms, whereas before CCC^{+VOICE} it is 191ms. Since vowel duration is already at its minimum before voiceless sequences there should be no significant shortening, and there is not (CC^{-VOICE}=149ms,CCC^{-VOICE}=153ms).

This suggests that each consonant in a cluster shortens vowel duration by 10-20ms. This may also explain why FC_1C_2 clusters are shorter than their FC_1 counterparts: For example, duration before [s] is 200ms. With [st] it is 188ms, since the extra consonant shortens the vowel. The addition of another consonant (i.e. [sts]) should shorten it further, to 168-178ms. In this context, the vowel was found to be 160ms.

4.2.2 Problems

Much remains to be explained about the proposals above. For example, why is it that overanticipation occurs at all? There is a more serious problem with the proposal, however: there is nothing inherent in the explanation that predicts that the vowel shortening should be *additive*. To elucidate this point, a [-continuant] segment shortens a vowel by 65ms, and a [-voice] segment shortens it by 60ms. However, why should a [-continuant, -voice] segment shorten duration by 125ms (60ms+65ms)? If shortening is really overanticipation of a target and features are independent, then a [-continuant, -voice] segment should shorten a vowel by 65ms – i.e. its overanticipation of [-voice] should *overlap* its overanticipation of [-continuant].

This suggests that the duration before [z] is not the target value. An explanation that takes ~215ms as the target (as in the descriptive model) seems to be more correct.[5] However, this means that there is no shortening before [-continuant], but rather lengthening before [+continuant] – a type of 'underanticipation'. Evidently a more precise notion of over- and underanticipation needs to be formulated.

PAUL DE LACY
14 May 1998

References

Chen, Matthew. 1970. "Vowel length variation as a function of the voicing of the consonant environment." *Phonetica* 22:129-159.

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English.* Cambridge, MA.:MIT Press.

Crystal, Thomas H. & Arthur S.House. 1988a. "Segmental durations in connected-speech signals: Syllabic stress." *Journal of the Acoustical Society of America* 83.4:1574-1585.

Crystal, Thomas H. & Arthur S.House. 1988b. "The duration of American-English vowels: an overview." *Journal of Phonetics* 16:263-284.

Crystal, Thomas H. & Arthur S.House. 1988c. "The duration of American-English stop consonants: an overview." *Journal of Phonetics* 16:285-294.

Crystal, Thomas H. & Arthur S.House. 1988d. "Segmental durations in connected-speech signals: Current results." *Journal of the Acoustical Society of America* 83.4:1553-1573.

[5] DiSimoni (1974) argues from language acquisition data that vowels *lengthen* before a [+continuant], not shorten.

DiSimoni, F.G. 1974. "Evidence for a theory of speech production based on observations of the speech of children." JASA 56: 1919-1921.

Fischer-Jørgensen, E. 1968. "Les occlusives francaises et danoises d'un sujet bilingue." *Word* 24:112-153.

Hubbard, Kathleen. 1994. *Duration in Moraic Theory*. PhD(UCB).

Klatt, Dennis H. 1973. "Interaction between two factors that influence vowel duration." *Journal of the Acoustical Society of America* 54.4: 1102-1104.

Klatt, Dennis H. 1976. "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence." *Journal of the Acoustical Society of America* 59:1208-1221.

Keating, P.A. 1988. "Underspecification in Phonetics." *Phonology* 5: 272-292.

Laefur, Christiane. 1992. "Patterns of voicing-conditioned vowel duration in French and English." *Journal of Phonetics* 20:411-440.

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, MA.: MIT Press.

Mack, Molly. 1982. "Voicing-dependent vowel duration in English and French: Monolingual and bilingual production." *Journal of the Acoustical Society of America* 71.1:173-178.

Peterson, Gordon E. & Ilse Lehiste. 1960. "Duration of Syllable Nuclei in English." *Journal of the Acoustical Society of America* 32.6:693-703.

van Santen, J.P.H. 1992. "Contextual effects on vowel duration." *Speech Communication* 11:513-546.

Appendix 2: Word List, with attested words

1. C

p	t	b	d	f	s	v	z	ʧ	ʤ	n	l
tip	pit	bib	kid	tiff	kiss	live	tizz	titch	ridge	tin	till

2. LC

l t	l d	l s	l z	l ʧ	l ʤ	l n
tilt	tilled	pulse	tills	filch	bilge	kiln

3. NC

n t	n d	n s	n z	n ʧ	n ʤ
tint	tinned	prince	tins	pinch	tinge

4. AC

ʧt	ʤd
bitched	budged

5. FC

fs	vz	st	zd
tiffs	lives	kissed	whizzed

6. SC

pt	bd	ts	dz
tipped	ribbed	kits	kids

7. LCC

l p t	helped
l b d	bulbed
l t s	tilts
l d z	welds
l s t	pulsed
l v d	shelved
l f s	shelfs
l v z	shelves
l ʧ t	belched
l ʤ d	bulged
l n t	(see §3.3.1)
l n d	filmed
l n z	kilns

8. NCC

n t s	n d z	n s t	n z d	n θ s	n ʧ t	n ʤ d
tints	winds	winced	bronzed	tenths	pinched	tinged

9. FCC

skt	sts	zdz	fθs	vðz
whisked	mists	housed’s	fifths	twelvths

10. SCC

p s t	d z d	t θ s	d ð z
lapsed	adzed	eighths	breadths