Phonological ‘voicing’, phonetic voicing, and assimilation in English*

Wouter Jansen
Division of Speech and Language Therapy and Rehabilitation
UCE Birmingham
Wouter.Jansen@uce.ac.uk

July 10, 2007

Abstract

This article investigates certain aspects of regressive voicing assimilation by means of a quantitative acoustic study of British English obstruent clusters. It is found that, the phonologically voiceless obstruents /t, s/, and contrary to many impressionistic descriptions, the phonologically voiced sounds /z/ and to some extent, /d/, trigger certain forms of voicing assimilation in preceding obstruents. The precise patterning of the phonetic effects of the processes involved suggests that they should be modelled in terms of the coarticulation of the gestures that underpin the acoustic correlates of phonological voicing contrast.

1 Introduction

It is well-known that the phonetic interpretation of what is sometimes loosely referred to as ‘voicing contrast’ is subject to positional, dialectal, and crosslinguistic variation. For example, utterance initially and in post-obstruent contexts, many dialects of English contrast a series of long lag VOT (conventionally > 35 ms) or

---

[*] This is the author’s version of a work that was accepted for publication in Language Sciences. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Language Sciences 29, Issues 2-3, March-May 2007, pages 270-293 DOI: 10.1016/j.langsci.2006.12.021.

‘voiceless aspirated’ plosives [pʰ, tʰ, kʰ] with a series of plosives that sometimes have a negative VOT, but more often have a short lag positive VOT (< 35 ms). Word medially, the first series retains its long lag VOT in the onset of stressed syllables, but in the onset of unstressed syllables the amount of VOT often decreases to within the short lag bracket. The second series may be partially or wholly voiced in medial context, depending, among other things, on the phonetic context. Languages such as French or Polish on the other hand, contrast a series of short lag VOT stops with a series of negative VOT or ‘prevoiced’ stops across initial and medial context. (see Keating (1984) for one of the principal instrumental studies on this topic).

The variable interpretation of ‘voicing contrast’ has driven linguists to draw a terminological distinction between ‘phonological voicing’ (sometimes referred to as ‘big’ VOICE or represented by a binary phonological feature [±voice]) and ‘phonetic voicing’, i.e., vocal fold vibration and/or its acoustic and perceptual manifestations (sometimes referred to as ‘small’ voice). From this perspective, both English and French have a series of [-voice] or VOICELESS plosives, which might be symbolised as /p, t, k/, and a series of [+voice] or VOICED plosives, which might be represented as /b, d, g/. English and French are then said to differ as to their use(s) of phonetic voicing to implement these categories. Thus, French implements utterance-initial [-voice] stops as [p, t, k] and their [+voice] counterparts as [b, d, g], whereas English expresses the same categories as [pʰ, tʰ, kʰ] and [b̥, d̥, and g̊] respectively.[1]

Interdialectal and crosslinguistic variation in the phonetic implementation of voicing contrasts is of importance to the study of voicing assimilation (VA) as it has been observed that [+voice] stops only trigger regressive voicing assimilation (RVA) in languages that employ the ‘French-type’ negative VOT imple-

---

[1] Work on laryngeal phonology along the lines of Harris (1994); Iverson & Salmons (1995, 1999) (on which more below) rejects the use of a single binary feature [±voice] in formal accounts of two-way VOT/voicing contrast in obstruent systems. It replaces [±voice] by two monovalent features, which yield three laryngeal categories that roughly mirror the three VOT categories identified by Lisker & Abramson (1964). The laryngeal unmarked class defined by this system is assigned to the short lag stops of both ‘English-type’ systems, where they are traditionally regarded as [+voice], as well as ‘French-type’ languages, where they represent the traditional [-voice] category. Regardless of the merits of this sort of analysis at the phonological level, it is too coarse-grained at the phonetic level because ‘English-type’ short lag stops are phonetically distinct from ‘French-type’ short lag stops. In fact, as shown by for example Kingston & Diehl (1994), ‘English-type’ short lag are phonetically similar in all respects but their voicing (in some contexts) to the negative VOT stops of ‘French-type’ languages. Similarly, ‘French-type’ short lag stops pattern with the long lag stops of ‘English-type’ languages in all respects except their voicing (in some contexts). This means that the classes defined by traditional [±voice] represent phonetic categories of obstruents, even if full linguistic phonetic descriptions of such categories will also have to appeal to the three conventionally recognised VOT categories, and most likely a more finely-grained set of VOT/voicing classes (e.g., Cho & Ladefoged 1997).

mentation (e.g., Kohler 1979; see Jansen 2004, chapter 4 for a more detailed discussion). This observation is illustrated by the distribution of RVA to [+voice] plosives across the languages and dialects of Germanic. Within this group only the varieties employing a 'French-type' implementation of [±voice] are reported as exhibiting RVA to [+voice] plosives: Afrikaans (Wissing, 1991), (Western and Southern) Dutch (e.g., Cohen et al. 1972), Scottish English (Kohler, 1979; Wells, 1982), (West) Frisian (Riemersma, 1979; Tiersma, 1985), Rhineland German (Kohler, 1979), and (Eastern varieties of) Yiddish (Katz, 1987; Jacobs et al., 1994).

This apparent phonetic conditioning of regressive voicing assimilation to [+voice] plosives raises a number of questions with regard to the nature of the process. An important empirical question is whether RVA is conditioned by the (canonical) phonetic interpretation of obstruents more generally, or whether this phonetic conditioning is a more restricted phenomenon, e.g., limited to specific phonetic or phonological categories. A part of this broader issue that would seem to be particularly interesting is the behaviour of [+voice] *fricatives* in languages such as English. Unlike the corresponding plosives, these fricatives tend to be produced more or less voiced in utterance-initial contexts. If, as is suggested by the behaviour of [+voice] plosives, RVA to [+voice] obstruents is conditioned by the presence of phonetic voicing, this would imply that the [+voice] fricatives of English and similar languages are able to trigger regressive voicing assimilation.

A theoretical issue that is raised by the assimilatory behaviour of [+voice] plosives is how its (apparent) dependence on the role of phonetic voicing should be accounted for in a formal framework. At least two approaches to this issue have been proposed in the literature. One is to assign distinct (lexical) phonological representations to the two types of [+voice] stops and to make the rule responsible for RVA sensitive to this distinction (Harris, 1994; Iverson & Salmons, 1995, 1999). The second approach cuts out all phonological mediation by viewing RVA as the coarticulation of (the gestures underpinning) phonetic voicing categories in obstruent sequences (cf. Slis 1985; Ernestus 2000; Jansen 2004). Because it models the process of regressive voicing assimilation at the articulatory phonetic level, differences in the phonetic implementation of [+voice] (and [-voice]) have a direct and automatic effect on the occurrence and phonetic manifestation of the process under this alternative approach.

This paper addresses aspects of the empirical and theoretical issues sketched above using an experimental acoustic study of English obstruent clusters. It shows that at least two phonetic correlates of phonological [voice], viz. phonetic voicing and the perturbation of F_1 on preceding vowels, are subject to a form of regressive assimilation. Since the remaining correlates, and in particular preceding vowel duration, do not assimilate, this process must be regarded as non-neutralising. In addition, the results of the experiment provide partial support for the hypothesis that RVA is phonetically conditioned as it is found that /z/, but not /d/, influences the phonetic voicing of a preceding obstruent.

The theoretical upshot of this study is rather dependent on whether the phenomena observed below are considered to be within or outside the scope of phonological models of voicing assimilation. Given the fact that the form of RVA uncovered here does not affect vowel duration and given that the process is non-neutralising more generally, it seems likely that many phonologists would consider it to be a 'low-level phonetic' process that lies outside the scope of phonological theory.

Nevertheless, the results of the experiment need to be accounted for in some way. The key argument for the account I will sketch in section 5 rests on the observation that there appears to be a direct connection between the phonetic features of assimilation triggers and the phonetic features they are able to influence in preceding sounds. This suggests that the assimilation process(es) involved should be modelled in terms of coarticulation of the articulatory gesture that underpin the acoustic correlates of the lexical phonological feature [±voice].

Before proceeding with a survey of the descriptive literature on English voicing assimilation, a minor terminological point is in order. To avoid all confusion of 'big'(phonological) and 'small' (phonetic) voicing, I have used the terms *tense* and *lax* and the feature [±tense] to refer to the former. Thus, the 'phonologically voiceless' obstruents of English, as well as, e.g., French are classified as *tense* or [+tense], whilst the obstruents often known as 'phonologically voiced' are described here as *lax* or [-tense]. Note that this is intended solely to keep the distinction between phonological and phonetic categories maximally clear; it is certainly not meant to imply that tense and lax are useful concepts in dealing with phonetic substance.

2 Previous descriptions of the phonetics of English obstruent clusters

Impressionistic accounts usually describe standard varieties of English as possessing little or no regressive voicing assimilation across word boundaries. Indeed, Jones (1956) warns native speakers of French and Dutch against making the mistake of applying RVA in English obstruent clusters (e.g., Jones's §851), a point which is echoed by Gimson (1994).

More importantly for present purposes, with a few exceptions, it is generally assumed that only tense obstruents are capable of triggering voicing assimilation in English. For example, Gimson (1994) claims that at the boundaries of 'close-knit' groups of words, lax fricatives (but normally not lax stops) may devoice completely when preceding a tense obstruent. Interestingly, according to Gimson, this form of RVA may also shorten the preceding vowel, although this phenomenon is judged to be "relatively rare" (Gimson 1994:257). This description of RVA to English tense fricatives is reminiscent of the account of a regressive devoicing rule found in Yorkshire dialects of English provided by Wells (1982). According to Wells's description, this rule is triggered by all tense obstruents and affects voicing as well as durational correlates of [tense].

A further well-documented example of regressive voicing assimilation in English concerns the behaviour of function words ending in lax fricatives, which often assimilate in voicing to a following tense obstruent. For example, post-pausal and post-sonorant weak forms of <is> are usually realised with [z] before lax obstruents and sonorants, but with [s] before tense obstruents (Lakoff, 1972; Selkirk, 1980):.

(1) Voicing of weak <is> in English (examples from Selkirk 1980)

Orthography	Phonetic form
<Is Jack going?>	[(ɪ)zd̥͡ʒ̊ækg̊oʊɪŋ]
<Is Will going?>	[(ɪ)zwɪɫg̊oʊɪŋ]
<Is Pete going?>	[(ɪ)spʰiːtg̊oʊɪŋ]

When instances of regressive assimilation to *lax* obstruents are mentioned in impressionistic accounts, these almost invariably involve word internal contexts. One instance is the voiced pronunciation of plosive + alveolar fricative clusters in words containing orthographic <x> such as <exam>, [ɪgzæm]; <exhibit>, [ɪgzɪbɪt]; <excerpt> [ɛgzɜːpt]. These clusters presumably originate from an older (and invariant) form [ks] which was subsequently affected by the process of [s]-voicing that is also responsible for the pronunciation of e.g., <desire> as [d̥ɪzaɪə]. According to Borowsky (2000) the (historical) voicing of word-medial prestress /s/ in turn triggers voiced realisations of the preceding /k/. The same author discusses a second example of apparent word internal RVA in English: the optional voiced realisation of the final alveolar fricative of prefixal <dis> before lax stops, as in [dɪzg aɪz] for <disguise>.

However, a number of additional observations cast doubt on the idea that the optional voicing of the medial fricative in <disguise> is correctly attributed to a process of assimilation. Perhaps the most important of these observations is the following. In word-medial, pre-consonantal contexts the contrast between [±tense] fricatives is almost non-existent in English: near-minimal-pairs such as <saffron>-<chevron> are exceedingly rare. Non-alveolar fricatives are generally voiceless in this context, and all fricatives are voiceless when the following sound is a tense obstruent. However, alveolar sibilants (orthographic <s>) tend to be voiced before both lax obstruents and sonorant consonants (2).

(2) Realisation of English medial alveolar sibilants

a. followed by a sonorant:

<osmosis>	[ɒzmoʊsɪs]	<Bosnia>	[b̥ɑznɪə]
<Oslo>	[ɒsloʊ][ɒzloʊ]	<gosling>	[g̊ɒzlɪŋ]
<Israel>	[ɪzrɛl]	<Bosworth>	[b̥ɒzwɜθ]

b. followed by a lax plosive:

<Neasden>	[niːzdən]	<Marsden>	[mɑːzdən]
<Osborne>	[ɒzbɔn]	<Osgood>	[ɒzgʊd]
<wisdom>	[wɪzdəm]	<Lisbon>	[lɪzbən]

The voicing of alveolar sibilants in this context might be treated as an assimilation process but there is a number of reasons not to pursue this treatment. The first is that the process almost exclusively applies to alveolar sibilants (cf. <Lisburn> and <Oslo> with [z] vs. <Fishburne> with [ʃ] and <athlete> with [θ]) whilst regressive voicing assimilation processes are generally rather inclusive in their choice of targets. Second, there are voicing processes that specifically target alveolar sibilants in contexts where laryngeal contrast is suspended and outside what are normally regarded as assimilation contexts. The voicing of word-initial prevocalic alveolar fricatives in (standard) German and in certain dialects of Dutch is an example. Third, in the absence of contrastive voicing on sonorants current phonological theory generally treats these sounds as (surface-)underspecified for laryngeal structure and therefore incapable of triggering voicing assimilation. Fourth, (in standard varieties of) English lax stops do not trigger RVA under word sandhi.

It is perhaps better, therefore, to interpret the voicing of pre-consonantal alveolar sibilants in English as the reflex of a relatively general, non-assimilatory process, which is blocked by a following tense obstruent on assimilatory grounds. This interpretation is bolstered by the observation that devoicing of lax fricatives before tense obstruents is attested independently for English (cf. 1 above).

Perhaps because the impressionistic literature depicts voicing assimilation under word sandhi as a marginal phenomenon in English, instrumental investigation of English obstruent clusters with mixed underlying [tense] specifications has been limited. However, the (quantitative) instrumental evidence that is available in the literature contains some clues suggesting that the picture sketched above may be incomplete.

The quantitative (acoustic) study of laryngeal contrast and voicing in American English fricatives conducted by Stevens et al. (1992) shows a clear effect of following context on the voicing of tense and lax fricatives. For example, in their corpus lax fricatives (/v, z/) have on average 29 ms of voicing preceding a tense fricative (/f, s/), which increases to 58 ms before another lax fricative or a vowel. Unfortunately, since Stevens et al. (1992) do not provide separate means for fricative + vowel sequences and homogeneous (tense + tense and lax + lax) clusters, there is no baseline measure to determine whether the observed differences in voicing stem from assimilation to the tense fricatives or from assimilation to the lax fricatives, or from both. Furthermore, Stevens et al. (1992) do not provide tests of the statistical significance of the differences in the mean voicing values they observe.

An acoustic study of English voicing assimilation by Myers (2002) on the other hand, does allow for the effects of tense and lax contexts to be teased apart, and provides statistical tests. Interestingly, Myers finds that RVA patterns symmetrically, with both tense and lax obstruents influencing the voicing of preceding obstruents. However, whilst the set of potential assimilation triggers in Myers's stimulus set includes a lax fricative (/z/) as well as a lax stop (/g/), his (statistical) analysis does not distinguish between these two environments. Consequently, it is impossible to address the question regarding the phonetic conditioning of RVA that was raised in the introduction above.

Table 1: Voicing duration (ms) and ratio of the closed phases of unstressed English /t/ and /d/ followed by a C_2 in the onset of a stressed syllable, as reported by Thorsen (1971).

C_1C_2	C_1 voicing		$C_1 + C_2$	C_1 voicing	
	Duration	Ratio		Duration	Ratio
/tl/	35	0.53	/dl/	56	1.00
/tk/	31	0.66	/dk/	44	0.86
/tg/	41	0.87	/dg/	51	0.91
/ts/	33	0.60	/ds/	50	0.79
/tz/	64	0.82	/dz/	62	0.96

However, an early and all but forgotten study by Thorsen (1971) (and one that I have only become aware of after most of the work reported below had been completed) does shed light on precisely this matter. Thorsen's study investigates voicing and other phonetic features of [tense] in English alveolar stop C_1 + consonant C_2 sequences straddling word and morpheme boundaries in three different prosodic contexts, and crucially, reports measurements for sonorant C_2s separately from measurements for stop and fricative C_2 contexts. The latter feature allows for the effects of tense and lax obstruents to be assessed against a 'neutral' baseline environment. Some of the mean values reported by Thorsen are represented in table 1.

The data in table 1 provides a number of interesting insights in the phonetics of the tense-lax contrast in English word-final stops. First, there seems to be a clear voicing contrast, both in absolute terms (21 ms) and in relative terms between /t/ and /d/ before sonorant /l/. Second, the absolute and relative extent of this contrast decreases in all other contexts: most so before /z/ and least so before /s/, which is what would be expected if voicing were subject to a symmetric assimilation process. Third, and perhaps most important, are the ways in which the voicing of C_1 consonants deviates from the baseline contexts in 'mixed' clusters. For example, before /z/ the voicing duration of /t/ increases by 19 ms relative to the baseline, whilst there is a 12 ms decrease in the voicing of /d/ when preceding /k/. These observations are consonant with impressionistic descriptions insofar as they support the idea that tense obstruents are able to trigger a certain amount of devoicing in a preceding lax obstruent, but go beyond such descriptions in suggesting that /z/ is able to trigger assimilation too.

Furthermore, from the perspective of this paper, it might be significant that, relative to the baseline environment, the effect of /g/ on the voicing of a following /t/ appears to be considerably smaller than the effect of /z/. This could be taken as evidence that regressive voicing assimilation is indeed conditioned by the voicing of the sounds that trigger the process. However, the fact that there also is a differential effect of /s/ and /k/ on the voicing of a preceding /t/, where this would not seem to be expected on phonetic grounds, casts doubt on such an interpretation.

A final interesting aspect of the study by Thorsen (1971) is that it finds no effects of assimilation on the duration of C_1 consonants or on the duration of the vowels preceding them. In fact, the latter parameter, generally regarded as the most important cue to the tense-lax contrast in English postvocalic obstruents, perfectly preserves underlying distinctions between /t/ and /d/.

3 Methods

3.1 Subjects

Subjects were 4 native speakers (2 male, 2 female) of British English aged between 24 and 35, and living in or near to London at the time of recording. None of the subjects had a history of speech or hearing impairment. They were not paid for their participation in the experiment. 3 subjects, K6, L7 (both female) and R10 (male) were speakers of a south-eastern variety of British English, while the speech of the remaining subject J11 (male), displayed some characteristics of his native Lincolnshire, although without strong local features. All 4 subjects were non-rhotic.

3.2 Materials

The stimuli for the experiment consisted of consonant clusters combining a /k, g, ŋ/ C_1 and a /t, d ,s, z, r/ C_2. Velar stop C_1s were preceded by a long central mid open vowel [ɜː] (V_1)[2], whilst /ŋ/ followed low back rounded /ɒ/. C_2 was always followed by a vowel.

The main reason to use velar rather than alveolar C_1 was that word-final /t/ is often realised as a glottal stop in British English. A different place of articulation was chosen for the C_1 consonants for segmentation purposes; the choice for velar stops over labial ones was determined by the desire to control for the preceding vowel. The choice to use alveolar C_2s was made partially because of the exceptional behaviour of lax labiodental fricatives with regard to RVA in a number of languages (e.g., Hungarian, Russian), and partially because some claims about the phonetic basis for the nature of laryngeal contrast in fricatives have been made with specific reference to sibilants (Balise & Diehl, 1994).

The clusters were located at the internal boundary of noun + noun (N_1 + N_2) constructions and further embedded within a carrier phrase (*How does _ translate?*) designed to attract nuclear stress to the second noun. Both N_1s and N_2s were disyllabic with an initial lexical stress. Thus, the rhythmic structure of the stimuli and nuclear accent placement were controlled to maximise the potential effect of RVA, which has been shown to depend on lexical stress in Dutch (see Slis 1986). Given the sparsity of English words beginning with /z/ no attempt was made to control for the lexical frequency of the target words N_1 and N_2. For each of the 15 different consonant clusters 4 stimuli were prepared. Some sample stimuli are given in (3), with target consonant clusters in a slanted font. A full list of stimulus sentences appears in the appendix at the end of this paper.

(3) English sample stimuli

a. How does patchwor*k d*uvet translate?
b. How does headstro*ng z*ealot translate?
c. How does Hambur*g d*airy translate?

Stimuli containing the sonorant consonants /ŋ, r/ (realised by all subjects as [ŋ, ɹ] in word-final and word-initial contexts respectively) were included to create baseline conditions for the comparison of the relative effects of tense vs. lax C_2 on the properties of a preceding obstruent.

---

[2] This vowel is transcribed in square brackets in order to side-step questions about the underlying representation of orthographic <V + r + C> sequences, as in, e.g., <work>. Note that all subjects realised such sequences with a long vowel rather than [V + ɹ]

3.3 Procedure

The stimuli were presented to the subjects in a quasi-randomised order to avoid consecutive stimuli with identical consonant clusters. The subjects were asked to repeat each stimulus three times at a comfortable rate and to read a stimulus again if they made a mistake or produced a hesitation. In total, 3 (C_1) * 5 (C_2) * 4 (stimuli) * 3 (repetitions) * 4 (speakers) = 720 utterances were recorded.

Recordings were made onto minidisk in a sound-proofed room using a Brüel and Kjær condenser microphone (Type 4165) and measuring amplifier (Type 2609), and digitised at 22.5 kHz. Segmentation and acoustic measurements were carried out using PRAAT. 23 utterances had to be discarded because they contained small speech errors or (hesitation) pauses between C_1 and C_2 and 37 utterances were excluded because an underlying /k/ appeared to be realised without any oral closure (i.e., as a glottal stop). In addition, all (remaining) clusters starting with a /ŋ/ are excluded from the discussion below because they are largely irrelevant to the hypotheses under consideration, leaving a total of 425 utterances for analysis.

3.4 Segmentation and measurements

Segment boundaries were determined by visual inspection of waveforms and broadband spectrograms based on Fast Fourier Transforms (*FFT*) on a 5 ms Gaussian window (spectrogram bandwidth 260 Hz). The boundary between a vowel and a following plosive C_1 was placed where there was an abrupt change in the higher frequency energy, as illustrated by figure 1. The boundary between a C_1 plosive and a following C_2 was placed at the end of the release burst of the plosive, where *release burst* was defined as the initial transient and any following frication that could be assigned to C_1 rather than to a C_2 fricative.

59% (101 out of 171) of plosive-plosive clusters had a clear C_1 release and could therefore be internally segmented according to these criteria. In the remaining utterances where this was not the case, no boundary was placed and voicing and duration characteristics were measured for the cluster as a whole. In a few cases, mainly involving /g/ followed by /z/, the initial plosive was followed by a short period of schwa-like voicing. These intervals were treated as voiced releases (analogously to the 'embryonic vowels' often observed after word-final lax stops in French), and consequently their duration and voicing were assigned to C_1. In another set of tokens the release was completely obscured by the noise of a following fricative. Here the boundary was set at the onset of the noise signal.

The offset of C_2 constriction was defined as the offset of frication for /s, z/ and the onset of the release burst for /t, d/. The first measurement point for F_0

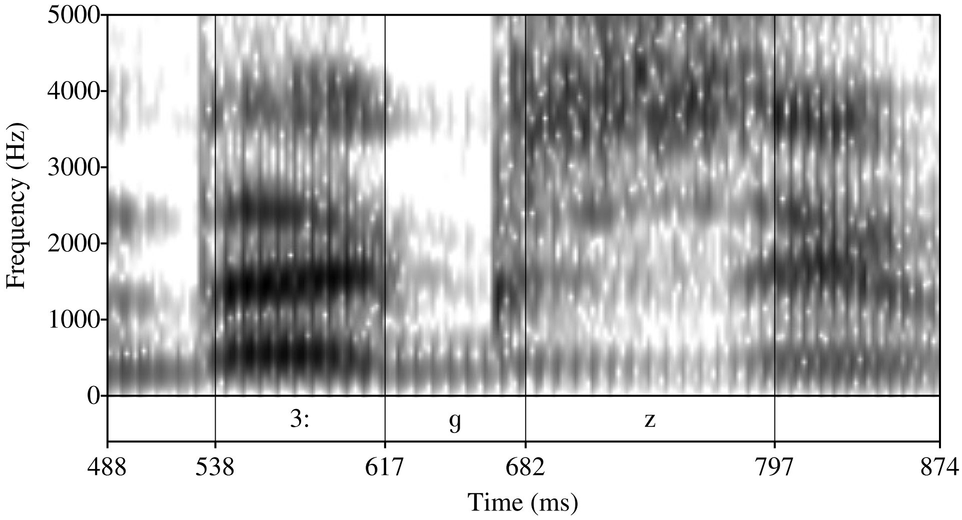

Figure 1: Sample spectrogram of a /gz/ cluster. Speaker: R10 (male).

was placed at 10 ms from the offset of frication for fricatives and at 10 ms after the onset of post-release voicing for plosively.

Voicing measures were determined on the basis of periodicity in the waveform and the presence of a voice bar in the spectrogram. Note that on the basis of these criteria the /gz/ cluster in figure 1 is voiced throughout. VOT was defined in the standard way in terms of the timing of voicing onset relative to the onset of the release burst in plosively.

The measurements that were made on the basis of the hand-segmented speech samples, as well as the relevant derived measures are listed in table 2, ordered by speech segment. Vowel (V_1) duration is a crosslinguistically common cue to [±tense] in postvocalic contexts, and is generally considered to be particularly salient in English (Chen, 1970; Flege & Hillenbrand, 1987). F_0 and F_1 values were extracted at 10 ms intervals between 50 and 10 ms preceding the onset of C_1 closure, using the autocorrelation and Burg algorithms embedded in PRAAT 4.0.

C_1 closure duration and release duration were measured separately for two reasons. The first is that they are not necessarily part of the same cue in articulatory or perceptual terms. In theory it is therefore possible that only one of these two features turns out to cue [tense] in the subjects' speech, and in this case simply measuring overall C_1 duration might lead to a distorted picture of the reflexes of [tense] marking on C_1 and/or C_2 in terms of segmental duration. The second, more practical reason for considering C_1 closure and release duration separately is that when a stop is followed by another obstruent, the release noise of the former may be partially or wholly obscured by by the latter. Again, this might distort the interpretation of C_1 overall duration. On identical grounds, C_1 voicing measures are reported separately for closure and release phases.[3]

Table 2: Acoustic measurements and derived measures.

			Segment				
	V_1		C_1		C_2		V_2
(a)	Duration	(d)	Closure duration	(h)	Closure duration (stops)	(l)	F_0 10-50 ms after C_1 offset
(b)	F_0 50-10 ms before C_1 onset	(e)	Release duration	(i)	Overall duration (fricatives)		
(c)	F_1 50-10 ms before C_1 onset	(f)	Overall duration	(j)	Voicing duration (fricatives)		
		(g)	Voicing duration (2 m.)	(k)	VOT (stops)		

Since this article attempts to address the possible phonetic conditioning of voicing assimilation, it is essential that the phonetic features of the potential assimilation triggers (C_2) are assessed too. Apart from voicing/VOT, this assessment comprises measurements of segmental duration and F_0 perturbations in the following vowel.

4 Results

4.1 Phonetic features of C_2

The data in table 3 and figure 2 indicates that the subjects use voicing distinctions to signal the distinction between tense and lax stops and fricatives as would be

---

[3] Only absolute voicing durations are reported in this article. My main motivation for focusing on absolute values of duration and voicing rather than voicing ratios is that the latter type of measure combines two acoustic features of [tense] in way that inflates the distance between two set of obstruents if both its components behave as they do in ‘typical’ cases of intervocalic tense-lax contrast. The relatively short duration and large amount of voicing of lax obstruents both contribute to a relatively high voicing ratio, whilst the long duration and lack of voicing of tense obstruents both contribute to a low voicing ratio. However, if either absolute duration or absolute voicing duration behaves contrary to the ‘typical’ pattern, the effects of underlying [±tense] or RVA on one feature may be (partially) cancelled by the other and voicing ratio ceases to be a reliable measure.

expected of an aspirating language. The stops /t/ and /d/ can be characterised as voiceless aspirated vs. voiceless, whilst the contrast between /s/ and /z/ is realised as voiceless vs. (partially) voiced. Thus, the subjects appear to exhibit the voicing patterns that are most common for English and the West Germanic languages and dialects more generally.

The mean VOTs for /t/ (70 ms) and /d/ (14 ms) fall into the standard ranges for the long lag and short lag categories, and the difference between them is highly significant according to a t-test: t(99) = 16.18, $p < .001$. All tokens of tense /s/ are completely voiceless, whilst /z/ has a substantial amount of voicing (77 ms). The mean voicing ratio for this obstruent is .78 (standard deviation: .22), which is fairly high in comparison with earlier studies such as Haggard (1978) or Smith (1996). Unsurprisingly, the mean difference in absolute voicing duration is statistically highly significant: t(173) = -23.62, $p < .001$.[4]

Table 3: Duration and voicing of C_2. Closure duration and VOT of /t, d/, and overall duration and duration of the voiced interval for /s, z/. All values in ms, and pooled across preceding contexts (/k, g/). Standard deviations in brackets.

C_2	VOT	Closure duration	N
/t/	70 (15)	56 (16)	44
/d/	14 (19)	71 (15)	57
	Voicing	Duration	N
/s/	0 (0)	132 (18)	92
/z/	77 (31)	99 (17)	83

F_0 microprosody seems to signal the distinction between tense and lax C_2 obstruents, too. Figure 2 plots the mean F_0 at five measurement points from 10 to 50 ms into the vowel following C_2 for the two male subjects R10 and J11. It shows how 10 ms into the vowel, F_0 values for /t, s/ on the one hand and /d, z, r/ on the other are roughly 20-25 Hz apart, and then gradually converge as time progresses. Both the magnitude of the F_0 differences and the grouping of lax (passively devoiced) /d/ and (actively voiced) /z/ with sonorant (passively voiced) /r/ are in line with earlier observations in the literature (see e.g. Kingston & Diehl 1994). A one-way ANOVA for *C_2 laryngeal specification* (tense obstruents vs. lax obstruents vs. sonorant) confirms that there is a highly significant effect of the phonological status of C_2 on F_0 at the onset of a following

---

[4] All data on stop + stop clusters in this section pertain to sequences that could be internally segmented, unless indicated otherwise. The result is a fairly large discrepancy in the number of cases for plosive and fricative C_2s.

vowel: F(1,171) = 24.05, $p < .001$. Tukey and Scheffe post hoc tests indicate that lax /d, z/ and /r/ are both significantly distinct from tense /t, s/ (both $p < .001$) but not from each other.[5]

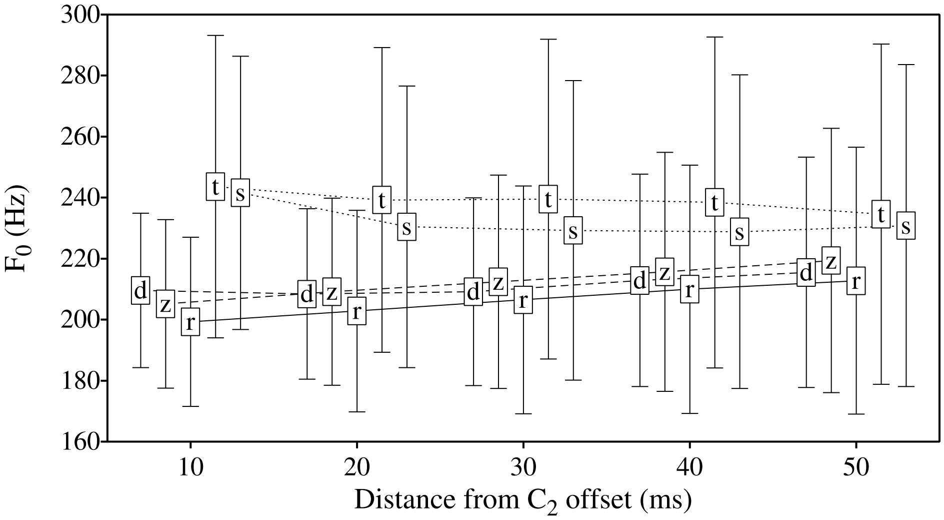

Figure 2: F_0 (Hz) 10-50 ms into the vowel following C_2, for female speakers only. Both internally segmented and unsegmented stop + stop clusters included. Error bars represent the mean ±1 standard deviation.

Finally, fricative (frication) duration but not stop closure duration behave according to the typical [±tense] pattern. On average, /s/ is 33 ms longer than /z/, and this difference is highly significant according to a t-test, t(137) = 12.51, $p < .001$. The 15 ms difference in closure duration between /t/ and /d/ is also statistically significant (t(99) = -4.94, $p < .001$) but patterns in the 'wrong' direction. As closure duration is not known to be a cue to [±tense] in word-initial contexts this finding hardly topples any theories, and it is hard to say whether any meaningful interpretation can be assigned to it.

---

[5] Utterances from the two male speakers were excluded from figure 2 and the ANOVA because of a considerable difference in overall F_0 level between the male and female subjects. However, the behaviour of the male subjects with regard to post-C_2 F_0 perturbations is highly similar to that of the female subjects, with a maximal difference between /d, z, r/ and /t, s/ of approximately 35 Hz.

4.2 Voicing of C_1

Figure 3 represents the mean voicing duration of /k, g/ (closure and release) across C_2 contexts. The bottom two bars of this figure show that the tense and lax velar stops are distinct in terms of voicing in the baseline pre-sonorant environment. For instance, the overall difference in voicing between /g/ and /k/ is 21 ms (43 vs. 22 ms). A T-test indicates that this difference is statistically significant: $t(77) = -6.50$, $p < .001$. Separate analysis of voicing during the closure and release stages give similar results. Thus, the present experiment suggests that obstruent voicing plays an auxiliary role alongside vowel duration in cueing the tense-lax distinction in word final stops.

However, overall C_1 voicing values in the remaining environments shows that the voicing distinction between /k/ and /g/ is not robust in three out of four pre-obstruent contexts, at least in part due to assimilation to following sounds. Relative to the baseline environment, there is a decrease in the amount of voicing of /g/ before tense /t/ and /s/. Conversely, there is an increase in the overall voicing duration of /k/ as well as /g/ before lax /z/. In all three of these environments, the extent of voicing contrast between /k/ and /g/ is drastically reduced: the residual contrast ranges between 3 and 6 ms. By contrast, the distinction in voicing observed in the baseline environment remains almost perfectly intact before lax /d/.

The decrease in voicing of /g/ before /t/ and /s/ and the virtual neutralisation of the voicing contrast with /k/ in these environments suggests that the tense obstruents trigger regressive voicing assimilation in preceding obstruents. Similarly, the increase in voicing of both /k/ and /g/ before /z/, and the near-neutralisation of voicing contrast in this context, suggests that the lax alveolar fricative, too, is capable of triggering RVA in preceding obstruents.

Moreover, the rather marked contrast in behaviour between this sound and its plosive counterpart /d/ provides a clear indication that the assimilation process involved is phonetically conditioned in that the ability of lax obstruents to trigger RVA seems dependent on their voicing. Just as (canonically) voiced lax plosives (e.g., in French or Dutch), (partially) voiced /z/ triggers assimilation; voiceless /d/ on the other hand, has little or no effect on the voicing of a preceding obstruent.

A similar grouping of C_2 contexts into three 'assimilation classes' (devoicing, no assimilation, voicing) appears to emerge if the amount of voicing during the closure phase of /k, g/ is considered in isolation. C_1 closure voicing is likely to be a slightly safer gauge of assimilation, since (the acoustic signature of) the release may be overlapped by a following obstruent, thus obscuring 'true' voicing values that allow for safe comparison with the baseline context. As illustrated in figure

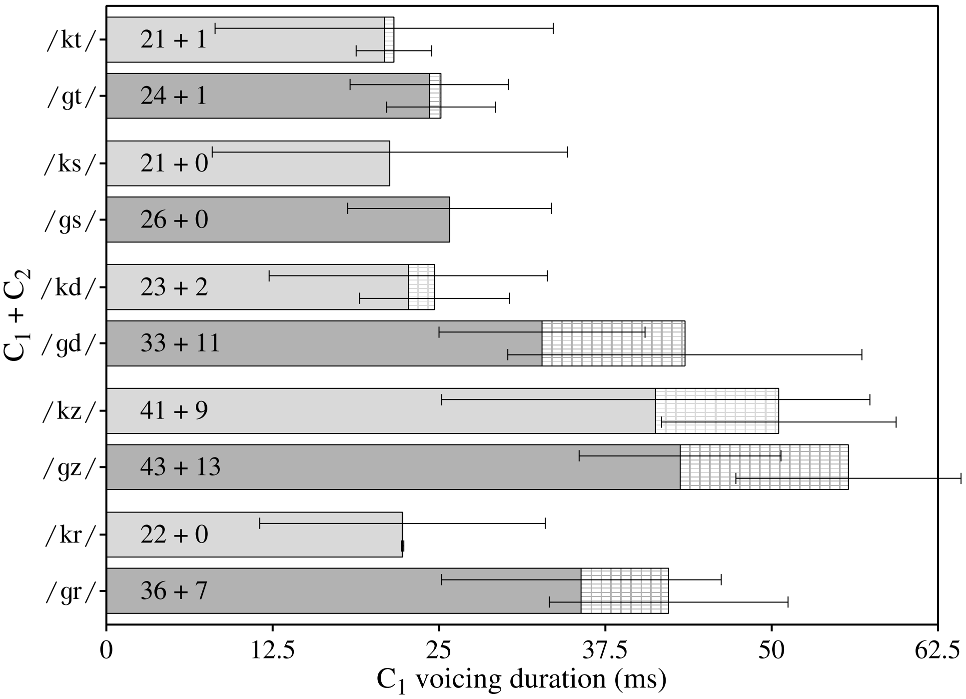

Figure 3: C_1 voicing duration across C_2 contexts. All measures in ms; error bars represent the mean ± 1 standard deviation. The top panel represents mean C_1 closure and release voicing durations separately: for each bar the left-hand segment indicates the extent of voicing during the closure phase and the right hand segment represents the duration of voicing during the release stage. Exact values for these means are printed in the leftmost segment for typographical reasons. This diagram shows a marked increase in voicing, relative to the baseline environment, for both /k/ and /g/ before /z/. There is a small decrease in the voicing of /g/ when it precedes /t, s/. These observations suggest that /z/ and, to a lesser extent, the two tense obstruents, trigger RVA. Before /d/, /k/ and /g/ behave more or less as in the baseline environment, which is an indication that the lax plosive is unable to trigger voicing assimilation.

3, there is a 14 ms distinction between /k/ and /g/ in the baseline environment. The amount of voicing during the closure phase of /g/ decreases preceding /t, s/, virtually neutralising the distinction with /k/ in these contexts. Before /z/ the closure voicing of both /k, g/ increases, again virtually erasing the distinction found in the baseline context. At the same time however, this distinction remains largely intact before /d/.

To test these impressionistic observations, a two-way ANOVA was carried out on the C_1 closure voicing data for *C_1 laryngeal specification* (/k/) vs. /g/) * *C_2 assimilation class* (devoicing /t/ + /s/ vs. neutral /d/ vs. voicing /z/). The results of this analysis show highly significant main effects of *C_1 laryngeal specification*, F(1,270) = 12.62, p< .001, and *C_2 assimilation class*, F(2,270) = 80.99, $p < .001$. The interaction of *C_1 laryngeal specification* and *C_2 assimilation class* is not significant. The main effect of *C_2 assimilation class* is consistent with the notion that some form of assimilation affects the amount of closure voicing of C_1, whilst the effect of *C_1 laryngeal specification* indicates that this process does not erase the voicing contrast found before /r/ completely. It seems likely that this effect is mostly due to the patterning of C_1 voicing before /r/.

Interestingly, Tukey and Scheffe post hoc tests show that the mean C_1 closure voicing before /z/ is significantly different from the amounts of voicing found before /d/ as well as /t, s/ (all $p < .001$), whereas the means for the latter two types of context are not significantly different. The first observation further reinforces the idea that English /z/ is capable of triggering RVA; the second does not necessarily imply that /t, s/ are unable to trigger assimilation, but is most likely due to the fact that, unlike /z/, these sounds only have a noticeable effect on the voicing of /g/ and therefore a smaller *overall* effect on C_1 voicing.

4.3 Duration of C_1

Obstruent duration is another well-recognised correlate of [±tense], even if not all acoustic intervals or contexts provide reliable measures (see e.g. Crystal & House 1988; Stevens et al. 1992). As figure 4 illustrates, there is a duration contrast between /k/ and /g/ before /r/. Both the closure stage and the release phase of /k/ (50 + 33 ms) are longer than the corresponding stretches of /g/ (44 + 21 ms). T-tests show that the 6 ms difference between the means for closure duration is weakly significant, t(77) = 2.41, p< .02, and that the difference in release burst duration is highly significant, t(77) = 4.54, $p < .001$.

C_1 duration parallels the behaviour of C_1 voicing in that both seem to contribute to the expression of [tense] in the baseline environment, but this parallelism does not extend to pre-obstruent contexts. The patterning of C_1 duration does not mirror the three-way grouping of C_2 environments that seems evident in the C_1 voicing data. Focussing again on the closure stage of C_1 stops, perhaps the only change that could be interpreted in terms of assimilation is the shortening (relative to the baseline) of /k/ before /d/ from 50 to 41 ms. However, before /z/, which triggers assimilation of phonetic voicing in /k/, there is hardly any change at all (50 vs. 51 ms). Similarly, there is no lengthening of /g/ before tense /t, s/ to match the devoicing of this sound in the same environments. Moreover, in 3 out of 4 obstruent C_2 contexts, there is a (small) positive difference between /k/ and /g/ in C_1 closure duration that is roughly similar to the difference observed in the baseline context. It seems unlikely therefore, that the mechanism that controls the acoustic extent of C_1 closure is the same as the device controlling the extent of voicing into C_1.

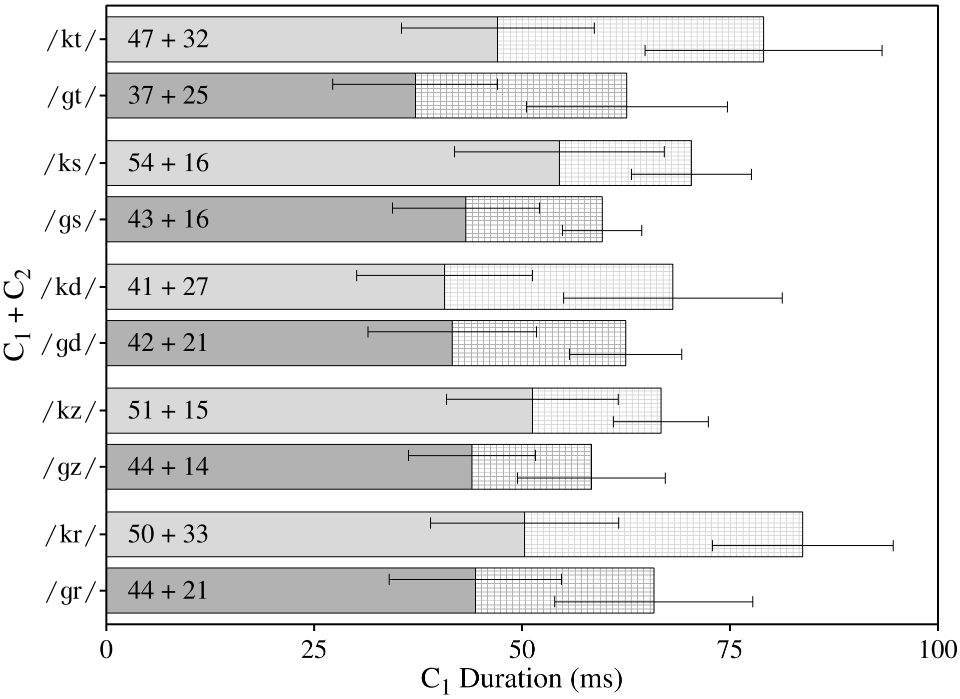

Figure 4: C_1 duration across C_2 contexts. All measures in ms; error bars represent the mean ± 1 standard deviation. The diagram represents mean C_1 closure and release durations separately: for each bar the left-hand segment indicates the closure duration and the right hand segment represents the duration of the release phase. Exact values for mean C_1 closure and release duration are given in the leftmost segment for typographical reasons. This diagram provides little or no evidence for an effect of RVA on the closure duration of C_1, because there is little or no systematic shortening (relative to the baseline context) before /t, s/ or shortening before /z/.

The results of a two-way ANOVA are largely consistent with these impressions. This ANOVA tested the effects of the classifications used above in the analysis of C_1 closure voicing, i.e., *C_1 laryngeal specification* (/k/) vs. /g/) * *C_2 assimilation class* (devoicing /t/ + /s/ vs. neutral /d/ vs. voicing /z/). It reveals significant main effects of *C_1 laryngeal specification*, F(1,270) = 15.24, $p < .001$, *C_2 assimilation class*, F(2,270), $p < .005$ as well as a significant interaction between *C_1 laryngeal specification* and *C_2 assimilation class*, F(2,270) = 4.94, $p < .01$. The main effect of *C_1 laryngeal specification* reflects the difference in closure duration of /k/ and /g/ before tense obstruents and /z/. The main effect of *C_2 assimilation class* indicates an assimilatory effect of C_2 on C_1, but the interaction of *C_1 laryngeal specification* and *C_2 assimilation class* suggests that this is mostly due to the shortening of /k/ (and not /g/) when followed by /d/.

The idea that with regard to C_1 closure duration, _/d/ forms a distinct environment with regard to the implementation of /k/ is further reinforced by Tukey and Scheffe post hoc tests, which show that the mean duration of C_1 closure is significantly different from the means before tense obstruents and /z/ (all $p < .015$), whilst the difference between the latter two environments is not significant on both tests.

4.4 V_1 duration

Figure 5 represents the duration of the vowel [ɜː] across C_1 and C_2 contexts. The two bottom bars of this figure indicate that the distinction between word-final /k/ and /g/ is signalled by the duration of V_1 in a familiar fashion: the tense stop is preceded by a shorter vowel (72 ms) than its lax counterpart (99 ms). A t-test shows that this difference is statistically significant: t(77) = -5.47, $p < .001$.

Unlike C_1 voicing and duration, this vowel length contrast is robust across C_2 contexts, and it seems hard to interpret the variation that can be observed in terms of assimilation. It is true that there is a decrease in vowel duration before /t/, which might suggest an influence of the [+tense] alveolar stop, but note that the next ‘shortest’ environment is _/d/.

This apparent lack of assimilation of vowel length is consistent with a two-way ANOVA for *C_1 laryngeal specification * C_2 assimilation class* on the V_1 duration data. This ANOVA reveals a highly significant effect of *C_1 laryngeal specification*, F(1,270) = 63.432, $p < .001$ but no other effects. A second ANOVA, for *C_1 laryngeal specification * C_2 laryngeal specification* (tense /t, s/ vs. lax /d, z/), also failed to detect any effect of C_2 on the duration of V_1. Unsurprisingly, this ANOVA again revealed a highly significant effect of *C_1 laryngeal specification*, F(1,270) = 88.78, $p < .001$, but no other significant effects.

4.5 Effects on F_0 and F_1

The remaining two phonetic features that were investigated in this study are the perturbations of F_0 and F_1 during the last 50 ms of V_1. F_0 and F_1 tend to be raised

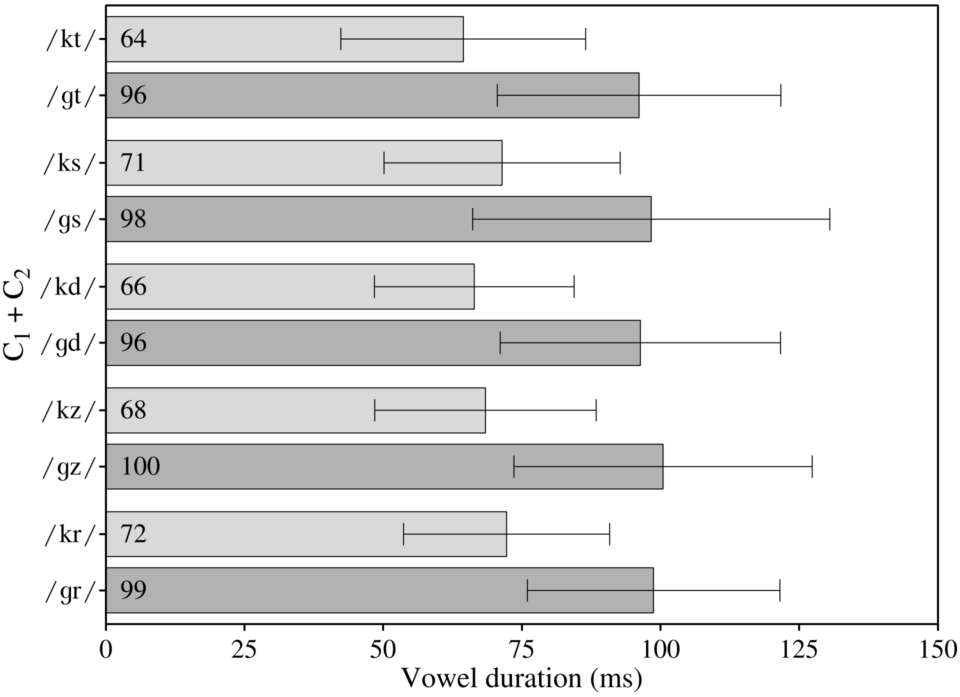

Figure 5: V_1 duration across C_2 contexts. All measures in ms; error bars represent the mean ± 1 standard deviation. This diagram shows that vowel duration reflects the status of the immediately following plosive (C_1) but is impervious to the value of [tense] on C_2 consonants: /k/ is preceded by a vowel of relatively short and highly similar duration across C_2 contexts whilst /g/ is preceded by longer vowels of near-identical length.

on vowels flanking tense obstruents regardless of their VOT, and Kingston & Diehl (1994, 1995) hypothesise that these perturbations of F_0 and F_1 (and voicelessness) act in unison to create an upward shift in the overall balance of spectral energy in the low frequency region.

The effect of [tense] on the F_0 of a following vowel is clearly illustrated by the data summarised in figure 2 above, but there are no corresponding effects of C_1 or C_2 on the F_0 of the vowel preceding V_1. This may come as no surprise in the light of previous studies showing that F_0/F_1 perturbations are much more marked in post-obstruent than in pre-obstruent contexts (see Kingston & Diehl 1994 for a survey).

However, as illustrated by figure 6 and table 4, C_1 and C_2 do have effects on the F_1 of the preceding vowel. Figure 6 plots the first formant contour of the vowel [ɜ] at 10 ms intervals between 50 and 10 ms preceding the onset of C_1. The downward slope of this contour is steeper before /g/ than before /k/, which results in a 26 Hz difference (476 vs. 502 Hz) at 10 ms before the onset of C_1. Both the shapes of the F_1 trajectories and the relative positions of their end points agree with data reported elsewhere in the literature Stevens et al. 1992; Kingston & Diehl 1994. Note that the difference at 10 ms is statistically significant according to a t-test: t(77) = 3.13, $p < .005$.

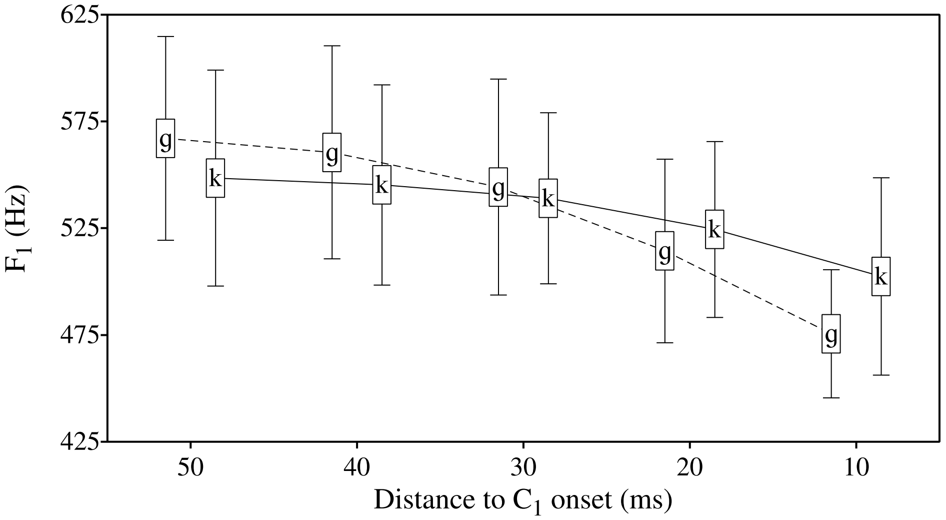

Figure 6: F_1 (Hz) at 5 points from 50-10 ms preceding C_1 onset in /k/ + /r/ and /g/ + /r/ sequences. Error bars represent the mean ±1 standard deviation.

Table 4 reports the mean F_1 at 10 ms from the onset of C_1 across C_1 and C_2 contexts. Interestingly, this data appears to show a relatively consistent effect of C_2 on the vowel preceding C_1 and therefore suggests a form of regressive assimilation. Relative to the baseline environment /k + r/ (502 Hz), F_1 is lowered somewhat before /k/ followed by lax /d/ (486 Hz) and /z/ (489 Hz), whilst it is raised before tense /t/ (508 Hz) and especially /s/ (519 Hz). Similarly, the F_1 of [ɜː] is raised somewhat if C_2 is a tense obstruent. At the same time, F_1 is consistently higher before /k/ than before /g/ across obstruent C_2 contexts, which suggests that any assimilatory influence is only incompletely neutralising.

A two-way ANOVA for *C_1 laryngeal specification * C_2 laryngeal specification* (tense /t, s/ vs. lax /d, z/) is consistent with the idea that F_1 is subject to partial laryngeal assimilation to C_2. This ANOVA reveals highly significant effects

Table 4: F_1 preceding obstruent clusters. F_1 (Hz) at 10 ms before the onset of C_1. Standard deviations in brackets.

C_1C_2	F_1 at C_1 - 10 ms	N	C_1C_2	F_1 at C_1 - 10 ms	N
/kt/	508 (42)	31	/gt/	508 (34)	26
/kd/	486 (41)	26	/gd/	478 (20)	18
/ks/	519 (43)	47	/gs/	487 (28)	45
/kz/	489 (43)	36	/gz/	479 (34)	47
/kr/	502 (46)	32	/gr/	476 (30)	47

of both *C_1 laryngeal specification*, F(1,270) = 10.75, $p < .001$, and *C_2 laryngeal specification*, F(1,270) = 24.75, $p < .001$, but no interaction. The second effect indicates that the value of F_1 indeed reflects the [tense]-value of C_2, whilst the first effect indicates that the F_1 distinction between /k/ and /g/ found in the baseline environment is partially preserved in pre-obstruent contexts despite the effects of assimilation.

5 General discussion

The aim of this paper, as stated in the introduction, was to shed some light on the nature and modelling of voicing assimilation phenomena. With regard to the empirical side of this enterprise, the first issue I identified was the possibility that voicing assimilation is generally conditioned by the phonetic properties of the sounds that trigger the process; recall that this hypothesis was prompted by the observation that lax stops and affricates such as /b, d, d͡ʒ, g/ only trigger assimilation if they belong to the ‘French’ prevoiced type.

The results of the experiment discussed above lend at least partial support to the idea that this phonetic conditioning extends to the behaviour of English lax /z/. As shown in section 4.1, this sound is implemented as partially voiced by the four subjects, and as demonstrated in section 4.2 it triggers an increase in the amount of voicing of a preceding velar stop. /z/ could therefore be said to trigger regressive voicing assimilation, at least under an acoustic definition of the process (see further below). The lax plosive /d/, on the other hand, is realised with a short lag VOT, i.e., voiceless, and has no effect on the voicing of a preceding sound relative to the baseline pre-sonorant context.

It might seem that the patterning of F_1 during the last 50 ms of the vowels preceding the target clusters, which was reported in section 4, contradicts this apparent connection between phonetic voicing during the trigger obstruent and the occurrence of RVA. The first formant of vowels followed by a tense obstruent in C_2 position was found to be raised somewhat, whilst the presence of both /z/

and /d/ in this position has a (weaker) lowering effect. Consequently, this data suggests that /d/ is able to trigger RVA, even if in a more restricted way than /z/ and /t, s/, all of which influence C_1 voicing as well as F_1.

So far I have implicitly treated the term *voicing assimilation* as a monolithic concept covering all correlates of [tense], i.e., 'big' voice. If the effects of C_2 on the correlates of [tense] are considered separately, however, the contradiction identified in the previous paragraph may well turn out to be an artefact of a monolithic concept of voicing assimilation. The set of stimuli used in the present experiment did not allow for an assessment of the effect of C_2 on the F_1 of a following vowel, since the lexical quality of that vowel was not controlled for. But lax stops tend to depress F_1 regardless of their (canonical) VOT, and /d/ clearly acts as a F_0 depressor for the speakers investigated here (see figure 2). This suggests that /d/ acts as an F_1 depressor for these speakers, too. If this is indeed the case, the effects of /z/ and /d/ on a preceding velar stop are fully consistent with their phonetic properties: /z/ is partially voiced and an F_1 depressor and able to 'transmit' both features to a preceding stop, whilst /d/ only lowers F_1 and is able to 'transmit' only this feature to a preceding stop. In the absence of information regarding the effect of /d/ on a following vowel, this theory remains somewhat speculative, but note that the mismatch between the C_1 voicing and F_1 effects does not contradict the idea that RVA is phonetically conditioned.

Whilst this study demonstrates certain effects on the implementation of word-final obstruents that are best regarded as (forms of) RVA, it is clear that the process(es) involved do not extend to all correlates of [tense] and, partially as a result of this, operate in a non-neutralising fashion. Thus, even if the closure duration of /k/ is shortened before /d/, there is no consistent mapping between the value of [tense] or the voicing of C_2 and the closure duration of the preceding velar stops to warrant labelling it as voicing assimilation. The durational contrast between vowels preceding /k/ and those followed by /d/ is hardly affected by the phonological or phonetic status of C_2. Note that the lack of an effect on vowel duration (contrast) may well play an important role in the perception, reported in impressionistic studies, that lax obstruents do not trigger RVA in English. Moreover, even if there is an assimilatory effect of C_2 on the F_1 of preceding vowels, this effect does not completely erase the distinction between /k/ and /g/ found in the baseline environment.

The absence of full neutralisation of underlying distinctions in C_1 position means that it is impossible to model the form of RVA observed in this study in terms of a textbook-style phonological rule that switches the values of a feature [voice], at least if this feature is used to represent lexical distinctions between tense and lax obstruents. Treatment along these lines entails that RVA is neutralising at the phonological level and therefore implies that it is neutralising at the phonetic level as well.

Beyond this rather broad point, it is hard to draw any conclusions for the present study for generative phonological theories of voicing assimilation phenomena, as such theories rarely come accompanied by explicit models of phonetic interpretation. As a result they do not derive specific enough predictions about the phonetic manifestation of VA to be tested against the data gathered here. More importantly, perhaps, there is the issue of whether the form of voicing assimilation investigated above falls within the scope of phonological theories in the first place: some phonologists might want to dismiss it as a 'low-level phonetic process' that occurs outside the phonological component of the grammar. I think it is important to note that this position presupposes that there exists a class of voicing assimilation rules with a different (and perhaps neutralising) phonetic signature that is within the scope of phonological theory. The experimental phonetic literature on voicing assimilation contains a few clues that there may indeed be phonetically different varieties of voicing assimilation. Thus, Thorsen (1966); Charles-Luce (1993) and Jansen (2004) show that in French, Catalan, and Hungarian respectively, RVA affects the vowels preceding obstruents targeted by the process, a phenomenon which was shown in 4.4 above to be absent from the English data.[6]

Regardless of their ramifications for phonological theory, the results obtained by the present experiment are largely consistent with a coarticulation-based theory of regressive assimilation developed in Jansen (2004). According to this theory, the differences between spontaneous and non-spontaneous voicing and devoicing form the key to understanding voicing assimilation phenomena. The notion of spontaneous or *passive* voicing is relatively well-known (see for example Chomsky & Halle 1968) and refers to the mechanical interaction between a voiced carrier signal and place and manner of articulation. Passive voicing occurs during sounds of which the articulation does not create sufficient aerodynamic inhibitions to interrupt the carrier signal, such as (most) sonorants; passive *devoicing* can be defined along the same lines to capture sounds of which the articulation does provide aerodynamic obstacles that lead to temporary inhibition of voicing (utterance-initial obstruents, as well as medial and final obstruents and obstruent clusters longer than a certain critical interval: see Westbury & Keating (1986); Stevens (1998)).

There is ample evidence to suggest that some classes of obstruents are accompanied by articulatory gestures designed to counteract the effects of passive devoicing: some of these gestures slow down the build up of air pressure behind the oral constriction (e.g., expansion of the pharyngeal cavity), others lower the minimal transglottal pressure difference that is required for vocal fold vibra-

---

[6] This does not mean that generative treatments are consistent with the experimental literature more generally. For example, there is a range of studies that show that RVA to be incompletely neutralising, regardless of the the presence of vowel length effects: e.g., (Thorsen, 1966, 1971; Charles-Luce, 1993; Burton & Robblee, 1997; Barry & Teifour, 1999; Jansen & Toft, 2002).

tion (Ladefoged, 1973; Stevens, 1998). These obstruents might be called *actively voiced*. Similarly, there is evidence that certain types of obstruent are articulated in such a way as to minimize the effects of passive voicing; consequently the sounds involved might be labelled *actively devoiced*.

Because their prevoicing in utterance-initial and post-obstruent environments would be hard to account for in terms of passive voicing the [-tense] stops of languages such as French and Dutch are best regarded as actively voiced. Their English counterparts on the other hand, are better regarded as passively (de)voiced (cf. Harms 1973), which implies that the former, but not the latter, are accompanied by 'voicing-enhancing' measures of the type described by Ladefoged (1973) and Stevens (1998). The fact that /z/ is partially voiced in post-obstruent contexts in the data surveyed above suggests that this sound is actively voiced, and therefore accompanied by voicing-enhancing articulations, too. Finally, the tense plosives and fricatives of English are all accompanied by relatively large glottal abduction gestures (Yoshioka et al., 1981; Stevens, 1998), at least in word-initial environments. As far as the plosives are concerned, the abduction gesture is responsible for aspiration and long lag VOT found in these contexts. The main role of glottal abduction in tense fricatives may be to create the aerodynamic conditions for the production of sustained noise at the oral constriction, but in either case, glottal abduction is one of the most effective articulations to produce active devoicing. All this implies that the C_2 obstruents investigated here fall in the voicing categories represented in table 5:

Table 5: Voicing categories for English alveolar obstruents.

Actively devoiced	/t, s/
Passively (de)voiced	/d/
Actively voiced	/z/

These classes correspond exactly to the 'assimilation classes' used in section 4.2 above to capture the effects of C_2 on the voicing of C_1 (cf. figure 3). Thus, actively devoiced /t, s/ devoice preceding /g/, passively (de)voiced /d/ has no effect on the voicing of a preceding sound, whilst only actively voiced /z/ causes an increase in the voicing of preceding velar stops. According to the theory developed in Jansen (2004) this correspondence between the active or passive (de)voicing of C_2 and the effects on a preceding sound is the byproduct of anticipatory coarticulation of the (de)voicing-enhancing measures that underpin the three voicing classes. Coarticulation here means the general smoothing process that governs transitions between articulatory targets in speech production. It seems that the extent of coarticulation in this sense is subject to language-specific, speech style, and prosodic factors, but it is nevertheless regarded as universally present whenever two or more sounds are articulated sufficiently close together in time (see e.g., Farnetani 1997 for an overview). It follows that the articulatory gestures responsible for the production of voicing contrasts are subject to coarticulation regardless of the presence of phonological voicing assimilation rules in the grammar.

For example, according to this theory, the voicing-enhancing measures accompanying actively voiced English /z/ ‘spill over’ into a preceding obstruent by means of anticipatory coarticulation and thereby increase the amount of voicing during the sound in question. Similarly, the devoicing gestures associated with /t, s/ are expected to be anticipated during the articulation of a preceding sound, and thus to shorten the voiced interval of such a sound somewhat.

No attempt will be made here to formalise this account, although note that coarticulation models such as *Articulatory Phonology* (Browman & Goldstein 1986 et seq.) would seem to offer an excellent framework for this purpose. Nevertheless, it appears that some of the properties of the voicing assimilation process(es) observed above can be captured in a very straightforward way in a coarticulation-based account. For example, there is evidence to suggest that the production of the vowel length contrasts that cue postvocalic [±tense] is not mechanically related to the articulation of voicing contrasts (or the articulation of obstruents more generally), even if both correlates conspire to create a unified perceptual category (Kluender et al., 1988). If this is indeed the case, a coarticulation-based model predicts that coarticulation in obstruent sequences does not have any effects on the duration of preceding vowels, simply because no gestures related to vowel length control are implemented during the realisation of such obstruent sequences.

Finally, given that /d/ and /z/ are both F_1 depressors, the observation that both have a small effect on the F_1 of preceding vowels across an intervening velar stop independently of voicing interactions, may follow from a coarticulation-based approach. However, since the articulatory underpinnings of F_0/F_1 perturbations by tense and lax obstruents remain unclear (Kingston & Diehl, 1994), thus must remain speculative.

6 Conclusions

The aim of this article was to address certain aspects of RVA by means of a quantitative acoustic study of English obstruent clusters. It was found that, contrary to many impressionistic descriptions, the lax plosive /d/, and in particular the lax fricative /z/ exert influences on the phonetic properties of preceding velar stops that are best regarded as forms of voicing assimilation or, more broadly, ‘laryngeal’ assimilation. These assimilatory effects do not completely erase the underlying distinction between word-final /k/ and /g/, nor do they extend to vowel length distinctions, and therefore the form(s) of assimilation found in this study are clearly non-neutralising. In this respect, this article contributes to a growing body of research which indicates that voicing assimilation phenomena are non-neutralising more generally (Thorsen, 1966, 1971; Charles-Luce, 1993; Burton & Robblee, 1997; Barry & Teifour, 1999; Jansen & Toft, 2002).

It might be argued that the results obtained by this investigation are of limited relevance to the phonological analysis of voicing assimilation phenomena, because these results reflect a low-level phonetic process that lies outside the scope of phonological modelling. This position might be justifiable on the grounds of a (very partial) phonetic typology of voicing assimilation phenomena, but it does not dispense with the need for an account of the data gathered here. The precise patterning of the phonetic effects of /t, s, d, z/ on preceding /k, g/ suggests that this account must be based on the coarticulation of the gestures that underpin the acoustic correlates of [tense]. The most clear-cut piece of evidence in favour of this approach is the observation that effects on C_1 voicing match the (canonical) voicing rather than the lexical specification of C_2 obstruents.

7 Acknowledgements

This article is a modified version of chapter 5 of my University of Groningen PhD dissertation (Jansen, 2004). This work was in part sponsored by grant 200-50-068 from the *Nederlandse Organisatie voor Wetenschappelijk Onderzoek* (NWO). Thanks are due, first and foremost to my subjects, to Bernard Howard at the School for Oriental and African Studies for technical support, to Zoë Toft for help during the recording sessions and comments; and to the audience at Toulouse, Dicky Gilbers, and John Nerbonne for further helpful comments. Many thanks also to Patrick Honeybone and an anonymous reviewer for numerous and often incisive comments. All remaining errors are my own.

Appendix: stimulus sentences

/k/ + /t/

How does Falkirk tonic translate?
How does Selkirk topping translate?
How does brickwork tunnel translate?
How does patchwork tartan translate?

/**k**/ + /**d**/

How does patchwork duvet translate?
How does Selkirk devil translate?
How does Falkirk dagger translate?
How does brickwork depot translate?

/**k**/ + /**s**/

How does Falkirk singer translate?
How does Selkirk saga translate?
How does patchwork surface translate?
How does brickwork siphon translate?

/**k**/ + /**z**/

How does Falkirk zipper translate?
How does brickwork zester translate?
How does patchwork zebra translate?
How does Selkirk zygote translate?

/**k**/ + /**s**/

How does patchwork rigging translate?
How does Falkirk river translate?
How does Selkirk raven translate?
How does brickwork rafter translate?

/**g**/ + /**t**/

How does Hamburg tenant translate?
How does Limburg timber translate?
How does Lindberg tactic translate?
How does Strindbergh temper translate?

/**g**/ + /**d**/

How does Limburg daisies translate?
How does Hamburg dairy translate?
How does Lindberg diary translate?
How does Strindbergh Danish translate?

/**g**/ + /**s**/

How does Hamburg satin translate?
How does Limburg singer translate?
How does Lindberg summon translate?
How does Strindbergh sermon translate?

/g/ + /z/

How does Limburg zombie translate?
How does Hamburg Zulu translate?
How does Strindbergh zenith translate?
How does Lindberg zephyr translate?

/g/ + /r/

How does Limburg relish translate?
How does Strindbergh rigour translate?
How does Hamburg rifle translate?
How does Lindberg rumour translate?

References

Balise, R. & Diehl, R. (1994). Some distributional facts about fricatives and a perceptual explanation. *Phonetica*, 51, 99–110.

Barry, M. & Teifour, R. (1999). Temporal patterns in Arabic voicing assimilation. In *Proceedings of the XIVth International Congress of Phonetic Sciences*, volume 3 (pp. 2429–2432). San Francisco.

Borowsky, T. (2000). Word faithfulness and the direction of assimilations. *The Linguistic Review*, 17, 1–28.

Browman, C. & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, 3, 219–252.

Burton, M. & Robblee, K. (1997). A phonetic analysis of voicing assimilation in Russian. *Journal of Phonetics*, 25, 97–114.

Charles-Luce, J. (1993). The effects of semantic context on voicing neutralisation. *Phonetica*, 50, 28–43.

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

Cho, T. & Ladefoged, P. (1997). Variations and universals in VOT: evidence from 17 endangered languages. *UCLA Working Papers in Phonetics*, 94, 18–40.

Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper.

Cohen, A., Ebeling, K., & van Holk, A. (1972). *Fonologie van het Nederlands en het Fries*. Den Haag: Nijhoff.

Crystal, T. & House, A. (1988). Segmental durations in connected speech: current results. *Journal of the Acoustical Society of America*, 83, 1553–1573.

Ernestus, M. (2000). *Voice assimilation and Segment Reduction in casual Dutch*. PhD thesis, Free University of Amsterdam/Holland Institute of Generative Linguistics.

Farnetani, E. (1997). Coarticulation and connected speech processes. In W. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences*. Oxford: Blackwell.

Flege, J. & Hillenbrand, J. (1987). A differential effect of release bursts on the stop voicing judgements of native French and English listeners. *Journal of Phonetics*, 15, 203–208.

Gimson, A. (1994). *Gimson's Pronunciation of English*. London: Arnold, 5th edition. Revised by A. Cruttenden.

Haggard, M. (1978). The devoicing of voiced fricatives. *Journal of Phonetics*, 6, 95–102.

Harms, R. (1973). Some nonrules of English. Distributed by the Indiana University Linguistics Club.

Harris, J. (1994). *English Sound Structure*. London: Blackwell.

Iverson, G. & Salmons, J. (1995). Aspiration and laryngeal representation in Germanic. *Phonology*, 12, 369–396.

Iverson, G. & Salmons, J. (1999). Glottal spreading bias in Germanic. *Linguistische Berichte*, 178, 135–151.

Jacobs, N., Prince, E., & van der Auwera, J. (1994). Yiddish. In E. König & J. van der Auwera (Eds.), *The Germanic Languages*. London: Routledge.

Jansen, W. (2004). *Laryngeal Contrast and Phonetic Voicing. A Laboratory Phonology Approach to English, Hungarian, and Dutch*. PhD thesis, University of Groningen.

Jansen, W. & Toft, Z. (2002). On sounds that like to be paired (after all): an acoustic investigation of Hungarian voicing assimilation. *SOAS Working Papers in Linguistics*, 12, 19–52.

Jones, D. (1956). *An Outline of English Phonetics*. Cambridge: Heffer, 8th edition.

Katz, D. (1987). *Grammar of the Yiddish Language*. London: Duckworth.

Keating, P. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, 60, 286–319.

Kingston, J. & Diehl, R. (1994). Phonetic knowledge. *Language*, 70, 419–454.

Kingston, J. & Diehl, R. (1995). Intermediate properties in the perception of distinctive feature values. In B. Connell & A. Arvaniti (Eds.), *Papers in Laboratory Phonology IV. Phonology and Phonetic Evidence*. Cambridge: CUP.

Kluender, K., Diehl, R., & Wright, B. (1988). Vowel length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*, 16, 153–169.

Kohler, K. (1979). Parameters in the production an the perception of plosives in German and French. *Arbeitsberichte Institut Für Phonetik Kiel*, 12, 261–291.

Ladefoged, P. (1973). Features of the larynx. *Journal of Phonetics*, 1, 73–83.

Lakoff, G. (1972). The arbitrary basis of transformational grammar. *Language*, 48, 76–87.

Lisker, L. & Abramson, A. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20, 384–422.

Myers, S. (2002). Gaps in factorial typology: The case of voicing in consonant clusters. MS., Rutgers Optimality Archive (ROA 509-0302).

Riemersma, T. (1979). *Sylabysjerring, nazzeljerring, assymljerring*. Ljouwert: Koperative Utjowerij.

Selkirk, E. (1980). *The Phrase Phonology of English and French*. New York: Garland.

Slis, I. (1985). *The voiced-voiceless distinction and voice assimilation in Dutch*. PhD thesis, University of Nijmegen.

Slis, I. (1986). Assimilation of voice in dutch as a function of stress, word boundaries and sex of speaker and listener. *Journal of Phonetics*, 14, 311–426.

Smith, C. (1996). The devoicing of /z/ in American English: Effects of local and prosodic context. *UCLA Working Papers in Phonetics*, 92, 125–150.

Stevens, K. (1998). *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Stevens, K., Blumstein, S., Glicksman, L., Burton, M., & Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustical Society of America*, 91, 2979–3000.

Thorsen, N. (1971). Voicing in British english T and D in contact with other consonants. *Annual Report of the Institute of Phonetics, University of Copenhagen*, 5, 1–39.

Thorsen, O. (1966). Voice assimilation of stop consonants and fricatives in French and its relation to sound duration and intra-oral air pressure. *Annual Report of the Institute of Phonetics, University of Copenhagen*, 1, 67–76.

Tiersma, P. (1985). *Frisian Reference Grammar*. Dordrecht: Foris.

Wells, J. (1982). *Accents of English*. Cambridge: CUP. 3 Vols.

Westbury, J. & Keating, P. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics*, 22, 145–166.

Wissing, D. (1991). Regressiewe stemassimilasie in afrikaans. *South African Journal of Linguistics*, Supplement 11, 132–156.

Yoshioka, H., Löfqvist, A., & Hirose, H. (1981). Laryngeal adjustments in the production of consonant clusters and geminates in American English. *Journal of the Acoustical Society of America*, 70, 1615–1623.