12/8/07

The phonetics of schwa vowels

Edward Flemming
Department of Linguistics & Philosophy, MIT

1. Introduction

Schwa is often characterized as a weak or reduced vowel. This is based on a number of generalizations about the cross-linguistic behavior of schwa: Schwa is the outcome of neutralization of vowel quality contrasts in a number of languages including English (e.g. Chomsky & Halle 1968:110ff.), Dutch (Booij 1995) and Southern Italian dialects (Maiden 1995). It is also commonly restricted to unstressed syllables due to vowel reduction and/or resistance to being stressed, e.g. English, Dutch, Indonesian (Cohn 1989). Schwa is often singled out by deletion processes, e.g. in Dutch (Booij 1995), English (Hooper 1978), French (Dell 1973), Hindi (Ohala 1983).

The basis for the weakness of schwa has been the subject of much research by phonologists (cf. Van Oostendorp 2000 for a recent proposal), but much less attention has been devoted to the question of what the phonetic characteristics of schwa vowels are. As observed by Lass (this volume), the answer is far from clear. Schwa is often taken to be a mid central vowel, in accordance with the denotation of the schwa symbol [ə] in the International Phonetic Alphabet. On the other hand, it is frequently observed that the quality of schwa in languages like English and Dutch varies substantially across contexts. The nature of schwa has implications for the analysis of the phonological patterns mentioned above. For example, the assumption that schwa is a mid central vowel has lead to the notion that vowel reduction involves approximation to the center of the vowel space. The observation that schwa is contextually variable is more consistent with an alternative characterization of vowel reduction as assimilation of a vowel to its segmental context (e.g. Lindblom 1963). According to this line of analysis, schwa is a vowel that lacks a well-defined target, and so assimilates strongly to surrounding segments, resulting in substantial variation in the vowel quality of schwa.

The goal of this chapter is to clarify the nature of schwa vowels, primarily through phonetic studies of English schwa, but drawing on evidence from other languages where it is available. We will see that both kinds of schwa vowels exist: true mid central vowels and contextually variable vowels. In fact both kinds of schwa vowel are found in English. But the two kinds of schwa appear to differ in their phonological patterning: mid central schwa usually minimally contrasts with higher vowel qualities (e.g. [i, u]), whereas variable schwa occurs primarily in contexts where all vowel quality contrasts can be neutralized. We will also briefly consider the implications of the existence of two kinds of schwa vowels for the analysis of vowel reduction. Although both mid central and variable schwa can arise via vowel reduction in unstressed syllables, both result from assimilation to context, so there is no support for the notion of vowel reduction as approximation to a mid central quality. The different outcomes represent different degrees of assimilation to context. Moderate reduction results in raising of low vowels towards mid central schwa as a result of assimilation to the narrow constrictions of adjacent consonants, while more extreme reduction results in a vowel quality that is strongly assimilated to its context, and is therefore contextually variable. This is variable schwa.

We will begin by illustrating the difference between mid central and variable schwa vowels in English, exploring the phonetic properties of variable schwa in some detail. Then we will turn to the broader implications of recognizing two kinds of schwa.

2. Two kinds of schwa in English

Flemming & Johnson (2007) found that there are significant phonetic differences between schwa vowels in word-final position, as in *china* or *comma*, and schwa vowels in other positions, as in *suppose* or *probable*. Word-final schwa vowels have a relatively consistent vowel quality, usually mid central, while word-internal schwa is relatively high and varies contextually in backness and lip position. This is illustrated in figure 1, based on data from that study. The plot shows the first two formant frequencies of word-final schwa vowels from words like *Rosa, sofa, umbrella,* and word-medial schwa vowels from two and three syllable words like *suggest, today, probable,* as produced by nine female speakers of American varieties of English. The mean formant values of full, primary stressed vowels produced by the same speakers are also plotted to provide a frame of reference.

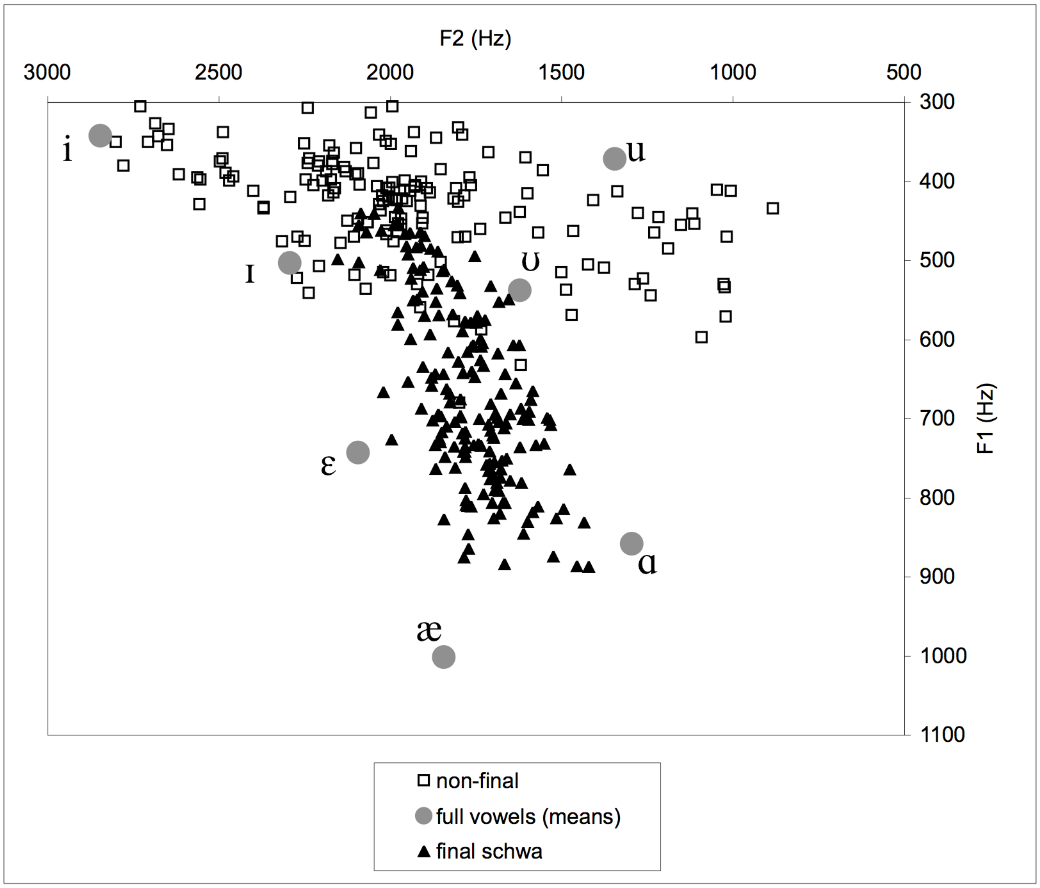

Fig. 1. Formant frequencies of tokens of final schwa (filled triangles) and non-final schwa (open squares), and the mean formant frequencies of the full vowels (gray circles). Data from nine female speakers of American English.

The mean F1 of word-final schwa vowels is 665 Hz, while mean F2 is 1772 Hz, which corresponds to a mid central vowel, IPA [ə]. The plot in figure 1 shows wide variation around this mean F1, but much of this is between speaker variation. This is shown in figure 2 where the mean F1 of final schwa is plotted for each speaker, together with mean F1 of the highest vowel, [i] (from the word *heed*), and the lowest vowel, [æ] (from the word *had*) to a give a sense of each speaker's formant range. It can be seen that F1 of final schwa is correlated with F1 of [æ], suggesting that much of the spread in F1 observed in figure 1 is due to differences in the vocal tract sizes of the subjects - larger vocal tracts yield lower formants overall. However, it can also be seen that subjects vary in the location of final schwa relative to the high and low reference vowels, indicating individual differences in vowel quality. For example, the second subject from the left has F1 of final schwa much closer to F1 of low [æ] than the other subjects. This variation is audible, and covers a range of central vowels from higher mid [ɘ] to lower [ɐ].

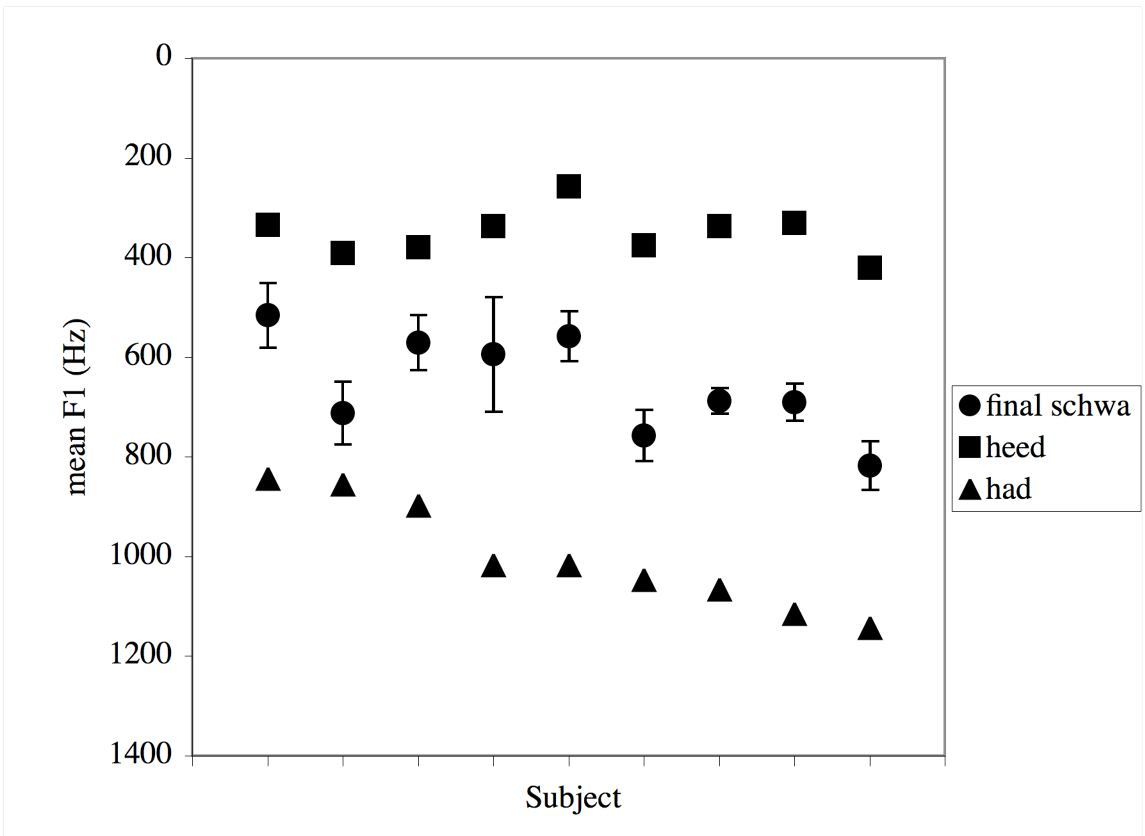

Fig. 2. Mean F1frequencies of word final schwa (circles), [i] of *heed* (squares), and [æ] of *had* (triangles), plotted for each subject. The final schwa means are marked with 1 standard deviation error bars.

The medial schwas, on the other hand, are high vowels – the average first formant frequency of 428 Hz lies between the tense and lax high vowels – with wide variation in F2, ranging from high values close to front [i], down to F2 values lower than those found in the [u] of *who*. Flemming & Johnson argue that this variability results from assimilation of schwa to the surrounding context. For example, the schwas with lowest F2 are found in *probable* [pɹɑbəbɫ̩]. The tongue body is probably back during this schwa due to assimilation to preceding back [ɑ] and the following velarized lateral, and the lips are likely to be constricted due to partial assimilation to the labial closures of the preceding and following stops. Both a back tongue body and labial constriction serve to lower F2. Higher F2 values are found in schwa adjacent to palato-alveolars which have a relatively front tongue body position, e.g. *suggest* [sədʒɛst], *prejudice* [pɹɛdʒədɪs].

However, to establish that the variability of non-final schwa is due to assimilation to context, it is necessary to examine the effects of context in a more controlled and systematic fashion. Kondo (1994) investigated schwa variation in an experimental study of the schwa vowel of the English indefinite article *a*, in phrases like *pick a kitten*, using speakers of RP English. She systematically varied the consonants and vowels preceding and following schwa and measured the resulting variation in the formant frequencies of schwa. She found that variation in schwa F2 was largely predictable from context in ways that were consistent with the assimilation hypothesis. The next section reports an experiment to investigate word-internal schwa vowels using a similar methodology, but considering a wider range of contexts.

3. Variable schwa in English

To study contextual variability of word-medial schwa, we need to examine the realization of schwa in a wide range of contexts. To this end, the experimental materials consisted of nonce words of the form [b V_1C_1əC_2V_2t], where V_1 was one of [i, æ, u], C_1 and C_2 were each one of the stops [b, d, g], and V_2 was one of the vowels [i, A, u], resulting in a total of 81 words. Subjects were instructed to produce the words with same stress pattern as the words *propagate* and *parakeet*. So the resulting words were [ˈbudəˌgit], [ˈbæbəˌdut], etc. The words were read in the frame sentence ‘X. Do you know what an X is?” by four native speakers of American English, two male and two female. The repetition of the target word in the carrier phrase was intended to facilitate fluent production of the nonce words, so only the second ‘rehearsed’ rendition was analyzed.

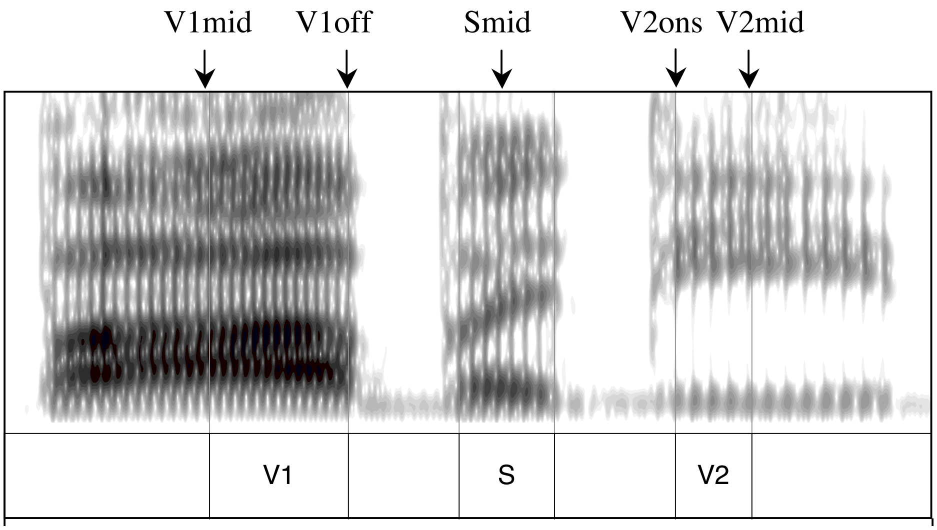

Fig. 3. Spectrogram of an utterance of [ˈbæbəˌdit], illustrating the points at which formant measurements were made.

F1 and F2 were measured at the mid point of the voiced portion of schwa (Smid), at the steady states of V_1 and V_2 (or the temporal mid point if there was no steady state) (V1mid, V2mid), at the offset of V_1 (V1off) and at the onset of voicing in V_2 (V2ons). These measurement points are labeled on a spectrogram of the word [ˈbæbəˌdit] in figure 3. Figure 4 shows scatterplots of the formant frequencies from the middle of the schwa vowels taken from all contexts. Only two subjects are shown, but the other two subjects showed very similar patterns. As in figure 1, we observe that medial schwa is highly variable in quality, particularly in F2 where it covers the full range from the front to the back of the vowel space. But the question we wish to address now is whether this variability results from assimilation to context.

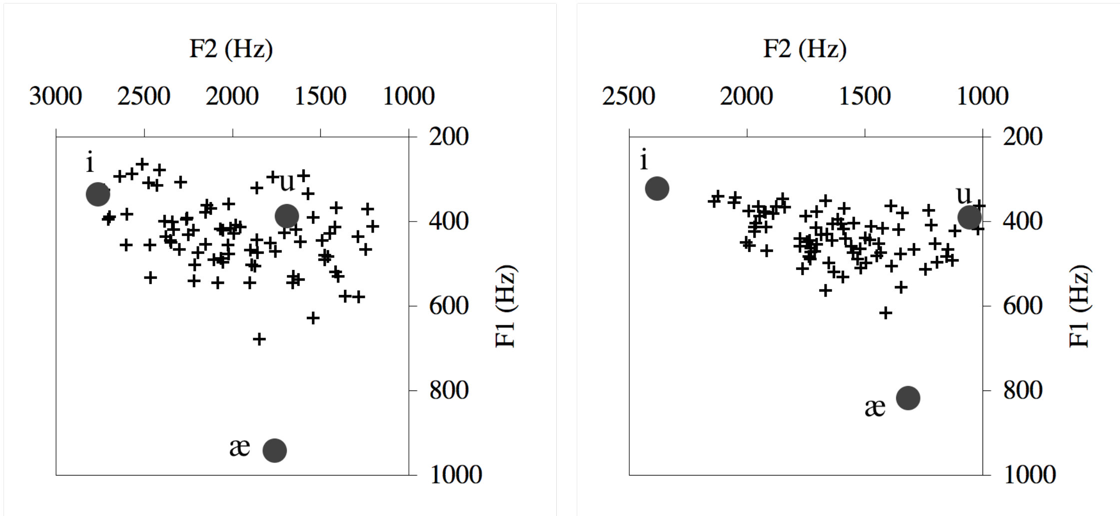

Fig. 4. Formant frequencies of schwa vowels from all contexts, two speakers.

The adjacent vowels and consonants all have substantial effects on the realization of the schwa vowel. This is illustrated by figure 5 which shows mean F2 values at all five measurement points for three words, [bigəgit], [bugəgɑt], and [bibəbit]. The effect of vowel context on F2 at the midpoint of schwa (Smid) is apparent from a comparison of [bigəgit] and [bugəgɑt]. Both have the same consonants, but the first has front vowels preceding and following schwa whereas the second has back vowels. Schwa F2 is high in the context of front vowels, which also have high F2, and much lower in the context of back vowels, which have low F2. In other words, schwa is assimilating to the surrounding vowels. The comparison between [bigəgit] and [bibəbit] illustrates how large the effect of consonants can be. These words have the same vowels, but differ in the consonants that precede and follow schwa. F2 in schwa is much lower when the surrounding consonants are labial than when they are velar. Again, this is plausibly an assimilatory effect: the lips are not opening fully during schwa in [bibəbit] due to partial assimilation to the preceding and following lip closures. Lip constriction lowers all formants, including F2, resulting in a much lower F2 than in [bigəgit].

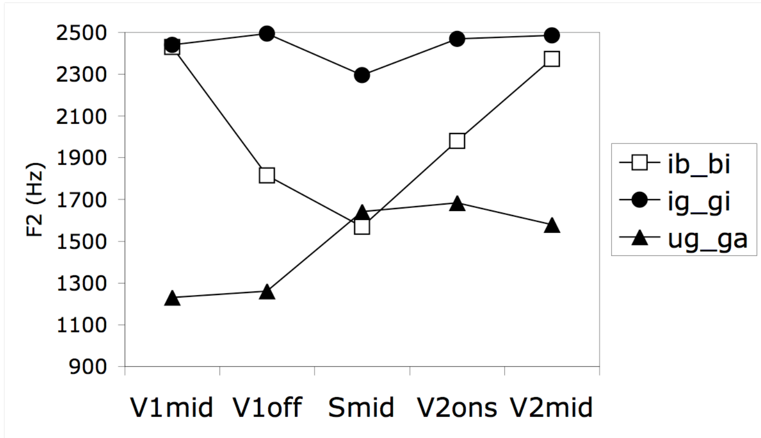

Fig. 5. Mean F2 at five time points in the words [bibəbit] (diamonds), [bidədit] (triangles), and [bigəgit] (circles).

In fact, most of the variation in schwa F2 can be attributed to assimilation of schwa to its context. We show this by constructing a model of schwa assimilation that can predict most of the observed variation in schwa F2. The basic idea behind the model is that if schwa is maximally assimilated to its context then it should be realized as a smooth movement from the articulatory position for the preceding consonant to the position for the following consonant (cf. Browman & Goldstein 1992, van Bergem 1995). The vocal tract must also be opened sufficiently to realize a vocalic sound, but in other respects such a vowel would essentially be the result of interpolation between preceding and following context. Accordingly the middle of schwa, where our measurements were taken, should occupy a position that is intermediate between these two endpoints.

However, consonants and vowels are coarticulated, so the positions of the articulators during the consonants also depends on the adjacent vowel. For example, in a sequence like [bi] the lips must be closed for the labial stop, but the tongue body anticipates the position of the following vowel (e.g. Löfqvist & Gracco 1999). The effects of this coarticulation on the F2 frequency adjacent to a stop are typically very predictable: F2 adjacent to a stop is a linear function of F2 at the center of the adjacent vowel (Klatt 1987). That is, the relationship between $F2_C$, the frequency of F2 adjacent to the stop closure, and $F2_V$, the frequency of F2 at the center of the adjacent vowel can be expressed as a linear equation of the form shown in (1), where slope *a* and intercept *b* depend on the consonant. If there were no coarticulatory effect of the vowel on F2 adjacent to the consonant, i.e. $F2_C$ had the same value regardless of the vowel context, then a would be 0. At the other extreme, if a is close to 1, then the consonant is strongly coarticulated with the vowel, so any change in vowel F2 is closely tracked by $F2_C$. The locus equations of labial and velar stops have been found to have steep slopes ($a \approx 0.7$-1), while coronal stops have shallower slopes ($a \approx 0.3$-0.5) (Sussman et al 1993).

(1) $F2_C = aF2_V + b$

So the model of F2 in schwa is built on the idea that F2 adjacent to C_1 and C_2 should depend on the place of articulation of the consonant and F2 in the adjacent vowel, as in (1). The F2 during schwa should then be a smooth transition between these two values, so F2 in the middle of schwa should be a weighted average of these endpoints. This model can be formulated as in (2), where $F2_{Smid}$ is the frequency of F2 at Smid (the middle of schwa), $F2_{V1}$ and $F2_{V2}$ are F2 at V1mid and V2mid respectively, a_{C1} and a_{C2} are the slopes associated with C1 and C2, and b_{C1} and b_{C2} are the intercept terms associated with C1 and C2 – i.e. there is one value of each of these coefficients for each of the three consonants, [b, d, g]. The weighting term α takes a value between 0 and 1, and represents the relative strength of the effects of the preceding and following contexts on schwa F2.

(2) $F2_{Smid} = \alpha(a_{C1}F2_{V1} + b_{C1}) + (1-\alpha)(a_{C2}F2_{V2} + b_{C2})$

However, the experiment does not manipulate the value of α, so we cannot distinguish its contribution from the contributions of the a and b coefficients. Accordingly, the model that was actually fitted to the data is as shown in (3). This is derived from (2) by setting $a'_{C1} = \alpha a_{C1}$, $b'_{C1} = \alpha b_{C1}$, etc.

(3) $F2_{Smid} = a'_{C1}F2_{V1} + b'_{C1} + a'_{C2}F2_{V2} + b'_{C2}$

This model was fitted to the data for each subject using Stata's 'anova' program (StataCorp 2007). The resulting models had r^2 ranging from 0.73 to 0.86. r^2 is the ratio of the variance accounted for by the model to the total variance in the data, so an r^2 of 0.86 means that the model accounts for 86% of the variation in schwa F2[1]. In other words, the wide range of schwa F2 values observed in scatter plots like figure 3 can largely be explained as assimilation of the schwa vowel to its context.

To illustrate how the model operates, the parameter values from the best fitting model for one subject are shown in the table in (4). To calculate the predicted schwa F2 for an utterance of [bæbəbut], we need the values of $F2_{V1}$ and $F2_{V2}$, which are 1512 Hz and 1099 Hz respectively. These values, and the values of the coefficients from table (3) are then substituted into the model equation in (3), as shown in (5), calculating a predicted F2 of 1297 Hz, which is close to the measured value of 1325 Hz.

(4)

	a'_{C1}	b'_{C1}	a'_{C2}	b'_{C2}
b	0.129	349	0.199	534
d	0.062	729	0.183	862
g	0.129	611	0.279	611

(5) $F2_{Smid} = 0.129 \times 1512 + 349 + 0.199 \times 1099 + 534 = 1297$ Hz

The height of schwa, as indicated by F1, also varies according to vowel context, as illustrated in figure 6. This figure shows mean F1 values in schwa and in the middles of V_1 and V_2 for three sets of words. It can be seen that F1 is higher in schwa when it is preceded and followed by vowels with high F1, (V_1 = [æ], V_2 = [ɑ]), i.e. schwa is lower when surrounded by low vowels. F1 is lower where the surrounding vowels have low F1 ($V_1 = V_2$ = [i]), and mixed vowels result in intermediate schwa F1. However the range of variation in F1 is much smaller than in F2: Schwa is mostly high, and is rarely lower than mid even

[1] It is not informative to fit a single model to the pooled data from all of the subjects, because much of the variance in the pooled data set is due to differences between subjects in their overall formant ranges.

when surrounded by low vowels. This is due to assimilation to the surrounding consonants. All the stops require complete closures of the vocal tract, and any constriction above the pharynx lowers F1, so if schwa assimilates to the surrounding stops in terms of constriction degree, the result is a vowel with low F1.

The influence of surrounding vowels on schwa F1 suggests that schwa assimilates to its context with respect to some articulators more than others. For example, lip constriction in [b] is usually a complete closure regardless of vowel context, but jaw height (Keating et al 1994) and tongue body position vary as a function of adjacent vowel height, so the influence of adjacent vowels is presumably mediated by the vowel-to-consonant coarticulation in these articulators.

A model similar to (3) accounts for 54%-72% of the variance in schwa F1. The lower r^2 of the F1 models may simply reflect the fact that there is less coarticulatory variation in F1, so the amount of variation due to noise (such as measurement error and random variation) constitutes a larger proportion of the total F1 variance. This interpretation is supported by the fact that the Root Mean Squared error of the predicted values is actually smaller for the F1 model than for the F2 model – i.e. the predicted F1 values are, on average, closer to their observed values than is the case for the predicted F2 values.

Overall these results are very similar to those reported by Kondo (1994) for the schwa of the indefinite article *a*. She found that schwa F2 varied substantially across contexts, and that this variation was largely predictable from the context. Schwa F1 varied even less than in the present experiment: all the schwa vowels were high (F1 values in the vicinity of 300 Hz).

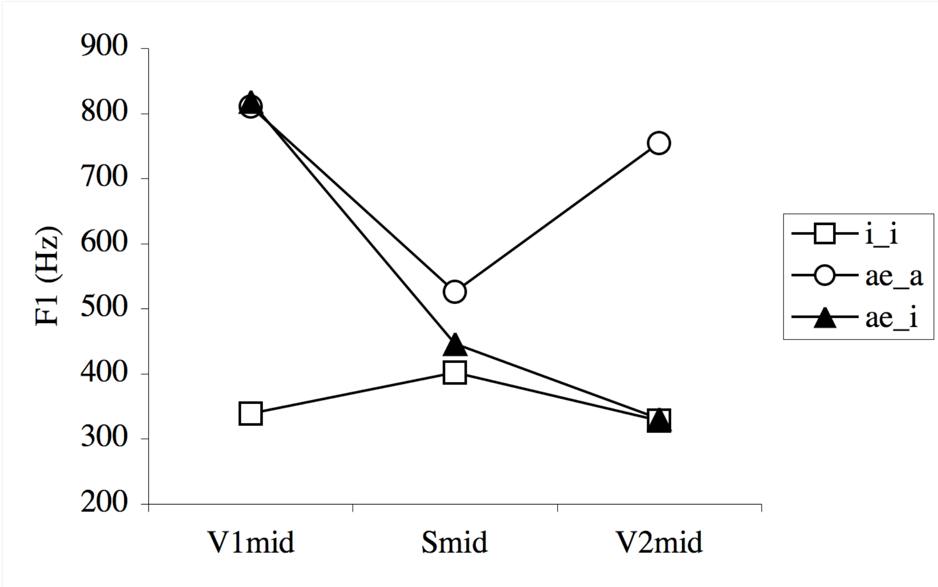

Fig. 6. Mean F1 at the midpoints of V1, schwa, and V2 in selected classes of words.

3.1 Why is schwa quality so variable?

We have seen that English non-final schwa vowels are very variable in quality, particularly F2, and that most of this variation can be explained in terms of assimilation to context. But we still need to understand why this kind of schwa is more variable than a full vowel. Flemming (2004) argues that two related factors are involved: word-medial schwa is (i) very short, and (ii) it does not minimally contrast with other vowel qualities. These two factors are related because the short duration of non-final unstressed syllables motivates the neutralization of vowel quality contrasts in these contexts.

The outline of the analysis is as follows: To realize a particular vowel quality in a word, it is necessary to move from the articulatory position of the previous segment to the target for the vowel and then on to the position for the following segment. As the duration of the vowel decreases, it can become difficult to complete the required movements, especially if the vowel target is far from the targets for the preceding or following segments, because the articulators would have to move too fast to complete the movements in the time available. Lindblom (1963) shows that speakers tend to fall short of vowel targets as vowel duration decreases: at shorter durations vowels assimilate more and more to their contexts. Lindblom dubs this phenomenon 'target undershoot'. As a result, in positions where vowel duration is very short, all vowels in that context are liable to be strongly assimilated to the surrounding segments, and thus similar to each other in quality. Given a requirement that contrasting sounds should be perceptually distinct (Liljencrants & Lindblom 1972, Flemming 2004), this makes short, unstressed syllables a poor location for vowel quality contrasts, and accordingly these contrasts are often neutralized.

So one reason why schwa is expected to be subject to stronger coarticulatory effects than other vowels is that it is shorter, and thus more subject to undershoot – i.e. assimilation to its context. The medial schwa vowels in Flemming & Johnson's (2007) study average 64 ms, while Kondo (1994) reports that schwa in the indefinite article averages 34 ms. By comparison, tense vowels can be as long as 300 ms in citation forms (Peterson & Lehiste 1960) and are on the order of 150 ms in fluent speech (van Santen 1992).

A second factor that is suggested to allow greater variability in schwa is that it generally occurs in contexts where it does not contrast with other vowel qualities, at least in American English[2]. The idea is that there is no motivation to resist the pressure to assimilate to context if there is no need to realize vowel quality contrasts. If vowel targets specify the realization of contrasts so, for example, the targets for [i] are the properties that it must have to differentiate it from contrasting vowels that could appear in the same context, such as [I], [u] etc, then in contexts where there are no vowel quality contrasts, vowels should

[2] As discussed in Flemming & Johnson (2007), there are derived contrasts between word-final and variable schwa when inflections are added to schwa-final words, e.g. adding a possessive suffix to schwa-final *Rosa*, creates Rosa's, which is a minimal pair with *roses* in most dialects. However, this contrast is not possible morpheme-internally. There are also other vowels that can appear in non-final unstressed syllables, but they only appear in specific segmental contexts where schwa is not permitted. For example, [i] and [oU] can appear in unstressed syllables before vowels, e.g. *Whittier, Ottawa* (for some speakers, e.g. Hayes 1995:14f.), but schwa cannot appear before another vowel. Some dialects may have contrasts between two unstressed, non-final vowels. RP English has minimal pairs like *Lennon* and *Lenin* (homophonous in most American English accents), where the second syllable appears to be completely unstressed in each case, but I do not know of any instrumental investigation of this contrast.

lack vowel quality targets. In the absence of a specific target, it is predicted that schwa should be realized with a minimum of articulatory effort, which would plausibly yield a smooth transition between the preceding and following sounds.

Browman & Goldstein (1992) investigated the idea that schwa variability results from lack of any intrinsic vowel quality target using articulatory data collected by X-ray microbeam. They examined articulatory records of utterances of the form [pV_1pəpV_2pə], where V_1 and V_2 were all pairs of vowels drawn from the set [i, ɛ, ɑ, ʌ, u] to see if the tongue body movement in the first schwa was the result of interpolation between the preceding and following full vowels. They concluded that schwa is not targetless, but has a weak mid central target. This was indicated most clearly by the observation that in words like [pipəpipə], the tongue body moved downwards during the schwa, whereas interpolation between the preceding and following high vowels would have yielded a steady, high tongue body position.

However it is not entirely clear whether Browman & Goldstein's materials elicited word-final schwa or word-medial schwa. Nonce words like [pipəpipə] could easily be read as compound words, making the schwa vowels effectively word-final, and thus expected to be mid central vowels given the evidence discussed in section 1. So it is worth revisiting this question in light of the present study, since the materials used here were expressly designed to elicit word-medial schwas. It turns out that very similar evidence against targetless schwa can be found here. Where schwa appears between [i] vowels, F1 in schwa is almost always higher than in the preceding and following vowels (cf. fig. 6). This is not an expected effect of assimilation to the vowels or the consonants, because both of those influences tend to lower F1. So this pattern appears to involve movements of schwa away from its context, indicating the existence of an F1 (height) target.

A refinement of the analysis of schwa variability outlined above may able to account for the observation that non-final schwa varies substantially according to context but is not completely targetless. The prediction that non-final schwa should be targetless was based on the hypothesis that the phonetic targets associated with a segment are related to the system of contrasts that it enters into: if there are no vowel quality contrasts there is no need for a specific vowel quality target. But vowels have manner specifications as well as place of articulation specifications, so even if variable schwa lacks a quality target, it is still a vowel, not a consonant, and realizing a vowel requires an open vocal tract. Opening the vocal tract raises F1, so the apparent F1 target could be a side effect of a vowel manner target. Note, however, that this target would have to specify a more open vocal tract than is found in [i], so it would be demanding more than the minimal aperture to produce a vowel.

A final possibility to consider is that assimilation to consonants may not uniformly lower schwa F1, as assumed above. This possibility is suggested by a phenomenon known as the 'trough effect': the tongue body lowers during a labial stop surrounded by high vowels (e.g. [ibi, ubu]). This is referred to as the trough effect because the tongue body is high during the vowels, but lowers slightly during the labial, forming a trough in the tongue height trajectory. This effect is well-established, although its cause is the subject of much dispute (e.g. Houde 1968, Lindblom et al 2002). Whatever its basis, schwa could assimilate to the lowered tongue body position of the labial, resulting in the apparent failure of schwa to assimilate fully to the height of surrounding [i] vowels.

The trough effect cannot fully explain the observation that schwa F1 is higher than F1 of surrounding vowels in [biCəCit] words since it is only known to apply to labials. But the general point is that the trough effect shows that it is not safe to assume that assimilation to an adjacent stop should always lower schwa F1. Moreover the discrepancy between schwa F1 and F1 of [i] is larger where one of the consonants is labial, i.e. where the trough effect could be present, and largest where both consonants are labial (a mean difference of 100 Hz). Conversely, the discrepancy is smallest between velars (a mean difference of 45 Hz). Velars could not show any kind of trough effect since they must have a complete tongue body closure.

It is also important to be clear that the analysis of F2 variation presented above does not show that schwa lacks an F2 target, all it shows is that the extensive variation in schwa F2 is substantially predictable from a model that assumes it results from assimilation to context. This is consistent with the variation being either due to lack of an F2 target or due to substantial deviation from an F2 target (cf. Barry 1998). To demonstrate that there is no need for a schwa target to account for the data, it would be necessary to construct a model of coarticulation that is applicable to contexts without a schwa vowel, and then show that the same model can account for schwa variability without positing a schwa target.

That is, Browman & Goldstein (1992) argue that schwa can only be regarded as targetless if its tongue body position can be predicted from a linear combination of the tongue body positions of the preceding and following vowels, without any constant term in the model (cf. Choi 1995). A constant term would represent a schwa target, because it would allow for deviation from an interpolative trajectory between the preceding and following vowels. The model in (2) does not meet this standard because it contains two constant terms, b_{C1} and b_{C2}. We have interpreted these constants as consonant targets rather than as a schwa target, but the schwa is present in every utterance, so it is possible that a schwa target is included in the consonant constants – i.e. we would arrive at different constants if we modeled consonant-vowel coarticulation for stops between full vowels. So it would be necessary to model a wider range of contexts to support the claim that these constants actually represent general consonant targets.

In summary, variable schwa assimilates to its context so extensively because it is a very short vowel. It may also be that this schwa is particularly susceptible to coarticulation because it is unimportant for it to have any particular vowel quality since it does not minimally contrast in vowel quality, but it is not clear that variable schwa is completely targetless.

4. The nature of vowel reduction

We have seen that English has two types of unstressed vowels that are commonly transcribed as schwa: a mid central schwa that appears in unstressed word-final syllables and a variable schwa that appears in word-internal unstressed syllables. Variable schwa assimilates to its context, resulting in substantial contextual variation in vowel quality. Given the existence of two quite distinct types of schwa vowel, we must be cautious in accepting generalizations about schwa vowels as a class until we really know what kinds of vowels are involved. For example, it is clear that the two kinds of schwa vowel pattern quite differently with respect to vowel reduction processes, a fact that is obscured by the practice of transcribing both in the same way.

Phonological vowel reduction involves the neutralization of vowel contrasts in unstressed syllables, as in English reduction to schwa. Both mid central and variable schwas arise through vowel reduction, but mid central schwa is generally the unstressed counterpart of a low vowel, and arises in a moderate form of vowel reduction that does not affect all vowel qualities, but leaves mid central schwa contrasting with higher vowels. For example, in Girona Catalan there are six vowels in stressed syllables [i, e, ɛ, a, o, u], but in unstressed syllables the vowel inventory is reduced to three [i, ə, u], where schwa is a mid central vowel, as shown in figure 7 (Herrick 2003). The vowels /e, ɛ, a/ are reduced to [ə] in unstressed syllables, while /o, u/ neutralize to [u]. Variable schwa results from a more extreme form of vowel reduction that applies to all vowel qualities, potentially neutralizing all vowel qualities, as in English.

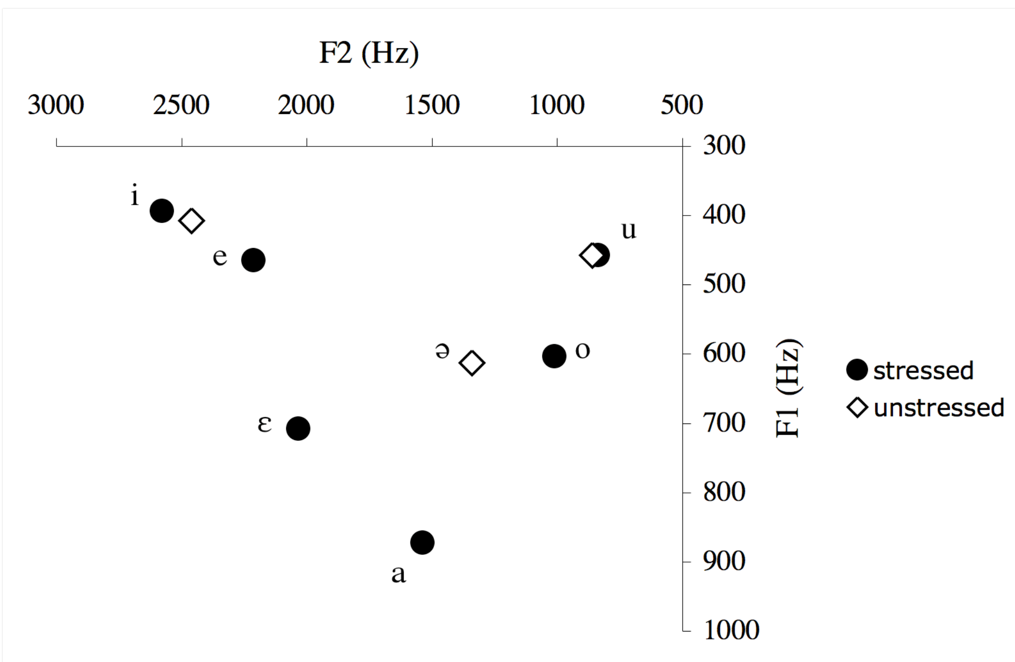

Fig. 7. Mean formant frequencies of Girona Catalan stressed and unstressed vowels. 3 speakers (data from Herrick 2003).

Other examples of moderate vowel reduction include Standard Russian (Padgett & Tabain 2003) and E. Bulgarian (Wood & Pettersson 1988). In both cases the reduced vowel inventory is [i, ə, u], i.e. mid central schwa contrasts with higher vowels[3]. Another example of reduction to variable schwa is found in Dutch (Booij 1995). In this case vowel reduction is optional, but like English vowel reduction it can apply to all vowel qualities, including high vowels. So it is more accurate to say that variable schwa occurs where the speaker is not concerned to realize any particular vowel quality, whether because of systematic or optional reduction. Studies by Koopmans-van Beinum (1994) and van Bergem (1994) show that Dutch schwa is very similar to English variable schwa: it is a high vowel in most contexts, although ranging down to mid, and is very variable in F2. Van Bergem shows that the F2 variation is due to assimilation of schwa to its context in an experiment similar to that described above.

As discussed in Flemming & Johnson (2007), the same correlation between the extent of vowel reduction and the nature of the schwa vowel involved extends to the two types of schwa in English. As already noted, English variable schwa results from neutralization of all vowel qualities. On the other hand, the mid central schwa found in word-final position contrasts with unstressed [i] and [oʊ] (e.g. Hayes 1995:14f.), as illustrated in (6). We can tell that the final vowels in these words are unstressed because they are preceded by flaps in American English and flapping only applies before unstressed vowels (or across a word boundary, as in *at ease*) (Kahn 1976).

(6)	ˈpɹ̠ɪɾi	*pretty*
	ˈbeɪɾə	*beta*
	ˈmɑɾoʊ	*motto*

Although moderate vowel reduction involves reduction to mid central schwa, this pattern does not support the conception of vowel reduction as a shift towards the center of the vowel space. Both kinds of vowel reduction can be analyzed in terms of assimilation to the context, the difference being in the extent of the assimilation (Flemming 2004). Reduction to variable schwa involves extensive assimilation to context in height, backness and lip position, as discussed in the preceding sections. But a lesser degree of assimilation to context

[3] In Russian reduction yields two vowels, [i, u], after palatalized consonants.

still results in raising of low vowels due to assimilation to the narrow constrictions of adjacent consonants. This often yields a mid central schwa vowel, as in the examples above. Lindblom (1963) shows that high vowels are relatively unaffected by undershoot until very short durations because it is not necessary to move so far to and from the adjacent consonant constrictions in order to realize a high vowel, so contrasts with higher vowels are generally preserved in these contexts.

One factor that shapes the extent of reduction is likely to be vowel duration, as discussed in section 3. This could be the basis for the split between word-final and non-final unstressed vowels in English. In Johnson & Flemming's (2007) study, the non-final schwa vowels averaged 64 ms in duration while the word-final schwa vowels had a mean duration of 153 ms. This probably overstates the usual margin of difference because the word-final schwas were subject to phrase-final lengthening, but it is likely that some duration difference is general, perhaps due to word-final lengthening. Greater vowel duration in final unstressed vowels means less undershoot, so it is possible to realize contrasts between mid central and higher vowels.

5. Coda

Lass (this volume) argues that his dialect of English contains at least seven different kinds of schwa – i.e. seven different vowel qualities that appear in unstressed syllables – and argues that the common practice of transcribing all of these vowels as [ə] is misleading. The evidence from the dialects studied here is broadly consistent with these claims, although I would argue that one of Lass's examples, the [ɔ] of *morality* bears secondary stress, preserved cyclically from the primary stress in *móral*. There is an alternative pronunciation of *morality* with a variable schwa in the first syllable (this is my own pronunciation), which results from elimination of the secondary stress.

The distinction that Lass notes between word-final and word-medial schwa vowels, describing an 'open mid central' vowel in word-final position and higher, context-specific unstressed vowels in word-medial syllables corresponds to the findings reported here. According to the data and analysis presented here, the latter are probably variable schwas, assimilated to their respective contexts. However, I would argue that it is a mistake to ask linguists to transcribe the realizations of variable schwa because they constitute a continuum of vowel qualities, as shown by plots like figures 1 and 4 above. It is of limited value to divide this continuum arbitrarily by assigning labels to regions of it. For the same reason the proposal that variable schwa should be identified with a variety of full vowel qualities is problematic: the fact that variable schwa sometimes resembles full vowel qualities is an inevitable side effect of the fact that it varies over such a large region of the vowel space. But equally, variable schwa frequently falls in between full vowel qualities, and these cases must be accounted for also.

What is needed is an analysis that derives the continuous acoustic and articulatory properties of variable schwa according to its context, and this paper takes some steps towards that goal. The analysis developed here shows that the contextual variability of non-final schwa need not imply that there is no underlying unity to it, contrary to Lass's proposed line of analysis. Variable schwa is not analyzed as a particular vowel quality, it is a very short vowel with at best a weakly specified vowel quality target. So in a sense what distinguishes variable schwa is a particular susceptibility to coarticulation rather than a particular vowel quality. The relative success in predicting the realizations of this vowel across a wide range of contexts supports the idea that there is a real sense in which it is a single category.

The mid-central schwa found in word-final unstressed syllables, on the other hand, is analyzed as a distinct category from variable schwa, and I concur with Lass that collapsing these two categories together under the label of schwa is a misleading. It also obscures phonological generalizations since the two types of schwa pattern quite differently in phonological processes such as vowel reduction.

References:

Barry, William (1998). Time as a factor in the acoustic variation of schwa. *The 5th International Conference on Spoken Language Processsing (ICSLP '98) vol. 5*. Sydney, Australia.

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

Browman, C., and L. Goldstein (1992). “Targetless” schwa: an articulatory analysis. D.R. Ladd and G. Docherty (eds.) *Papers in Laboratory Phonology II*. CUP. 26-67.

Choi, John (1995). An acoustic-phonetic underspecification account of Marshallese vowel allophony. *Journal of Phonetics* 23, 323-347.

Chomsky, Noam, and Morris Halle (1968). *The Sound Pattern of English*. New York: Harper and Row.

Cohn, Abigail. 1989. Stress in Indonesian and bracketing paradoxes. *Natural Language and Linguistic Theory* 7:167-216.

Flemming, E. (2004). Contrast and perceptual distinctiveness. Hayes, B., Kirchner, R. & Steriade, D. (eds.) *Phonetically-Based Phonology*, 232—276. Cambridge: Cambridge University Press.

Flemming, E., & S. Johnson (2007). Rosa’s roses: reduced vowels in American English. *Journal of the International Phonetic Association* 37, 83-96.

Fougeron, C., C. Gendrot, & A. Bürki (2007). On the acoustic characteristics of French schwa. *Proceedings of ICPhS XVI*, Saarbrücken, 941-944.

Herrick, Dylan (2003), An Acoustic Analysis of Phonological Vowel Reduction in Six Varieties of Catalan. PhD dissertation, University of California, Santa Cruz.

Hooper, Joan (1978) Constraints on schwa-deletion in American English; in Fisiak, Recent developments in historical linguisitics; 4th ed., pp. 183–207. Mouton, The Hague.

Houde, R. (1968) A study of tongue body motion during selected consonant sounds. *SCRL Monograph* 2. Speech Communications Research Laboratory, Santa Barbara.

Kahn, D. (1976). *Syllable-Based Generalizations in English Phonology*. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Keating, P.A., B. Lindblom, J. Lubker, & J. Kreiman (1994). Variability in jaw height for segments in English and Swedish VCVs. *Journal of Phonetics* 22, 407-422.

Klatt, Dennis H. (1987). Review of text-to-speech conversion for English. *Journal of the Acoustical Society of America* 82, 737-793.

Kondo, Yuko (1994). Targetless schwa: is that how we get the impression of stress timing in English? *Proceedings of the Edinburgh Linguistics Department Conference '94*, 63-76.

Koopmans-van Beinum, F.J. (1994). What's in a schwa? Durational and spectral analysis of natural continuous speech and diphones in Dutch. *Phonetica* 51, 68-79.

Lass, Roger (2007). On schwa: synchronic prelude and historical fugue. Donka Minova (ed.) *Weak Segments in English.*

Liljencrants, Johan, and Björn Lindblom (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48, 839-62.

Lindblom, Björn (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35, 1773-1781.

Lindblom, B., H.M. Sussman, G. Modaressi, & E. Burlingame (2002). The trough effecr: implications for speech motor programming. *Phonetica* 59, 245-262.

Löfqvist, Anders, and Vincent Gracco (1999). Interarticulator programming in VCV sequences: Lip and tongue movements. *Journal of the Acoustical Society of America* 105, 1864-1876.

Maiden, Martin (1995). Vowel systems. Martin Maiden and Mair Parry (eds.) *The Dialects of Italy*. Routledge, London, 7-14.

Ohala, Manjari (1983). *Aspects of Hindi Phonology*, volume II. MLBD Series in Linguistics, Motilal Banarsidass, New Delhi.

Oostendorp, M. van. 2000. Phonological Projection; A Theory of Schwa in Optimality Theory, Mouton - De Gruyter, Berlin.

Padgett, J. & Tabain, M. (2005). Adaptive dispersion theory and phonological vowel reduction in Russian. *Phonetica* 62, 14—54.

Peterson, G., & I. Lehiste (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32, 693-703.

Van Bergem, Dick R. (1994). A model of coarticulatory effects on the schwa. *Speech Communication* 14, 143-162.

StataCorp. 2007. *Stata Statistical Software: Release 10*. College Station, TX: StataCorp LP.

Sussman, H.M., K. Hoemeke, and F. Ahmed (1993). A cross-linguistic investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America* 94, 1256-1268.

Van Santen, J.P.H. (1992). Contextual effects on vowel duration. *Speech Communication* 11, 513-546.

Wood, Sidney, & Tore Pettersson (1988). Vowel reduction in Bulgarian: the phonetic data and model experiments. *Folia Linguistica* 22, 239-262.