

Dissertation Proposal:  
Two acoustic correlates of [voice] in American  
English-speaking children and what they tell us about  
the development of phonetics and phonology

Joshua Tauberer

October 8, 2008

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>The vowel duration correlate</b>	<b>3</b>
2.1	The grammatical status of the VLE . . . . .	4
2.2	An open issue for cross-linguistic comparisons . . . . .	8
<b>3</b>	<b>The /s,z/ contrast and /z/-devoicing</b>	<b>8</b>
<b>4</b>	<b>Acquisition of [voice] and its correlates</b>	<b>9</b>
4.1	Acquisition of the VLE in English . . . . .	10
4.2	Acquisition of the fricatives . . . . .	11
<b>5</b>	<b>Preliminary Work</b>	<b>11</b>
5.1	Experiment A: Confirmations of several adult hypotheses . . . . .	11
5.1.1	Methods . . . . .	12
5.1.2	Results . . . . .	13
5.1.3	Discussion . . . . .	15
5.2	Experiment B: Acquisition . . . . .	18
5.2.1	Methods . . . . .	18
5.2.2	Results . . . . .	19
5.2.3	Discussion . . . . .	20
<b>6</b>	<b>The Proposal</b>	<b>21</b>
<b>A</b>	<b>Word list used in experiment A</b>	<b>28</b>

## 1 Introduction

Phonological features often correlate with more than just a single articulatory property of speech output. de Jong (1995) for instance found in three American English speakers that the degree of tongue retraction and lip protrusion was correlated in the production of [+back] vowels. Though [back] is intended to convey the control over tongue position only (if it is intended to make any articulatory reference at all), it appears that [round] is recruited as a secondary feature to backing. Besides viewing [round] as a secondary phonetic feature (Stevens et al., 1986), another related perspective is that backing and roundness are together a part of the phonetic implementation of the abstract phonological feature [back] (Keating, 1984). The focus of the dissertation being proposed is the phonological feature [voice] in obstruents and its correlation with several independent gestures — phonetic voice itself, segment duration, and the duration of the preceding vowel — and the development of this correlation in American English-speaking children. It has been proposed in the past that the durational correlates serve to enhance the acoustic properties of [voice] (Stevens et al., 1986) or that vowel duration serves to enhance the perception of the duration of the consonant (Kluender et al., 1988). This rationale comes into question when vowel length changes persist despite the absence of voice in the (quite common) devoicing of /z/ (Smith, 1997) or in children preceding consonants that are omitted entirely (Weismer et al., 1981).

The dissertation will be concerned with the realization of [voice] in stops and the fricative pair /s,z/, and it will focus primarily on two correlates of [voice]: the duration of the preceding vowel and, for the fricatives, the duration of glottal voicing during the period of frication. In American English vowels are up to roughly 50% longer before voiced consonants than unvoiced consonants in some contexts (Chen, 1970). Acquisition research has shown that English-learning children exhibit this duration difference in their own speech at a remarkably early age, at least as early as 1;6–1;9 (Ko 2007 and the new results reported below), but the later development of voiced segments and especially fricatives (Stoel-Gammon, 1985) in word-final position suggests that the so-called ‘secondary’ aspects of voice may in fact be acquired by children before the primary feature, at least from the point of view of production. The phoneme /z/ is routinely ‘devoiced’ in adult speech, making it an interesting candidate for study, especially in children. Glottal voicing is sometimes present in /z/ but very rarely /s/ (Smith, 1997). Devoicing appears to be a low-level process of assimilation or lenition (Smith, 1997). Perfectly normal z-devoicing in children may have in past work obscured the presence of correctly articulated /z/. The high variability of the presence of voice in /z/ opens up a direction of study for the relation of voice and duration. (These two correlates are among many aspects of consonant voicing, among them Voice Onset Time, fundamental frequency and F<sub>1</sub> perturbation of the surrounding vowels, the presence of release bursts, and the duration of consonant closure and of frication. See Jansen 2004 for a summary, Stevens et al. 1986 for why they might be obligatorily or optionally redundant.)

I have already conducted preliminary work on both the vowel duration correlate of [voice] in adults and its time-course of acquisition, filling in several gaps in the literature. I propose now to pursue the interaction of the above phenomenon during acquisition. My dissertation will be exploratory from three angles and aims to answer two higher-level questions regarding the nature of acquisition of implementations of phonemes. The exploratory directions are:

- Do vowel length and other correlates of [voice], including voicing itself, trade-off in

production in adults, in the same way that a correlation was found between tongue position and lip rounding in [+back] vowels (de Jong, 1995)?

- What is the nature of /z/-devoicing in children: Does it mimic the adult pattern, and has it resulted in past research mistaking /z/ productions for [s]?
- What is the time course of the acquisition of the production of the individual articulatory correlates of, i.e. the secondary phonetic features or phonetic implementation of, phonological [voice]?

The two high-level questions to be answered are:

- Do the correlates of [voice] trade-off in production in children, both across utterances in a single session and longitudinally across sessions? As the articulatory gestures that make up a phone are acquired in production, does the child adjust his reliance on each of the gestures?
- The vowel duration and glottal voicing cues are observed in young children, but what can we learn about their linguistic status, as either e.g. deriving from stored phonetic detail in the lexicon or from being active linguistic knowledge about [voice]?

This will be a corpus study.

This proposal begins by discussing the two primary correlates of [voice] to be investigated, and in the case of vowel duration motivating its consideration as a phonological process in adults (sections 2 and 3). I then turn to background on the acquisition of [voice] and its phonetic correlates (section 4). Then, I present some preliminary findings based on my own work conducted over the last six months (section 5). Finally, I outline the dissertation work being proposed (section 6).

## 2 The vowel duration correlate

The post-vocalic consonant voicing effect or vowel length effect (VLE, House and Fairbanks 1953 among many others, also ‘pre-fortis clipping’ in Harris 1994 as cited by Jansen 2004) is the cross-linguistic phenomenon in which the duration of a vowel is longer when preceding a voiced versus voiceless obstruent. The phenomenon is often described formally but descriptively as the distinction between /bæt/ versus /bæ:d/ or  $V \rightarrow [+long] / \_\_\_ [+voice]$  in the style of Chomsky and Halle (1968), though it is by no means a settled issue whether the phenomenon ought to be described in terms of phonological lengthening, or whether the VLE is even a single phenomenon. The VLE operates within and across syllable (Chen, 1970) and word (Huff, 1980) boundaries.

The literature on the VLE has often portrayed it as a universal phenomenon but with a peculiar existence in English. Its apparent universality suggested early on that the VLE was a low-level possibly articulatorily effect, something necessitated by other factors such as the airflow differences in voiced and unvoiced obstruents, and so not a part of linguistic competence. The VLE has been claimed to exist at least in Danish, Dutch, French (see page 8), German, Hindi, Hungarian, Italian, Korean, Norwegian, Persian, Russian, Spanish, and Swedish (see for references Kluender et al., 1988), and of course English. (The VLE has been

found to not be present in Polish and Czech (Keating, 1985) and Saudi Arabian colloquial Arabic (Flege and Port, 1981).) However, English shows a much larger duration difference between voiced and unvoiced consonants than all other languages in which the VLE has been studied — suggesting instead that, at least in English, the effect is learned. Taking the results from Chen (1970) as a representative example, the ratio of the mean duration of vowels before voiced consonants to that before voiceless consonants in English is roughly 1.5 with an absolute difference between the mean durations of 92ms, while in French, Russian, Korean, etc. the ratio and difference are 1.2–1.3 and 28–53ms. I also report possibly for the first time below in my own production study of the VLE (section 5.1) that the two means are separated by 2–6 standard deviations, indicating the tokens are quite separable. There are, then, apparently two VLEs. There is first the cross-linguistically observed, smaller VLE and second the English-specific, larger VLE.

A number of factors affect the VLE, at different levels of linguistic structure. The effect is much larger in prepausal position (a ratio of 1.5) than elsewhere (1.4 word-finally, 1.1 word-medially) (Umeda 1975, but also see Luce and Charles-Luce 1985). The effect is also larger in monosyllables (1.5) than in the first, stressed syllable in disyllabic words (1.3) (Klatt 1973 but also see Port 1981). It is also larger for vowels with greater intrinsic duration (Luce and Charles-Luce, 1985). The VLE is greater in slow ‘tempo’ speech (1.18) than in fast ‘tempo’ speech (1.09) (Port 1981; Laeufer also notes work by Harris and Umeda in 1974). Laeufer also reported from other work as magnifying the VLE the vowel being stressed, and, at least in French, preceding fricatives (though this was not the case in the English results reported by Umeda 1975).

It is useful to put the VLE in the context of other aspects of vowel length. The VLE is just one of many factors that influence the duration of vowels: speaking rate, focus, intrinsic vowel duration, and the presence of (syllabic) primary stress, and there are the phenomena of pre-pausal, pre-boundary, and word-final lengthening (for a summary, see Klatt, 1976). These factors have made it difficult not only to ascertain the magnitude of the VLE in English, but to compare the VLE between languages. We can also compare the VLE to different types of durational phenomena in other languages. The magnitude of the VLE in English is still smaller than durational differences associated with a phonological length contrast, such as in Dutch (see for a summary Dietrich, 2006), for which van Bergem reported that long vowels were 1.8 times longer than short vowels. The VLE is outdone also by the Scottish Vowel Length Rule (Scobbie et al., 1999), which among other things involves lengthening by a factor of also 1.8 before voiced fricatives.

I turn next to several explanations that have been proposed for the VLE and evidence that in English it is a part of linguistic competence.

## 2.1 The grammatical status of the VLE

As I am suggesting here dissertation work on the acquisition of the VLE (and other phenomena), it is important that I motivate the consideration of the VLE as a learned component of the grammar of English. The most low-level explanation of the VLE would be that the VLE follows from constraints on articulation. Something inherent about laryngeal vibrations or some other difference in articulation of voiced and voiceless obstruents causes the duration of preceding vowels be longer. An articulatory explanation of this sort is essentially non-linguistic, that is, not a part of learned language competence, and while it would address

the lower-magnitude (i.e. smaller voiced/unvoiced ratio) cross-linguistically observed VLE, it would not explain why English has a much larger effect.

Articulatory explanations trace their roots back over half a century. As a straw-man argument, Chen (1970) starts with Otto Jespersen's articulatory distance hypothesis that vowel duration is a function of its distance to adjacent consonants, but quickly dismisses the possibility since vowels appear to have more in common with voiced consonants, being voiced themselves, before which they have the *longer*, rather than shorter, duration.

Chen proposed instead a story based on closure transition duration which goes something like this: Voiced and voiceless consonants differ in the state of the glottis during closure, voiced consonants having a closed glottis and voiceless consonants an open glottis. As a result, intraoral air pressure during consonant closure is higher for a voiceless consonant since the build-up of air pressure extends into the lungs, whereas for a voiced consonant it extends only to the glottis. More effort is then needed to form the closure in a voiceless consonant, and because this greater effort may begin during the transition to closure, the increased muscle effort may translate into increased velocity, a more rapid transition, and therefore a shorter vowel. Chen found in English that for the bilabial stops the duration of lower lip movement was longer by a mean of 27ms for voiced stops, which is in line with only the cross-linguistically observed VLE magnitude but was too small to account for the vowel length difference in English (92ms) or even French (53ms). Kluender et al. (1988) note that later research has shown that closure velocity is not reliably greater in voiceless consonants, at least not in the contexts that show the VLE.

But later work (on English) cast doubt on purely articulatory explanations. Fox and Terbeek (1977) found (though with a very small sample size) that vowel duration before flaps does not correspond to the presence of voicing during the flap (i.e. whether the flap is voiced or unvoiced). Additionally, the VLE persists in whispered speech at the same magnitude as in normally phonated speech (a magnitude as high as 1.8 in Sharf, 1964), and laryngectomized patients who phonate using the esophagus and need neither laryngeal nor pulmonary articulation also exhibit a very high magnitude VLE: 1.6 ratio in normal subjects, 1.7 in esophageal speech, the difference likely due to slower speaking rate in esophageal patients (Gandour et al., 1980).

Because stop closure duration also correlates with [voice] — *less* for voiced consonants — it had also been proposed that the duration of the vowel varies inversely with the duration of the stop in order to maintain constant syllable duration (Chen (1970) cites Kozhevnikov and Christovich (1967:107) and B. Lindblom (1967:21) for the hypotheses of compensatory temporal lengthening.) This hypothesis is both articulatory and phonological in nature. Chen investigated the possibility with, on the whole, no strong indication that there was such a thing as compensatory temporal lengthening of vowels in English. Kluender et al. (1988) reported other work on English and Swedish indicating that while there may be such a thing as compensatory vowel duration adjustments made inversely to other durations in the syllable, the compensation is neither total nor large enough to explain the VLE.

Taking another direction entirely, Kluender et al. (1988, p. 156) proposed that the VLE exists because speakers “select acoustic cues that have mutually reinforcing auditory effects. Talkers signal phonetic contrasts using a ‘conspiracy’ of cues that enhances the perceptual distinctiveness of features and segments.” (This is along the lines of the secondary features of Stevens et al. 1986.) In the case at hand, the VLE exaggerates the closure duration cue. By shortening vowels before voiceless consonants, the increased closure duration of voiceless

consonants is perceived to be longer than it otherwise would be. They note, further, that the VLE appears to be one instance in a class of vowel inverse-duration phenomena. In a variety of languages with contrastive consonant length vowels are shorter before phonemically long or geminate consonants, including Arabic, Italian, Finnish, and many others. Kluender et al.'s story though enticing is not an *explanation* since it would not be a contradiction to find a language with a closure contrast but not a vowel contrast, or even necessarily unexpected. (One would like to check Polish and Czech.) Kluender et al. (1988) report that they knew of only one language lacking a closure duration correlate, Arabic, for which they predicted no vowel duration correlate, and to their benefit it has been one of the few languages reported to not show any VLE. But a sample of one is not a good basis for making any generalizations. This hypothesis also does not address why in English the VLE might be greater than in other languages. Kluender et al. denied the difference. (More on this in section 2.2.)

If not articulatory, the grammatical status of the VLE might roughly fall in the range of being a secondary phonetic feature or gesture recruited to support phonetic [voice] (Stevens et al., 1986; Kluender et al., 1988), a part of the phonetic implementation of the phonological feature [voice] (Keating, 1984), or a separate length feature whose value is set by rule (rules in the style of Chomsky and Halle 1968). In the latter case, the rule may have different possible orderings relative to other rules related to voice (and conceivably length), such as the rule for English /t,d/ flapping in intervocalic context. Minimal pairs such as the well-known 'writer/rider' (both /raɪrɪə/) and 'latter/ladder' (both /lætər/) might or might not, a priori, show a vowel length difference. If they don't, the voicing contrast relevant for the VLE is neutralized by the alternation to the common allophone — the VLE would be a rule applied after the flapping rule. If the VLE shows up, then at the time the VLE takes effect information about the underlying form of the flap is still available — most sensibly because the flapping rule occurs later.

Joos (1942) claimed to find each pattern in Canadian English dialects at the time of his writing, strongly suggesting the VLE was a phonological rule-type process at least in the dialect with the /raɪrɪə/—/raɪrɪə/ distinction. (I do not intend to rely on an ordered-rule framework, but it is convenient for exposition.) Neither rule ordering precisely covers the facts observed apparently universally in English today. Vowels preceding flaps continue to show a vowel length difference, but with a difference on the order of the cross-linguistic pattern, rather than the usual English-level larger magnitude. The original data is due to Fox and Terbeek (1977), probably of Chicago speakers, though their study lacked a control condition with non-flapped consonants and it did not look across place of articulation, as a result collapsing the potential effects of both flapping and syllable structure, and they surprisingly did not report mean values. The facts were later investigated in New York City speakers by Huff (1980). Huff considered monosyllabic words with coda /t,d/ in a frame sentence that continues with an initial unstressed vowel, e.g. 'Say bite/bide again.' (both /saɪbaɪtəɡən/). This context induces flapping of the /t,d/. A voiced-voiceless vowel duration difference was found. Again, however, no non-flap baseline was included. The issue of lacking a proper control were resolved in the experiment I carried out, discussed in section 5.1, though it confirmed that before flapping the voicing contrast is considerably reduced.

My work failed to find a VLE before flapping, but if the results of Fox and Terbeek (1977) and Huff (1980) are correct that some small VLE remains after flapping, then one might call it an instance of 'incomplete neutralization', a topic unto itself (see e.g. Port and

O'Dell 1986, Port 1996, and for a study, though of disputable quality, relevant to my work (Baran and Seymour 1976). Before drawing any conclusions, I will report another case of neutralization that we might expect to interact with the VLE.

Regressive voicing assimilation (RVA), which, roughly, neutralizes a voice contrast in a preceding segment, also plays a role in understanding the VLE. An example of RVA (based on one from Jansen 2004) is the different realization of /z/ in 'Is Bob/Pete going?' [ɪzɒb...] vs. [ɪspɪt...]. Jansen (2004, 2007) reported results for "vowel—velar stop"—[consonant] sequences embedded within a pair of words in British English (e.g. 'brickw[ɔrk t]epot' vs. 'brickw[ɔrk t]unnel' vs. 'Hamb[ɜrg tɛnənt]' vs 'Hamb[ɜrg d]airy', etc. Note that the subjects were r-less.). By varying the (underlying) voicing of the first consonant ("C<sub>1</sub>") between /k,g/ and the voicing and manner of articulation of the final consonant ("C<sub>2</sub>") between /r,t,d,s,z/ (/r/ was considered a baseline), the effect of voicing assimilation on the correlates of voicing of the first consonant could be observed. Closure, release, and preceding vowel durations and  $F_0$  and  $F_1$  were measured. Results were near-categorical. While the duration of voicing during closure and release of C<sub>1</sub> was greater by 21ms or roughly a factor of 2 for /g/ than for /k/ before /r/, when preceding /t/ or /z/, for instance, the voicing duration difference was reduced to 6ms or a factor of 1.1.  $F_1$  of a vowel preceding a consonant is normally correlated with the voicing feature: it is lower before voiced consonants, and specifically in the context of the baseline /r/ reported lower by 26 Hz. When C<sub>2</sub> was /t/, the  $F_1$  difference was completely eliminated, and it was considerably reduced for /d/ and /z/ as well. But while voicing duration and preceding vowel  $F_1$  were nearly neutralized, there was virtually no effect on C<sub>1</sub>'s closure duration or preceding vowel duration. The fact that two cues were affected by assimilation, voicing duration and  $F_1$  on the preceding vowel, and two cues were not, closure duration and preceding vowel duration, is consistent with and in support of the lack of an articulatory connection between the VLE and at least some cues to the voicing contrast. At the same time it is interesting that vowel duration and closure duration pattern together, in light of Kluender et al.'s hypothesis that the former cue is used because it enhances the latter. The simplest explanation of the pattern seen here is that RVA is merely coarticulation (Mark Liberman, p.c.).

The flapping and RVA data help to narrow down the point in the grammar where the VLE is operating. According to the work showing that flapping reduces the duration difference of the VLE (Joos 1942 notwithstanding), we do not have evidence that the VLE is necessarily an ordered rule, only that it is not ordered before flapping. Despite a residual VLE that remains after flapping, I do not take this as evidence that the VLE must be a phonological rule ordered before the rule for flapping. It is possible this small residual VLE that remains is due to the type of VLE found cross-linguistically acting in parallel, or it is a case of incomplete neutralization. I do not pursue this any further. The remaining evidence, the whispered and esophageal speech data, Chen's (1970) failure to show an explanation in terms of articulatory force or compensatory temporal adjustment, and its dissociation from both the voicing and formant frequency correlates in regressive voicing assimilation, strongly suggest that the VLE is at least matter of linguistic competence. That leaves several possible explanations still available, but this is a sufficient understanding for the purposes of this proposal and dissertation.

## 2.2 An open issue for cross-linguistic comparisons

Though the evidence points to the English effect being different from the cross-linguistically observed effect, it is still an open question as proper cross-linguistic comparisons have been rare. I do not believe that there is much doubt that the English phenomenon is grammatical, but it remains to be shown robustly that English is unique.

Independent durational and prosodic differences between English and French concerned Laeuffer (1992) for they made previous work comparing the languages difficult to interpret. French vowels are often shorter than their English counterparts; English final stops are longer and are often, unlike in French, unreleased and when they are released their releases are shorter than those in French; French pre-pausal voiced stops are often (74%) followed by a vocalic release (a voicing cue unavailable in English that may relieve the VLE of its communicative load); and varieties of French, unlike English, lack syllabic stress. Controlling for some of these factors in the two languages, Laeuffer (1992) found a VLE ratio in French of 1.42 and in English 1.6, fairly close but still seemingly different. Laeuffer noted, however, that vowel durations were not consistent between the two languages (in one condition 150ms and 209ms, respectively), and this may account for the VLE difference. Klatt (1973), for instance, suggested a notion of incompressibility, which has the effect that durational changes will apply differently depending on the duration of the segment, limiting how short a segment can be. When French vowels had similar durations to English vowels, sentence-final position in French and sentence-medial position in English, the VLE magnitudes were much closer.

Laeuffer's (1992) work is important in several respects. First, it highlighted language-specific details that make cross-linguistic comparisons susceptible to confounding factors. Second, in light of the fact that independent vowel duration factors interact with the VLE (Klatt, 1973; Port, 1981) and that when vowel durations were matched between English and French the VLE magnitudes were very similar, it is not entirely clear that English and French differ in their respective VLE's. This casts doubt on English actually being unique. The extent to which English really differs from languages besides French reported to show a VLE is debatable and warrants significant further study.

## 3 The /s,z/ contrast and /z/-devoicing

The second primary cue of interest is voicing during frication, especially as it relates to /z/-devoicing, the extremely common realization of /z/ without glottal voicing throughout the duration of frication. Starting at least with Denes (1955) it was known that the /s,z/ distinction was made by listeners at least in part based on the duration of the segment. Decreased segment duration (and also increased preceding vowel duration, as we now expect) lead to increased identification of segments as /z/, even if the segment has no glottal voicing.

This is likely so because glottal voicing in /z/ is not robustly present. Smith (1997) documented the rate and magnitude of /z/-devoicing and compared the correlates of voicing in voiced /z/, devoiced /z/, and /s/. Based on electroglottographic data, in sentence-final position all /z/ tokens she collected were devoiced, meaning there was vocal cord vibration for less than 25% of the duration of frication. In word-final position followed by a voiceless consonant /z/ tokens were at least partially devoiced, meaning there between 25%-90% of the frication duration had vocal cord vibration. But, in word-medial position also followed by a

voiceless consonant /z/ tokens were generally partially or fully voiced (fully voiced meaning over 90%). /z/ was also overall less voiced in an unstressed syllable than in a stressed syllable, although this pattern was not found to be robust, and more often completely devoiced before voiceless /p/ than voiced /b/.

Despite being ‘devoiced’, the so-called partially devoiced and devoiced tokens of /z/ continued to differ from /s/ tokens in the same context in their frication duration and in the preceding vowel duration. /z/’s were approximately 20% shorter in frication duration and had approximately 20% longer preceding vowels. Maximum airflow during the frication also was different between /z/ and /s/ in the three categories of voicing, with even devoiced /z/ having less maximum airflow, suggesting that the glottal position was somewhere between the open state of unvoiced /s/ allowing greatest airflow and the constricted state of a fully voiced /z/ impeding airflow.

Similar to our interpretation of regressive voicing assimilation (page 7), Smith (1997) interpreted these results as indicating that /z/-devoicing was due to assimilation and/or lenition in laryngeal gestures and aerodynamic effects, such as decreased subglottal pressure over time during a sentence making voicing difficult to maintain.

## 4 Acquisition of [voice] and its correlates

While it is useful to talk about phonetic or phonological ‘knowledge’ in children, it is clear that there may be at least two aspects of this knowledge: perceptual capabilities and production capabilities. This dissertation is concerned primarily with the production capabilities of children, since it will be a corpus study.

Work by Dan Swingley, Christiane Dietrich, and Janet Werker, among others, has asked the questions of what types of phonetic contrasts the child can distinguish perceptually and store in their lexicon (that is, tied to words). Dietrich et al. (2007) showed that Dutch-speaking children (but not English-speaking children) at age 1;6 can learn words that differ in phonological vowel length and then notice when the words are incorrectly matched with visual stimuli. In English, of course, there is no phonological vowel length, and it is in the interests of the child to ignore duration differences when it has no linguistic import (e.g. not in the case at hand when it is a cue to voicing). As for distinguishing [voice], a categorical VOT distinction in perception in English is acquired at least as early as 1 month, as determined by a dishabituation experimental design. As van der Feest (2007) noted, a perceptual distinction does not indicate that the voice distinction has been encoded in words in the lexicon. Other work has shown that at age 1;5 English-learning infants can learn a distinction between two words that differ only in the voicing feature of an initial consonant. (For a summary see van der Feest 2007.) van der Feest’s (2007) own work showed that for Dutch-learning infants, at age 1;8 they do not detect a voice mispronunciation in learned words but at age 2 they detect words with segments mispronounced as voiced.

On the production end, which will be the concern for the remainder of this proposal, Snow (1997) found that for word-initial stops, Voice Onset Time was different between voiced and unvoiced stops starting at least by age 1;6 (the earliest age considered). Children at that age produced 41 ms more aspiration in voiceless stops, and 73 ms at age 1;9. (Snow claimed the lower bound of appropriate adult-like VOT difference is 60 ms.) I review the use of the VLE in the production of the [voice] contrast in the next section.

### 4.1 Acquisition of the VLE in English

The earliest work on the acquisition of the VLE appears to be in an unpublished doctoral dissertation by M.A. Naeser in 1970 at the University of Wisconsin, which Krause (1982) reported as finding a vowel length difference before voiced and voiceless consonants in the spontaneous and imitated speech of age-1;9 (and up) children. Both experimental and corpus studies since then have confirmed the conclusion. Taking the results in reverse chronological order by age, Baran and Seymour (1976) reported a VLE at age 5 in AAVE speakers. Both DiSimoni (1974) and Krause (1982) found a durational difference at age 3, a ratio of 1.86 for the three pairs of stops in Krause (1982), and at age 3;7 a ratio of 1.78 (/p,b,s,z/ only; estimating from a chart; reported not-significant) in DiSimoni (1974). At age 2;6 and 2;0 ratios of roughly 1.7 and 1.3, respectively, estimating from a chart (Buder and Stoel-Gammon, 2002). Ko (2007) reported a difference before the age of two, based on a corpus study. In my own work, described in section 5.2, a duration difference was found at age 1;6.

DiSimoni (1974) and Krause (1982) also both found a difference at age 6 (and 9 in DiSimoni 1974), but the developmental pattern diverged between the two studies. While both found that the durations of vowels before unvoiced consonants was relatively stable from age 3 to 9 (DiSimoni) or adult-age (Krause), DiSimoni reported an increase in duration for voiced consonants and Krause a statistically significant change in the reverse direction. The trend is thus not clear, but drawing conclusions about the developmental pattern from such a trend would be premature in any case. Overall developmental vowel duration changes will affect our expectations for VLE magnitude, since we know that the VLE voiced/unvoiced ratio increases with decreased tempo (Port, 1981). Lee et al. (1999) reported vowel duration decreases by 25% from 5 to 7 years of age and roughly a 3-4ms decrease per year until age 15. In light of this, the findings of roughly constant vowel duration preceding unvoiced consonants is quite a surprise, and it is not clear what to make of the changes reported in the VLE ratio.

Although the voiced-unvoiced ratio is consistently greater than 1.0 even at early ages, the grammatical status of the VLE in children is still not clear. The apparent VLE in children may not be a productive process, for instance. Instead, it might yet be articulatory in nature, despite the evidence against such a hypothesis for adults. Or, vowel durations may be encoded in the lexicon, either as phonetic detail or as a phonemic vowel duration contrast. Or, if the pattern is in fact productive, does it interact with the rule of flapping in the same manner as in adults?

An articulatory account is, actually, also unlikely at early ages. Children often omit word-final stops. In a study of three children age 3;10-7;6 with clinical problems that likely exacerbated their rate of word-final stop deletion to rates ranging from 0 to 97%, all three children showed the VLE in utterance-final context when the consonant itself was deleted. The youngest and most impaired child's ratio was 1.11, however — far lower than is expected. The two other children had ratios of 1.3 and 1.6 (Weismer et al., 1981). Ko (p.c.) reports, based on a continuation of her corpus work, preliminary findings that the VLE also persists when the child has devoiced the consonant (I believe primarily considering stops). These findings suggest that even in children the VLE is not articulatorily conditioned.

## 4.2 Acquisition of the fricatives

While the VLE pattern has been observed in children as early as age 1;6, fricatives besides [h] are not observed in children’s phonetic inventories until age 2;0 (Stoel-Gammon, 1985), where in children’s inventories means in at least half of the sampled children’s inventories in hour-long recording sessions. In word-initial position, voiced-unvoiced stop pairs appeared in inventories between 1;6–1;9 and fricative pairs between 1;9–2;0. In word-final position, which is the position of interest here in order to have a preceding vowel, phonetic inventories were smaller at each age. No voiced segments were observed through age 2;0, when the experiment ended. Though [t] was present at 1;6, [s] was only present at 2;0. This data roughly agrees with the corpus based on a single child used in my work reported in section 5.2, which found voiced-voiceless stop pairs and [ʃ,ʒ] at age 1;6, [s] at age 1;10, and [z] at age 2;1.

## 5 Preliminary Work

Two experiments were conducted in preparation for this proposal. The first experiment, a production study of adult speech, sought to verify claims in the literature about the VLE that were not sufficiently backed with a proper experimental design. Results roughly confirmed what was believed to be the case previously. The second experiment was a corpus study of child speech, age 1;6–4;0, intended to measure the magnitude of the VLE over time. While the experiment was (coincidentally) based on the same corpus used by Ko (2007), the methodology yielded a more comprehensive picture. It also sought to find a relationship, if any, between the developing VLE and changes in the child’s speaking rate from session to session, and to survey the data in preparation for further work.

Both experiments should be considered preliminary.

### 5.1 Experiment A: Confirmations of several adult hypotheses

The primary goal of Experiment A was to establish the effect of syllable structure on the magnitude of the VLE, and to separate from this the effect of flapping. I also addressed two other aspects of the VLE in light of the future research direction in the acquisition of the VLE: the effect of speaking rate and whether the VLE is a productive process.

The common wisdom regarding the VLE and flapping had been that flapping induces incomplete neutralization. This was supported by Fox and Terbeek (1977) and Huff (1980), as discussed in section 2.1. However, while both studies reported positive VLE magnitudes and smaller than what is usually found in English, neither study included a control group. Two problems arose in interpreting the results. First, because the VLE magnitude is highly variable depending on speaking task and phonological context, without a comparison to same-task data without a flap it was premature to make a strong statement that flapping in fact decreases the VLE. Second, because flapping occurs in a particular context, it had not been shown that the incomplete neutralization (descriptively speaking) has to do with the flapping or the context. It might have been that the VLE was reduced because the post-vocalic consonant was not word-final (Klatt, 1973) or because it syllabified with the vowel to its right, regardless of the consonant’s place and manner of articulation. An effect of syllable structure would be quite reasonable because in syllable-initial position other [voice]

cues such as aspiration become available, lightening the communicative load of the VLE. I answer these questions below.

The high variability of speaking rate in the developing child has prevented measures of the VLE magnitude in children from being properly compared with other measures at different ages. Because of the importance of understanding the interaction of speaking rate and the VLE, I sought to replicate some of the work of Port (1981), who showed that the VLE is larger in slow ‘tempo’ speech than in fast ‘tempo’ speech. I intend to use the results of manipulating this variable to understand whether changes in vowel durations in children are a result of changing speaking rate or the development of the VLE.

Because I hypothesize that at some positive age children exhibit the VLE not because of a productive process but instead resulting from encoding durational phonetic details in the lexicon, I anticipated future work directly addressing the question of whether the VLE is productive in children. For a fair comparison, I tested whether the VLE is productive in adults by having adults speak nonsense words. Though I fully expected the process to be productive, an effect of processing an unfamiliar word may nevertheless affect the magnitude of the VLE. Eliciting this data from adults was a necessary baseline for future comparisons with children.

### 5.1.1 Methods

Four undergraduate and graduate students, native speakers of English, participated in the experiment, three associated with the linguistics program at the University. The participants were compensated \$15 for their time. The participants were asked to read a list of sentences provided to them, which they had not seen before. Recordings were made in the University’s linguistic department’s Phonetics Lab’s sound booth at 44,000 Hz.

Each sentence was a frame sentence of the form “Say \_\_\_\_ for me.” containing a target word. Target words contained a VC sequence either in a monosyllabic word, necessarily in the first syllable of a disyllabic word (the ‘tautosyllabic’ condition), or crossing a syllable boundary in a disyllabic word (the ‘heterosyllabic’ condition). The words came in minimal or near-minimal pairs, differing in the voicing of C and potentially in the segmental content after C (when no minimal pair could be found). They were drawn from a list of 163 real English and novel nonsense words. The disyllabic real words were trochaic. Vowels were mixed and included both monophthongs and diphthongs. C ranged over the six stop consonants /p,b,t,d,k,g/. It was expected that flapping would occur for /t,d/ in the heterosyllabic target words, and nowhere else. We will call these ‘flap target words’. The target words are listed in Appendix A. Some words given to the participants were discarded due to experimenter error. Sample target words were:

Real: thought, thawed (monosyllabic); crapshoot, crabmeat (tautosyllabic); seater, cedar (heterosyllabic)

Nonsense: chack, chag (monosyllabic); geetmonk, geedmonk (tautosyllabic); nuckist, nugist (heterosyllabic)

The reading task was divided into five blocks. In each block, subjects were given each of the sentences to read three times, presented in a random order (although subject confusion resulting from double-sided instructions resulted in some words being read fewer or more times). Each block of the task presented the subjects with either the real words or the nonsense words and instructed them to speak in either a normal, slow, or fast speaking

rate. Prior to the recording subjects were informally told what a fast or slow speaking rate might sound like, with spoken examples from the experimenter (me; but not using any of the target words). The sequence of experimental blocks is given in Figure 1. Note that the nonsense words were only recorded at the normal speaking rate, and the second iteration of nonsense words in block 4 was a filler to allow the participants to find their normal speaking rate again between switching from slow to fast. The recordings of block 4 were discarded. Excluding block 4, an average of 967 tokens were recorded per participant.

Block	Word List	Speaking Rate
1.	Real	Normal
2.	Fake	Normal
3.	Real	Slow
4.	Fake	Normal (filler)
5.	Real	Fast

Figure 1: Experimental design for Experiment A.

Recordings were passed through the Phonetics Lab’s forced alignment toolkit (Yuan and Liberman). The forced aligner, based on the HMM tools provided by the HTK software, time-aligns phone boundaries according to an orthographic transcript and pronunciation dictionary. As the participants were asked to read from a list, a transcript was almost readily available — almost, owing to the fact that most participants strayed from the instructions somewhere. Phone boundaries for the target vowels were then manually corrected based on spectrograms, with vowels located by clear formant structure and a fast intensity rise and decline at the start and end. Although I attempted to minimize it, some target words contained sonorant consonants preceding the target vowels. In these cases, determining the boundary with the following vowel was difficult. Formant transitions were sometimes used, but often no clear boundary was evident and the boundary was left where the forced aligner located it.

### 5.1.2 Results

Within each (near-)minimal pair, vowel durations were averaged for each half of the pair, and a VLE magnitude was computed for each pair by speaker as the ratio of the mean voiced duration to the mean unvoiced duration in that speaker’s tokens. Only minimal pairs for which the speaker spoke each half of the pair at least twice were included. The mean VLE magnitude (now averaging across pairs) varied widely from speaker to speaker. Considering the real, monosyllabic words at a normal speaking rate only, the median voiced/unvoiced ratio ranged among the speakers from 1.25 to 1.70 (a one-way ANOVA indicated a highly significant main effect for speaker,  $p < .001$ ). This is likely a result of variability in speaking rate and flap production.

While the ratio indicates the magnitude of the difference on the whole, it does not indicate the separability of the two groups. High variance in vowel duration could mean that the distributions of durations greatly overlap and so a listener could not reliably determine the category from the token, even if the means are far apart. To measure the separability of the distributions, the number of standard deviations between the mean voiced and mean

unvoiced durations were computed. I call this the separability z-score below. (Because variance was seen to increase with duration, and the unvoiced and voiced categories differ in duration, for the computation of separability durations were considered on a log scale so that the unvoiced and voiced categories would have a similar variance. See Rosen 2005.) As with the ratio computation, the separability z-scores were computed individually by minimal pair and by speaker, and so each individual z-score data point ranges over roughly three tokens of each half of the minimal pair. By speaker, mean separability z-scores ranged from 2.5 to 5.9 standard deviations.

The first question was whether the apparent reduction in VLE magnitude seen for flapping (Fox and Terbeek, 1977; Huff, 1980) was in fact true, and whether it was due to flapping or a syllable boundary separating the vowel from the post-vocalic consonant. Indeed, I found also that in flap target words the VLE ratio is smaller. The mean VLE magnitude for real disyllable words at a normal speaking rate, excluding the flap target words, was 1.29. For the flap target words, the mean ratio was 1.07. This was a highly significant difference (two-tailed t-test,  $p < .001$ ). In the latter case at least some and perhaps many of the [t,d] segments were not realized as flaps — this was not checked comprehensively.

This difference was not due to the syllable boundary between the vowel and consonant. Excluding flap target words, the VLE magnitudes in the tautosyllabic and heterosyllabic groups (real words at normal speaking rate) were virtually the same, 1.23–1.24. Nor was the effect due only to place of articulation: in the tautosyllabic group there was no significant main effect. (See Figure 2.) Figure 3 shows the VLE magnitude by place of articulation in the heterosyllabic group only, showing the /t,d/ pairs to be different. (The difference was highly significant,  $p < .001$ ; Speaker 1 was excluded as she was impressionistically deemed to not consistently flap.)

The second question was what the effect of speaking rate is on the VLE. Recall that speakers were asked to speak at normal, slow, and fast speeds, and given brief examples, but their actual tempo was otherwise not enforced by the experimental design. Whether the speakers adjusted their tempo according to the instructions was determined by measuring mean vowel duration (the vowels in the relevant VC pairs) and the duration of the frame sentence minus the duration of the vowel. Both measurements show speakers were able to adjust their speaking rate, although the distributions for the three speaking rates were considerably overlapping. The VLE ratio also varied by speaking rate in the expected manner, from 1.19 in the fast condition to 1.35 in the slow condition (Figure 4). The difference was highly significant (one-way ANOVA,  $p < .0001$ ), although the differences between speaking rates are comparatively smaller than what we have seen elsewhere.

The last question was whether the phenomenon is productive. It appears to be so. The VLE magnitude was nearly identical for real and nonsense words (Figure 5). I must note, however, that participants were free to choose their own pronunciations for the nonsense words and frequently chose different vowels for different halves of a minimal pair. This leaves open the possibility that phonotactic regularities may make it appear that a VLE exists, if participants tended to choose vowels with a longer inherent duration in pre-voiced-stop context. Annotating the vowels chosen by the participants would be a useful next step.

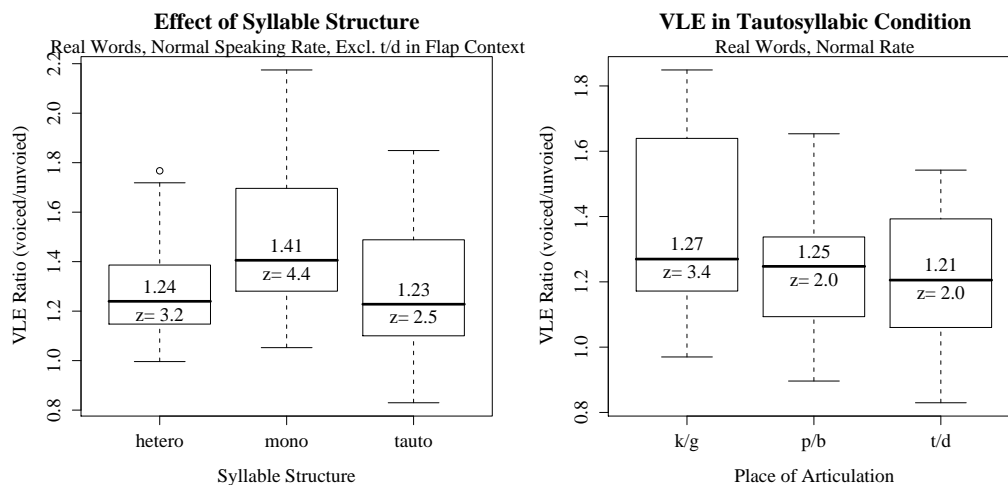


Figure 2: VLE magnitudes. Data points summarized by the boxes are the voiced/unvoiced ratio for each minimal pair separate by speaker. Left: By syllable structure, excluding flap target words in the heterosyllabic group. Right: By place of articulation in the tautosyllabic group. Median values are indicated above the median line in the boxes. Below the median line, the mean separability z-score is reported.

### 5.1.3 Discussion

To summarize the results so far, I have replicated the conclusions of past work, in some cases with a proper control condition. The VLE magnitude was previously found by Klatt (1973) to be greater in monosyllables (1.5) than in the first, stressed syllable in disyllabic words (1.3). I found similar ratios: 1.41 and 1.24, respectively. Fox and Terbeek (1977) and Huff (1980) previously reported small magnitudes for flaps. I found no significant effect of adding a syllable boundary between the vowel and consonant (tauto- vs. heterosyllabic conditions) or of place of articulation, except for the flap target words, which had a ratio of 1.05 (not significantly different from 1.0). I also found an effect of speaking rate, though it is fairly small, with increased speaking rate correlated with decreased VLE magnitude. Finally, I found that the VLE is a productive process by showing that the magnitude was the same in real and nonsense words, though with the caveat mentioned above.

With regard to flapping, the experiment here as well as that of Fox and Terbeek (1977) crucially relied on the assumption that speakers have the underlying forms for the stops that we think they do. For the flapped pairs in my experiment like ‘petal’/‘pedal’, it is not a trivial assumption that the speakers have an underlying /t/ and /d/ when the surface forms differ by perhaps a relatively small vowel duration (assuming they did not learn English in a dialect with vowel quality changes contingent on the following voice feature, Joos 1942, Huff 1980), and when the words have no (or common) morphological alternations that bleed flapping to reveal the underlying difference. Without a distinction in underlying form, no vowel duration difference would be found — just as was the case here. A better

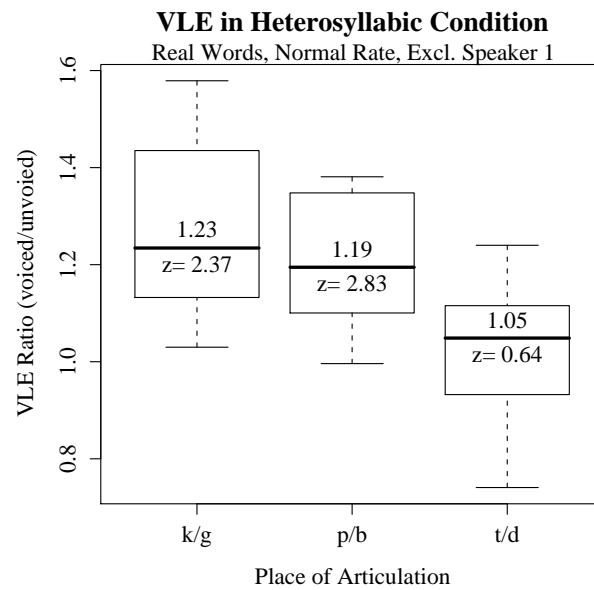


Figure 3: VLE magnitude by place of articulation in the heterosyllabic group. Speaker 1 was excluded from this graph since she (impressionistically) failed to flap more often than other participants.

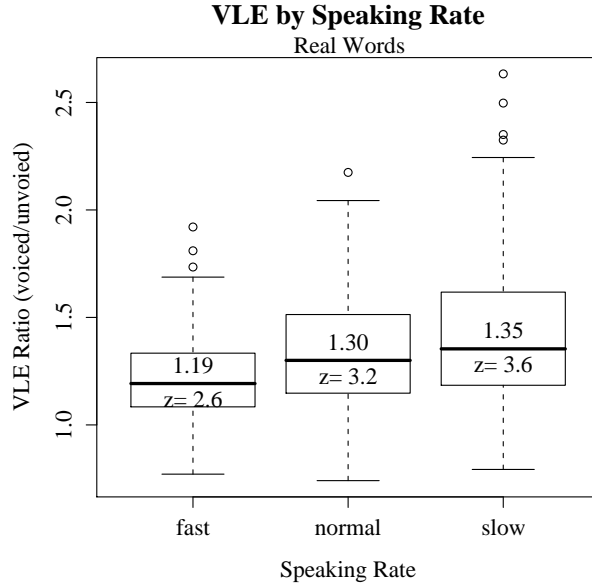


Figure 4: VLE magnitude by speaking rate.

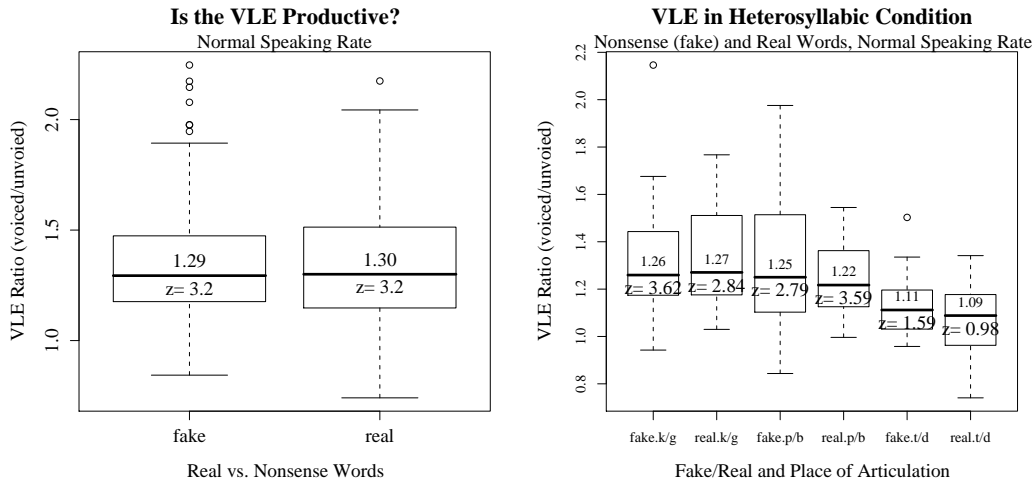


Figure 5: VLE magnitudes: Left: Real and nonsense words at a normal speaking rate. Right: Real and nonsense words in the heterosyllabic group, by place of articulation.

test was that used by Huff (1980), flapping across word boundaries, where the underlying form surfaces except when followed by a vowel-initial word. Huff found a fairly large ratio for the flapping context, 1.2, though we do not know what kind of confidence interval to apply and the number of tokens considered by Huff was relatively small. Huff may have found that flapping is not a case of incomplete neutralization but instead evidence that the VLE applies before flapping. I leave resolving this for future work.

## 5.2 Experiment B: Acquisition

I also conducted a corpus study of child language age 1;6-4;0 to track the development of the VLE over time. This work was (coincidentally) quite similar to the work by Ko (2007). The corpus used here was a subset of that used by Ko, a part of the CHILDES Providence corpus (Demuth et al., 2006). However, the approach differed. While Ko considered only minimal and near-minimal pairs, I aggregated vowel durations across words and then computed a ratio. Because even near-minimal pairs are difficult to find and still vary considerably within the pair from factors such as the excitingness of the word, it was decided to aggregate across words. Lastly, I sought to follow the VLE ratio over time and its relation to the child's speaking rate, which was not reported by Ko.

In preparation for further work, I also counted the occurrences of stops, fricatives, and flaps at different ages, based on the broad phonetic transcription provided with the corpus, and their distribution in content and function words and derivational and inflectional affixes. I also made preliminary measurements of frication duration and voicing duration in /s/ and /z/ segments.

### 5.2.1 Methods

Audio recordings of the child 'Lily' in the Providence corpus (Demuth et al., 2006) were extracted at ages 1;5.22, 1;6.11, 1;6.20, 1;10, 2;1, 2;8, and 4;0. The first three age levels were collapsed into a single 1;6 age to pool data. Recordings were matched with the transcript using time markers already present in the transcript. Only the first 50–100 utterances in each recording were used.

For each utterance in the transcript, the Phonetics Lab's forced alignment toolkit (Yuan and Liberman), based on the HMM tools provided by the HTK software, was used to align the phonetic transcription in the corpus. Normally the forced aligner is provided a word-level transcription, but in this case although a word-level transcription is provided in the corpus I provided the aligner with the (shallow) phonetic transcription directly, as the child speech diverged substantially from adult pronunciations. The alignment boundaries were then manually corrected in places relevant to the present study. Word and pause/utterance boundaries were added, and whether vowels contained primary stress, the underlying form of the consonant (in the adult form), and whether it was realized as a flap or glottal stop were annotated.

Vowel durations were collected only for vowels that contained primary stress and were followed by a stop without an intervening word boundary, and further were not the final vowel in the utterance. Durations at ages starting with 1;10 were normalized to remove the confounding effect of intrinsic vowel duration. (Before this age there was little sign that the child's vowels had intrinsic durations in a manner similar to that of adults.) The

normalization consisted of multiplying the vowel durations by a constant to scale all vowels as if they were [i]'s. The constant specific for each vowel was computed from the Boston University Radio News corpus as the ratio of the mean duration of stressed [i] to the mean duration of stressed occurrences of the vowel. I found that this procedure reduced the variation in the data.

Vowels were grouped by whether they preceded underlyingly (in the adult form) voiced stops and those that preceded underlyingly unvoiced stops. The stops may have been articulated as a glottal stop or flap, but must have been realized to be counted. The affricates /tʃ,ʃ/ were included as voiced and voiceless stops, respectively.

### 5.2.2 Results

Mean vowel durations are reported in Figures 6 and 7. In the table,  $p$ -values are derived from a one-tailed t-test comparing the unvoiced and voiced durations. At all ages, a significant difference was found between the durations of the vowels before voiced and unvoiced stops (at worst  $p < .03$ ), with voiced/unvoiced ratios ranging from 1.3 to 2.1. Both pre-voiced and pre-unvoiced mean vowel durations are decreasing during this time. In the figure, error bars indicate the worst effect on the ratio of one standard-error change in both the mean voiced and unvoiced durations.

Age	N	Unvoiced	Voiced	Difference	Ratio	$p$ -value
1;6	50/42	241	354	113	1.5	.0005
1;10	19/6	263	442	179	1.7	.03
2;1	28/11	167	293	126	1.8	.0002
2;8	18/10	113	237	124	2.1	.0005
4;0	31/22	113	146	32	1.3	.002

Figure 6: Mean vowel durations (ms) before unvoiced and voiced stops at different ages for the child ‘Lily’. The N column gives the number of unvoiced followed by the number of voiced tokens.

I also tested whether speaking rate correlated with VLE ratio. I noted earlier that tempo is known to affect the magnitude of the VLE (Port, 1981), and that Lee et al. (1999) reported vowel duration decreasing with age. This correlation may have explained the change in VLE during acquisition in Krause (1982) and DiSimoni (1974), with the VLE changes there not due to acquisition of the VLE itself but these other considerations. Here, I measured mean speaking rate at each age as the mean vowel duration. There was no evident relation between speaking rate and VLE ratio. The age with the longest vowels had exactly the median VLE ratio.

The counts of several phones in Lily’s speech are given in figure 8 for several ages, in anticipation of the proposed dissertation work. Note that the ages given represent a fraction of the available data, as recordings were made at many other ages.

Additionally, I looked at the distribution of a sample of 28 /s/ and /z/ phonemes (in the adult form) from age 2;7–3;2. in terms of whether they were a part of a morphologically simplex content word, closed-class function word (including ‘is/-s’), a possessive or plural derivational affix, or a verbal inflectional affix. The counts are given in Figure 9.

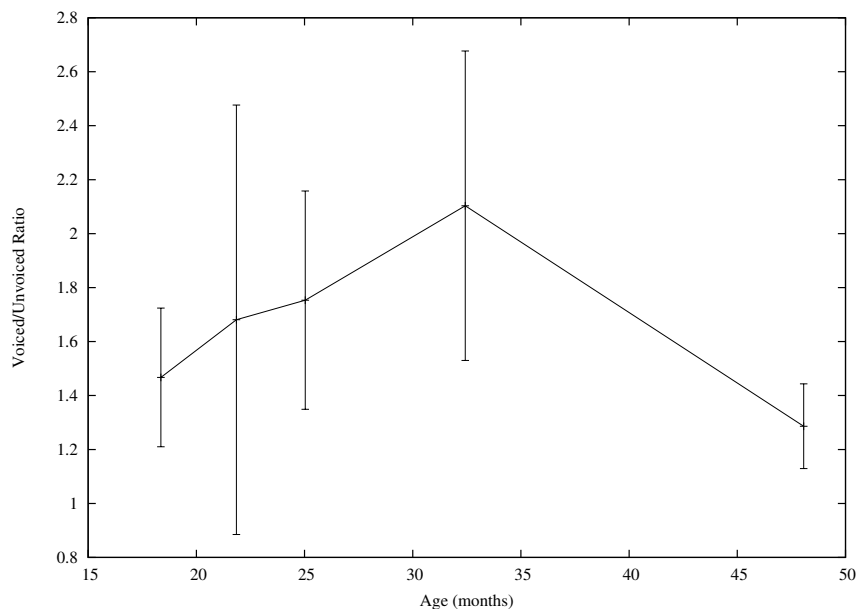


Figure 7: The voiced-unvoiced ratio at different ages for the child ‘Lily’, in utterance-medial context. Error bars indicate the effect on the ratio of one standard-error change in both the mean voiced and unvoiced durations.

Preliminary measurements of three correlates of voice in fricatives — preceding vowel duration, frication duration, and voicing duration during frication — from a sample of 45 tokens drawn from age 1;10–4;0 (data pooled for all ages) indicate that at least by age 4;0 Lily appears to be using all three correlates to indicate voice. No attempt was made to account for variability in speaking rate or focus, which we might expect in any case because of the different distributions of /s/ and /z/ indicated above. All three measures showed statistically significant differences between the tokens of the two phonemes, in the adult-like directions. Values are reported in Figure 10.

### 5.2.3 Discussion

The data reported here on the VLE largely agrees with the past work discussed in section 4.1, that there is a voiced-unvoiced difference going back to before age 2, and here as early as age 1;6. However, it is impossible to know from any work I know of whether the children are employing the VLE as a linguistic process. They may, instead, have phonetic detail stored in their lexicon, including target durations for segments as learned by copying the pattern from adult forms. If the VLE is a part of linguistic competence, its interaction with

Age	[t]	[d]	[r]	[s]	[z]
1;6	18	64	0	0	0
1;10	54	30	0	75	0
2;1	52	66	2	51	12
2;8	187	119	28	136	72
4;0	58	143	3	107	45

Figure 8: Counts of /t,d,r,s,z/ at various ages in Lily’s speech.

	N	Content	Function	Derivational	Inflectional
/s/	17	11	4	1	1
/z/	11	1	7	2	1
Total	28	12	11	3	2

Figure 9: Distribution of [s] and [z] phones age 2;7–3;2. Total count in the sample, and counts of those appearing in simple content and function words (including ‘is/its/s’, as possessive or plural derivational affixes, or as verbal inflection).

other linguistic and especially phonological processes, such as flapping, has also not been established.

Lily’s phone inventory also largely agrees with the past work discussed in section 4.2, indicating that further work on the realization of /s,z/ will be possible by studying productions around age 1;10–4;0.

Based on the preliminary measurements of the duration of the preceding vowel, frication, and voicing from a small sample across all ages, it would appear that there is enough data in the corpus to pursue this question in more depth and with a longitudinal design. The results from this sample show statistically significant differences in all three correlates, in the adult-like directions. But because the distributions of /s/ and /z/ between content and function words, the latter possibly often lacking any stress, were vastly different, one might object that comparing durations between /s/ and /z/ would not be reliable. However, for /s/’s the vowel duration was shorter but the frication duration was longer. An overall duration difference might explain the duration difference of one of the correlates, but not both.

## 6 The Proposal

I propose to carry out a corpus-based study, using the CHILDES corpora (continuing in part with the longitudinal portion used above, Demuth et al. 2006). The study will answer the five following questions.

*Do vowel length and other correlates of [voice], including voicing itself, trade-off in production in adults, in the same way that a correlation was found between tongue position and lip rounding in [+back] vowels (de Jong, 1995)?*

Before one can learn from acquisition data, one must have comparable adult data. The

	N	Vowel Duration (ms)	Frication Duration (ms)	Voicing Duration (%)
/s/	22	132	165	3
/z/	23	196	130	50
<i>p</i>		.03	.05	.0001

Figure 10: Correlates of [voice] for underlying (in adults) /s/ and /z/, age 1;10–4;0. Voicing duration is reported as a percent of frication duration. *p*-values are from a one-tailed t-test.

first component of this dissertation will be to flesh-out our knowledge of the interplay between voicing, segment duration, and airflow or frication amplitude in stops and /s,z/ in adults. Recall that de Jong (1995) found either a positive or negative correlation (depending on the speaker) from utterance to utterance between lip rounding and tongue position in the production of [+back] vowels, believed to be an interplay of articulations that is used to enhance or maintain the F<sub>2</sub> acoustic contrast of the segment. It was suggested that this may either relate to hypo/hyper-articulation or to the invocation of secondary phonetic features to fill in for other features that have been lenited. A similar interplay might be found between vowel duration, frication duration, frication intensity, and voicing. Although the correlates here do not affect a single acoustic cue such as F<sub>2</sub> above, we might nevertheless expect a production equivalent to cue trading. I will use corpus data, or if that fails then a production experiment, to measure whether these gestures are correlated amongst themselves within the two voice categories. The result will have implications for the interpretation of the acquisition data described below.

I intend to measure frication (high-frequency) energy as a stand-in for Smith’s (1997) measure of peak and mean airflow in [s,z]. A straight-forward check using adult corpus data will hopefully show that underlying /z/’s have lower high-frequency energy than underlying /s/’s, regardless of whether the phone is realized with voicing. To verify the relation between peak and mean airflow and the acoustic signal, a pneumotachographic mask can be acquired and a short experiment can be run to gather the required data. The measurement of high-frequency energy may be based on energy in a high band, or spectral tilt or center of gravity.

*What is the nature of /z/-devoicing in children: Does it mimic the adult pattern, and has it resulted in past research mistaking /z/ productions for [s]?*

Starting with recordings and transcriptions for children between age 2 and 4 from the corpus, I will measure the acoustic properties of the [s,z] phones and preceding vowel, including duration of frication, frication energy (measured as described above for adults), and preceding vowel duration. As Figure 8 indicates, this is the age in which sibilants are first used and when [z] phones are transcribed. The figure, which represents only a small subset of the recordings in the corpus, also shows that there ought to be sufficient tokens present for an analysis.

Amount of voicing is a somewhat difficult measure. For adult speech, Smith (1997) used an electroglottogram to record glottal movement directly to determine the start and stop time of periodicity during frication. In recordings from CHILDES corpus data, identifying the voice bar visually in a spectrogram is not always clear, particularly because of background noise and poor sound quality. I also desire to automate this process, which is

particularly useful when annotation standards need to be changed. For this I propose to develop or adapt a method of measuring voicing, looking first at the voice detection literature — possibly tailoring an approach to the data at hand, knowing something about the pitch range of the children and ambient noise. One possibility is to operationalize voicing as the energy in the low-frequency (< 600Hz) components of the segmented phone, relative to the mean energy in the same band in nearby sonorants in order to account for varying sound levels from recording to recording. This will be aided by the automatic forced alignment system I have in place for this corpus data, using the HMM and HTK-based Phonetics Lab’s forced alignment toolkit (Yuan and Liberman).

The [s,z] phones will be split according to whether they correspond with adult /s/ or /z/. Only morphologically simplex words will be considered, because the underlying [voice] value for the possessive and plural morphemes is not evident, since overtly they agree in voice with the preceding segment. The four measures above will be compared across the two /s/ and /z/ categories. This will tell us first whether the child has acquired the correct lexical entries, at least on the whole, if there is a statistically significant difference in the right direction for the four measures. For /z/, friction duration and energy should be less, and amount of voicing and preceding vowel duration should be greater. If this is the case, we can test whether (or when) the child has also acquired the rules to set the [voice] feature on the possessive and plural morphemes appropriately. If the child has acquired those rules, then those phones can be included in further analysis and treated as underlyingly (that is surface-phonologically) /s/ or /z/ as appropriate.

Then I will investigate whether the underlying-/z/ phones show the pattern of adult devoicing, that is that they differ from underlying-/s/ phones in friction energy, voicing, and preceding vowel duration, no matter whether there is complete, partial, or no actual voicing present, and whether the effects of utterance, syllable, and segmental context are similar to the effects in adults described in section 3 (Smith, 1997). We will then know whether children have a distinct production of /z/ before annotators have indicated the presence of [z] in children’s inventories and whether adult-like /z/-devoicing occurs in children.

*What is the time course of the acquisition of the production of the individual articulatory correlates of, i.e. the secondary phonetic features or phonetic implementation of, phonological [voice]?*

As discussed earlier, the VLE is seen preceding stops as early as age 1;6. While the presence of a voiced [z] was claimed to not occur until at least age 2;0, I suspect this age underestimates development of /z/. Because [z]’s, in adult speech, are commonly devoiced, we may find that /z/ is consistently realized with other cues much earlier. I will look at the time course of development of all of the cues mentioned above through the longitudinal corpus data.

*Do the correlates of [voice] trade-off in production in children, both across utterances in a single session and longitudinally across sessions? As the articulatory gestures that make up a phone are acquired in production, does the child adjust his reliance on each of the gestures?*

The data collective above will indicate whether the gestures of [voice] are correlated within the two voice categories in children or through time. That is, I will answer whether

there is an interplay between the gestures. Identifying tokens as either [+voice] or [-voice], however, cannot rely merely on the adult underlying representation. A child's having an incorrect underlying form, e.g. reversing /s/ and /z/, will introduce a correlation among the gestures if those gestures have already been learned to be associated with [ $\pm$ voice]. A more detailed analysis will be necessary. When considering /z/ tokens, only those tokens that show some /z/-like gesture (longer vocalic duration, shorter frication duration, etc.) can be considered, providing evidence that the child has the right underlying form for that token. Depending on the robustness of the data, the tokens may first need to be classified as having the correct underlying form according to the presence of some correct gesture before carrying out a correlation analysis on those tokens. Alternatively, a multi-dimensional factor analysis may be used on the data in whole to separate the dimensions of [ $\pm$ voice] and variation within each voice category.

From assessing the longitudinal development, I hope to test the hypothesis above, that the use of each of the correlates will be adjusted as the child masters the production of other correlates. Since it appears that the VLE is in use before voicing, we expect that the VLE magnitude will decrease over time once voicing starts to be used — that is, there should be a negative correlation between the magnitudes of the gestures over time. Since the VLE magnitude may be changing over time for independent reasons — changes in speaking rate, for instance — we will need to compare the VLE magnitude between stops and fricatives. Only a change in fricatives but not in stops could be attributed to the development of other cues. And because word-final stops have no other cues besides the VLE that I expect the child to need to learn, I do not believe there could be any developmental trends associated with the VLE specifically for word-final stops.

*The vowel duration and glottal voicing cues are observed in young children, but what can we learn about their linguistic status, as either e.g. deriving from stored phonetic detail in the lexicon or from being active linguistic knowledge about [voice]?*

I expect that by late ages the child will match the adult pattern of correlation among the [voice] gestures within [voice] categories — whether the gestures are correlated or not in adults is to be determined by the first part of this dissertation, and in children by a later part of the dissertation. But at early ages I may find a mismatch. If the gestures are uncorrelated in adults but are correlated in children, then the child must have some linguistic knowledge about the use of these gestures. The correlation could not be explained on either articulatory grounds or in terms of an exemplar model, since without an adult correlation there would be no correlation in the child's stored exemplars either. If adults have a correlation but children at some stage do not, then another conclusion can be reached. Because children's stored phonetic detail, e.g. exemplars, would include the correlations found in adult speech, child utterances without correlation are thus deviations from the exemplars. This suggests either that the child lacks the motor control to reproduce the exemplars, or that the child has linguistic competence of these factors, although perhaps not their association to [voice], so that they can control them independently without regard to stored phonetic detail.

A correlation between the use of these gestures over time also suggests that the correlates are active linguistic knowledge. If the VLE decreases as the use of voicing or frication duration increases to separate /s/ and /z/ in production, something again not attributable either to articulatory constraints or what the child hears in his environment, it would suggest

that the child is modulating individual gestures as separate variables related to [voice], rather than merely repeating stored phonetic detail.

The dissertation will involve the following tasks:

- Using existing adult corpus data or by means of a production experiment, comparing the magnitudes of the correlates of [voice] between fully, partially, and unvoiced /z/, looking for a correlation (i.e. something like cue trading) between the correlates within the /z/ category. This comparison was crucially omitted from Smith (1997). This task will take several weeks.
- Developing procedures to automatically measure the frication energy and presence and duration of voicing during frication. Previous work on voice detection will be relevant. An adult speech corpus and possibly new natural stimuli will be used to judge the success of the procedure developed. It will also be necessary to check whether the mean and peak airflow difference between /s/ and /z/ observed by Smith (1997) using a pneumotachographic mask can be detected from analysis of an acoustic signal. A procedure will be developed for measuring the intensity of frication to see if it can serve as a substitute for airflow in distinguishing /s/ and /z/. This process will take up to one month.
- Completing the forced phone alignment of the relevant portions of the CHILDES Providence corpus for child data, and manually correcting the alignments of the stops and pre-stop vowels. The phones [t,d,s,z] will be annotated as /t,d,s,z/ in the adult form. The vowels will be annotated as carrying or not carrying primary stress. The /s,z/ segments will be annotated as occurring within a morphologically simplex word or as an affix. Words will be annotated as being content and function words, and affixes will be annotated as derivational (possessive or plural) or inflectional (verbal ‘-s’). Word boundaries will be added. This process will take approximately one to two months.
- Analyzing the child data for correlations between the gestures and longitudinal trends, which will take up to one month.
- Writing the dissertation.

## References

- Baran, Jane, and Harry N. Seymour. 1976. The influence of three phonological rules of Black English in the discrimination of minimal word pairs. *Journal of Speech and Hearing Research* 19:467–474.
- van Bergem, Dick R. 1993. Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication* 12:1–23.
- Buder, Eugene, and Carol Stoel-Gammon. 2002. American and Swedish children’s acquisition of vowel duration: Effects of vowel identity and final stop voicing. *J. Acoust. Soc. Am.* 111:1854–1864.
- Chen, Matthew. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22:129–159.

- Chomsky, N., and M. Halle. 1968. *The sound pattern of english*. New York, NY: Harper & Row.
- Demuth, K., J. Culbertson, and J. Alter. 2006. Word-minimality, epenthesis, and coda licensing in the acquisition of English. *Language & Speech* 49:137–174.
- Denes, P. 1955. Effect of duration on the perception of voicing. *J. Acoust. Soc. Am.* 27:761–764.
- Dietrich, Christiane. 2006. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation. Doctoral Dissertation, Max Planck Institute for Psycholinguistics.
- Dietrich, Christiane, Daniel Swingley, and Janet F. Werker. 2007. Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences* 104:16027–16031.
- DiSimoni, Frank G. 1974. Influence of consonant environment on duration of vowels in the speech of three-, six-, and nine-year-old children. *J. Acoust. Soc. Am* 55:362–363.
- van der Feest, Suzanne V.H. 2007. Building a phonological lexicon: The acquisition of the Dutch voicing contrast in perception and production. Doctoral Dissertation, Radboud Universiteit Nijmegen.
- Flege, J.E., and R. Port. 1981. Cross-language phonetic interference: Arabic to English. *Language and Speech* 24:125–146.
- Fox, Robert, and Dale Terbeek. 1977. Dental flaps, vowel duration, and rule ordering in American English. *Journal of Phonetics* 5:27–34.
- Gandour, J., B. Weinberg, and D. Rutkowski. 1980. Influence of postvocalic consonants on vowel duration in esophageal speech. *Language Speech* 23:149–158.
- House, Arthur S., and Grant Fairbanks. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *J. Acoust. Soc. Am.* 25:105–113.
- Huff, C. 1980. Voicing and flap neutralization in New York City. *Research in Phonetics* 1:233–256.
- Jansen, Wouter. 2004. Laryngeal contrast and phonetic voicing: a laboratory phonology approach to English, Hungarian, and Dutch. Doctoral Dissertation, Rijksuniversiteit Groningen.
- Jansen, Wouter. 2007. Phonological ‘voicing’, phonetic voicing, and assimilation in English. *Language Sciences* 29:270–293.
- de Jong, Kenneth. 1995. On the status of redundant features: the case of backing and rounding in American English. In *Phonology and phonetic evidence*, ed. Bruce Connell and Amalia Arvaniti, 68–86. Cambridge University Press.
- Joos, Martin. 1942. A phonological dilemma in Canadian English. *Language* 18:141–144.
- Keating, P. A. 1984. Phonetic and phonological representation of stop consonant voicing. *Language* 60:286–319.
- Keating, P. A. 1985. Universal phonetics and the organization of grammars. In *Phonetic linguistics*, ed. V. Fromkin, 115–132. Academic Press.
- Klatt, Dennis H. 1973. Interaction between two factors that influence vowel duration. *J. Acoust. Soc. Am.* 54:1102–1104.
- Klatt, Dennis H. 1976. Linguistic uses of segmental duration in english: Acoustic and perceptual evidence. *J. Acoust. Soc. Am.* 59:1208–1221.

- Kluender, Keith R., Randy L. Diehl, and Beverly A. Wright. 1988. Vowel-length differences before voiced and voiceless consonants: an auditory explanation. *Journal of Phonetics* 16:153–169.
- Ko, Eon-Suk. 2007. Acquisition of vowel duration in children speaking American English. In *Proceedings of Interspeech 2007*. Antwerp, Belgium.
- Krause, Sue Ellen. 1982. Developmental use of vowel duration as a cue to postvocalic stop consonant voicing. *Journal of Speech and Hearing Research* 25:388–393.
- Laeufer, Christiane. 1992. Patterns of voicing-conditioned vowel duration in French and English. *Journal of Phonetics* 20:411–440.
- Lee, S., A. Potamianos, and S. Narayanan. 1999. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *J. Acoust. Soc. Am.* 105:1455–1468.
- Luce, Paul A., and Jan Charles-Luce. 1985. Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *J. Acoust. Soc. Am.* 78:1949–1957.
- Port, Robert F. 1981. Linguistic timing factors in combination. *J. Acoust. Soc. Am.* 69:262–274.
- Port, Robert F. 1996. The discreteness of phonetic elements and formal linguistics: A response to A. Manaster Ramer. *Journal of Phonetics* 24:491–511.
- Port, Robert F., and Michael L. O'Dell. 1986. Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13:455–471.
- Rosen, Kristin M. 2005. Analysis of speech segment duration with the lognormal distribution: A basis for unification and comparison. *Journal of Phonetics* 33:411–426.
- Scobbie, James M., Alice E. Turk, and Nigel Hewlett. 1999. Morphemes, phonetics and lexical items: The case of the Scottish Vowel Length Rule. In *Proceedings of the XIVth International Congress of Phonetic Sciences*, 1617–1220. San Francisco.
- Sharf, D.J. 1964. Vowel duration in whispered and normal speech. *Language Speech* 7:89–97.
- Smith, Caroline L. 1997. The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics* 25:471–500.
- Snow, David. 1997. Children's acquisition of speech timing in English: a comparative study of voice onset time and final syllable vowel lengthening. *J. Child Lang.* 24:35–56.
- Stevens, Kenneth N., Samuel Jay Keyser, and Haruko Kawasaki. 1986. Toward a phonetic and phonological theory of redundant features. In *Invariance and variability in speech processes*, ed. Joseph S. Perkell and Dennis H. Klatt, 426–463. Lawrence Erlbaum Associates.
- Stoel-Gammon, Carol. 1985. Phonetic inventories, 15-24 months: A longitudinal study. *Journal of Speech and Hearing Research* 28:505–512.
- Umeda, Noriko. 1975. Vowel duration in American English. *J. Acoust. Soc. Am.* 58:434–445.
- Weismer, Gary, Daniel Dinnsen, and Mary Elbert. 1981. A study of the voicing distinction associated with omitted, word-final stops. *Journal of Speech and Hearing Disorders* 46:320–328.

## A Word list used in experiment A

Below are the words used in the reading lists of experiment A.

Monosyllabic real words:

spite	spied
chat	chad
thought	thawed
bout	bowed
leak	league
sack	sag
pick	pig
peck	peg
buck	bug
tap	tab
ape	Abe
cup	cub
rope	robe

Tautosyllabic real words:

neatness	needless
seatbelt	seedling
fraction	fragment
doctor	dogma
pectin	pegboard
crapshoot	crabmeat
optics	object

Heterosyllabic real words:

seater	cedar
catty	caddy
petal	pedal
coating	coding
vicar	vigor
backing	bagging
chucking	chugging
flocking	flogging
hokey	hoagie
sopping	sobbing
seaport	seabed
staples	stables
soapy	sober

Monosyllabic nonsense words:

geet	geed
jite	jide
zat	zad
fot	fod
spote	spode
jeek	jeeg
chack	chag
skik	skig
nuck	nug
pauk	paug
spap	spab
skop	skob
peip	peib
gup	gub
foup	foub

## Tautosyllabic nonsense words:

geetmonk	geedmonk
jitehood	jidehood
zatback	zadback
fotful	fodful
spotestick	spodestick
jeekson	jeegson
chackpack	chagpack
skikmount	skigmount
nuckbon	nugbon
pauktill	paugtill
spapton	spabton
skoptrie	skobtrie
peipcat	peibcat
gupsnow	gubsnow
foupdram	foubdram

## Heterosyllabic nonsense words:

geety	geedy
jiteing	jideing
zattting	zadding
fotins	fodins
spowtuck	spowduck
jeeker	jeeger
chackal	chagal
ziken	zigen
nuckist	nugist
paukam	paugam
spapale	spabale
skoping	skobing
peiper	peiber
gupomt	gubomt
foupest	foubest