A TENTATIVE ARGUMENT FOR REGIONAL QUALITY TARGETS IN ASSIMILATING SCHWA

Olli-Pekka Marttila

May 2016

Minor Thesis

University of Turku

School of Languages and Translation Studies

Phonetics

TURUN YLIOPISTO

Kieli- ja käännöstieteiden laitos/Humanistinen tiedekunta

MARTTILA, OLLI-PEKKA: A tentative argument for regional quality targets in assimilating schwa

Sivuaineen tutkielma, 49 s., 4 liites.
Fonetiikka
Toukokuu 2016

Amerikanenglannin redusoitunut švaa-vokaali tuotetaan puheessa laadultaan huomattavan monimuotoisena. Kirjallisuudessa ei kuitenkaan vallitse selkeää käsitystä siitä, mitkä tekijät ehdollistavat tätä švaan laatuvaihtelua, ja etenkin sen mahdollinen sosiaalinen ulottuvuus on jäänyt aiemmissa tutkimuksissa varsin vähälle huomiolle. Tässä tutkielmassa pyritäänkin selvittämään, millainen vaikutus puhujan alueellisella murteella on hänen tuottamansa švaan akustiikkaan.

Tutkimukseen valittiin 21 informanttia muutamista Yhdysvaltain eteläisistä ja läntisistä osavaltioista. Informanttien haastattelunauhoitteista poimittiin analyysiin yhteensä 433 švaan F1- ja F2-formanttiarvot siten kuin ne esiintyivät sanassa *the*. Analyysissä havaittiin, että eteläisten puhujien švaat keskittyivät vokaalien [u,ʊ] tuntumaan, kun taas läntisten informanttien švaat asettuivat vokaalien [i,u] väliin. Tulokset antoivat aihetta kysyä, onko eteläinen švaa sulautunut takavokaaleihin [u,ʊ]. Kysymystä varten suoritettiin jatkotutkimus, johon valittiin yksi ensimmäisen tutkimuksen eteläisistä puhujista. Puhujalta mitattiin 66 švaan ja 45 [ʊ]:n formantit. Näitä tarkastellessa löydettiin tilastollista viitettä siitä, että puhujan švaa oli saattanut sulautua vokaaliin [ʊ].

Tulostensa pohjalta tutkielma ehdottaa, että *the*-sanassa esiintyvässä amerikanenglannin švaassa saattaa toteutua puhujan alueellista murretta mukaileva tavoitevaihe siitä huolimatta, että tämä vokaali on samanaikaisesti huomattavasti kontekstiinsa assimiloitunut. Tavoitevaihe näyttäisi lähtökohtaisesti sijoittuvan F2:ssa vokaalialueen keskelle, ollen kuitenkin eteläisten puhujien kohdalla altis siirtymään taaemmas kohti vokaalilaatuja [u,ʊ].

Asiasanat: schwa, dialect, sociophonetics, formants

Table of Contents

List of Figures and Tables

1. Introduction

Reduced vowels in American English are commonly transcribed with /ə/ (*schwa*), a symbol that represents a neutral, mid-central vowel quality. But in contrast to the conceptually established quality of /ə/, acoustic studies have found notable phonetic variation in the realizations of these reduced vowels in speech, their qualities spanning the vowel space from back to front and high to low. Although the existence of this variation is relatively well attested, as will be discussed further in Chapter 2, its correlates remain unclear.

This paper presents an investigation into the social basis of the acoustic variation in American English schwa,[1] seeking to find whether there is stratification in schwa quality along regional U.S. dialects. The investigation is realized as two consecutive studies—both reported in this paper—the first study evaluating acoustic differences in schwa between Southern and Western U.S. speakers, and the second study investigating the speech of a single Southern informant to pursue some of the findings made in the first study. In brief, this paper proposes the following research question *Q*.

Q: Does phonetic variation in American English schwa pattern regionally?

The results of the two studies suggest potential regional patterning in schwa quality, and particularly that some speakers in the South may merge very short schwas (in the definite article *the*) with certain full vowel qualities. A more descriptive summary of the results is given in Chapter 5, following a review of the relevant literature in Chapter 2, a report of the first study in Chapter 3, and of the second study in Chapter 4. In addition to concluding the paper, Chapter 5 suggests some avenues for future research.[2]

---

[1] The term *schwa* and the symbol ə are used in this paper to referer in general to unstressed vowel qualities commonly transcribed as [ə] or [ɨ], and not exclusively to the mid-central one.

[2] This report uses IPA symbols for transcription. When referring to abstract phonemes, they are encased in forward slashes //. When referring to those phonemes as produced by a speaker, they are encased in square brackets []; but so that a speaker's utterance of e.g. /u/ in *goose* is transcribed as [u] regardless of whether it was produced as a high-back quality.

2. Literature review

The usual description of schwa (cf. Lass 2009) evokes the mid-central /ə/ (IPA 1999:12), a quality not infrequently granted the label 'neutral vowel' (e.g. Kenyon & Knott 1949). On the other hand, the term "obscure vowel" (Krapp 1919:34) is perhaps a more fitting description of schwa as realized in speech, given that one finds there not a unitary, neutral quality but instead a vowel of notable acoustic complexity. Lass, for instance, describes in his American English accent "at least" seven phonetic variants of schwa: [ɘ,ɪ,ɪ̈,ɤ̈,ɔ,i,ɜ] (2009:52–53). The rest of this chapter will expound on the view of American English schwa as a highly variable quality, and while doing so, the argument will be posited that variation in schwa likely finds its correlates partly in the social domain, being thus potentially conditioned by e.g. the speaker's age, sex, and region.[3]

Following contemporary reports of American English schwa being so susceptible to coarticulation that it might in in fact be a vowel completely assimilated to its context, Browman & Goldstein (1990) set out to find whether schwa indeed exhibited such a lack of a quality target. To their aim, Browman & Goldstein undertook an X-ray study of the articulations of a single speaker of American English uttering nonsense words of the form /pV1pəˈpV2pə/, where V1 and V2 were experimentally varied through all combinations of /i,ɛ,ɑ,ʌ,u/. Though Browman & Goldstein did not analyze the second, final schwa in these nonsense words, they concluded of the first that "the tongue position associated with medial schwa cannot be treated simply as an intermediate point on a direct tongue trajectory from V1 to V2" (1990:205), proposing, in other words, that schwa was not fully assimilated to its context but instead possessed a quality target. On the nature of this target, Browman & Goldstein comment that "schwa seems to involve a warping of the trajectory toward an overall average or neutral tongue position" (1990:208).

A recent contrary view to that of Browman & Goldstein was proposed in Flemming & Johnson 2007. In their acoustic study, Flemming & Johnson observed nine speakers of American English uttering a set of English words in which, depending on the word, schwa occurred in either medial or final position. In medial positions, Flemming & Johnson found schwa to have a low first formant

---

[3] The speaker's geographic region is categorized in this paper as a social variable. While region is not as immediately recognizable as a social factor as e.g. age or sex, it is assumed here that forms of pronunciation identify the speaker as a member or non-member of a given social group, and thus that the speaker's region defines a social boundary (e.g. U.S. state) which stands to influence—relative to the other members and non-members of that boundary—the way in which the person speaks.

(*F1*), and a second formant (*F2*) which exhibited no specific target and was instead notably spread between [i,u] in its realizations. Flemming & Johnson saw these formant dynamics as symptomatic of the assimilating of medial schwa to its segmental context. In final position, on the other hand, Flemming & Johnson observed that schwa was drawn in F1 and F2 toward a seeming target at the mid-central position in the vowel space. In the rest of the present paper, the mid-central schwa will be referred to as *phonemic schwa* and the assimilating schwa as *phonetic schwa* (these were not terms used by Flemming & Johnson, however).

In their article, Flemming & Johnson (2007) argue that underlying the distinction between phonetic and phonemic schwa are two speaker-inherent factors: (1) the desire to maintain perceptual contrasts between vowel categories, and (2) the desire to minimize articulatory effort. Schwa in unstressed word-final position,[4] Flemming & Johnson propose, contrasts with certain other vowel qualities, necessitating its production as phonemic schwa, i.e. with a predictable target distinct from the qualities it contrasts with. In unstressed non-final position, on the other hand, vowel contrasts are neutralized, easing restrictions on the quality of schwa and allowing reduction in articulatory effort via the assimilation of schwa to its context.

Flemming (2009) adds to Flemming & Johnson's argument by proposing that there is an interplay between vowel duration and vowel contrasts, such that the desire for economy in articulation motivates the neutralization of contrasts in short segments, where contrasts would be physically more difficult to produce. Flemming builds, in other words, on Lindblom's (1963) undershoot model, which argues a positive correlation between the extent to which vowels reach their intended (target) qualities and the duration of the vowel, thereby that as duration decreases, the vowel increasingly fails to reach its target and grows more assimilated to the surrounding segments. Flemming proposes that the final position is conductive to a longer schwa (cf. e.g. Crystal & House 1988 on final lengthening), which in turn allows for more varied quality targets to be realized and thus category contrasts to form. By example, Flemming reports (2009:92) that medial schwas in Flemming & Johnson's (2007) study were on average 64 ms in duration, while the final schwas averaged 153 ms.

Flemming himself conducted an acoustic study (2009) in which he asked four speakers of American English to produce nonsense words of the form /ˈbVCəˌCVt/, where V cycled through /i,æ,ɑ,u/ and C through /b,d,g/ and in which schwa occurred in medial position. Flemming hypothesized

[4] Flemming & Johnson use the term *stem-final* rather than *word-final*, but the distinction they obtain in doing so is not relevant here.

that the formants of medial schwa at its temporal midpoint would be linearly predictable given the formants of the surrounding vowels and with certain assumptions about the places of articulation of the surrounding consonants—in other words, that medial schwa would be assimilated to its context. Flemming's hypothesis appeared validated, as he found a linear model of that kind to account for up to 86% of the variation he observed in schwa F2, and up to 72% of the variation in schwa F1. This suggests that the acoustic realization of phonetic schwa is a roughly linear glide between the preceding consonant and the following consonant. Aware of the conflict between his findings and those of Browman & Goldstein (1990) as to whether medial schwa assimilates to its context or targets a neutral quality, Flemming suggests (2009:88) that the way in which the nonsense words used by Browman & Goldstein were constructed may have mislead the speaker into producing those words as compounds, effectively rendering the medial schwas word-final.

To sum up the discussion so far, there appear to be two kinds of schwa in American English: (1) a phonemic schwa, which targets a neutral, mid-central position in the vowel space and may be specific to such unstressed positions that are conductive to vowel lengthening; and (2) a phonetic schwa, which lacks a distinct target and whose quality is instead largely determined from the qualities of the surrounding segments to which it assimilates, this assimilation being possibly motivated by the schwa's short duration.

Having said that, there are two reasons to view this typology with some suspicion. First, all three of the studies discussed above—Browman & Goldstein (1990), Flemming & Johnson (2007), and Flemming (2009)—are experimental: their data were obtained by recording speakers in a laboratory environment reading prepared lists of (often nonsense) words. While such methods allow for the control of certain confounders—e.g. of segmental context, by having speakers produce specific minimal pairs—they are not guaranteed to elicit the sort of speech that the informants would produce in *natural* (unscripted) conversation. This has been a concern perhaps notably in computer speech recognition, where recognition models trained on scripted speech have operated well on scripted material but produced notably higher error rates when presented with spontaneous conversation (Ernestus & Warner 2011 and references therein). A reason for this failure is that reduced pronunciation forms in spontaneous speech are more diverse (complex) than those found in scripted speech (*ibid.*). With that in mind, and given data of schwa elicited under experimental conditions, one must ask, how well do these observations generalize to natural speech?

Unfortunately, this author is not aware of any published research on the acoustics of American English schwa in unscripted speech. While such literature likely does exist, it does not appear to be in the mainstream, which may to some extent suggest a lack of interest in the field about the possible mismatch between what is observed of schwa experimentally and how it is produced in less formal settings. Whichever the case, it seems that one can argue that models of schwa derived experimentally—e.g. the distinction and distribution between phonemic and phonetic schwa proposed by Flemming & Johnson (2007)—should also seek to demonstrate their capacity to represent natural speech, unless these models are not intended as representatives of natural speech.

The second reason to be critical is that the typology does not account for social factors; that is, for instance, that men may speak differently than women, that older speakers might use qualitatively distinct pronunciations from those of younger speakers, or that speakers from a certain region might gravitate toward an allophone different from that favored in another region. Theoretically, it does not seem that schwa should be immune to such socially-patterned variation; however, none of the three schwa studies discussed above report having accounted for the potential effects of social factors on their results—e.g. whether it matters, in the case of Flemming & Johnson (2007), that most of their informants spoke Western dialects. Indeed, this author is again not aware of *any* acoustic research on American schwa that evaluates in particular the role of social factors on schwa quality (although, again, this does not mean that such research does not exist). Reports known to the author do exist in the literature of socially-conditioned phoneme-level variation in reduced vowels—for instance, Kurath & McDavid, Jr. (1961) describe regionally-based variation, Byrd (1994) finds possible differences between men and women, and Keating et al. (1994) suggest age-conditioned quality shifts—but even these phoneme-level descriptions appear scattered and perhaps somewhat coincidental.

Looking to generate more data on acoustic variation in American schwa, with a view specifically to the role of social factors in that variation, the author conducted an acoustic study (Marttila 2015) of 11 speakers of Southern American English, aged 19–84, from the Atlanta region in the state of Georgia. For that study, casual, unscripted interviews of the speakers were obtained in audio form from the Roswell Voices and Atlanta Survey Project corpora (LAP 2015). Hypothesizing that the extent to which phonetic schwa assimilates to its context is partly conditioned by the speaker's age, Marttila measured the F1 and F2 trajectories from onset to offset at 3 ms intervals of roughly 2

100 schwa tokens occurring in the carrier word *the*, selecting ca. 500 of those tokens for analysis.[5] In the analysis, limited for practical reasons to F2, Marttila found a quadratic correlation between speaker age and the extent to which schwa's F2 trajectory deviated from a linear glide: schwas uttered by the middle-aged informants tended on average to deviate most from this acoustic pattern of assimilation. These deviations were rather small, however, less than 15 Hz on average at the point of maximum difference from a linear glide, and did not notably influence the resulting quality of schwa even among the middle-aged informants.

The actual quality of schwa, however, as measured by Marttila (2015) from the temporal midpoint of schwa, revealed another quadratic correlation. The older speakers in particular exhibited ratios of schwa F2 to [u] F2 close to 1, while this ratio among the middle-aged speakers approached 1.5 (see Figure 8 in this paper). Marttila interpreted ratios close to 1 as reflecting a Southern variant of schwa, whose defining characteristic he suggested to be an interaction (or partial merging) with [u].[6] The higher ratios among the middle-aged speakers were seen by Marttila as typifying a non-Southern variant of schwa, whose use by the middle-aged Southerners was, in Marttila's view, potentially motivated by a negative perception of the Southern quality of schwa as a vernacular form.

Besides the hypothetical nature of his results (an acoustic study can measure speech output but not speaker intent or attitude), there are two limitations to Marttila's (2015) study that are relevant for this paper. First, the formant values of [u] for each speaker were the means of three repetitions of that vowel. For some speakers, however, these repetitions exhibited bimodality in F2, likely reflecting contextually-motivated allophonic variation (cf. e.g. Labov, Ash & Boberg 2006:152–153) which Marttila did not control for, but in any case casting doubt on the inference that schwa had coincided with [u] in F2 among some speakers, since the mean [u] F2 was thus potentially an ambiguous value. The second limitation is that Marttila's sample of speakers was regionally biased: the middle-aged informants were from Atlanta, while the young and old speakers were from Roswell, an edge city of Atlanta. If the differences observed by Marttila in schwa were indeed socially-based—which in itself is not certain—this bias would make it unclear whether it was age or regional dialect at play.

Despite their limitations, in proposing the existence of a potential quality target ([u]) in strongly-assimilating schwa, Marttila's (2015) observations raise the question of what it means for schwa to be

[5] The criteria by which these tokens were chosen were that they had to occur in non-disfluent contexts and be followed by a consonant other than [ɹ,l,w,j,h].

[6] The schwas typically had slightly higher F1 than [u], however.

acoustically assimilated. That is, if one reports (as did Flemming [2009] of American English, Kondo [1994] of British English, and van Bergem [1994] of Dutch) that the formants of an assimilating schwa display a linear glide between its flanking segments, one has described the formant dynamics of the schwa but not proof that the schwa had no quality target. Flemming was aware of this issue, as although he proposed that the assimilation to context of phonetic schwa signals its lack of target, he was careful to note that a target might in fact be encoded in the surrounding articulations to which the schwa is assimilated (2009:89–90).

Some further support for the notion that phonetic schwa could possess a quality target may derive from findings by Menezes et al. (2015). Analyzing—both acoustically and articulatorily—schwas in the carrier word *the* as uttered by seven speakers of Western American English, Menezes et al. found that the speakers differed notably from each other in the articulatory strategies with which they produced schwa, but not in their resulting acoustic output (F1 and F2). From this, Menezes et al. suggest that their speakers' schwas exhibited a shared acoustic target but lacked a similarly universal articulatory target. Although it is not clear that these schwas were indeed phonetic, since Menezes et al. did not report formant trajectories or durations, the definite article *the* as a high-frequency function word is conductive to considerable reduction (cf. Bell et al. 2009), and given that Marttila (2015) found schwa in *the* to be phonetic, it may not be unreasonable to assume that a phonetic schwa likewise obtained in the data of Menezes et al., and thus that they described an acoustic target in phonetic schwa.

A final point to consider in this chapter is whether phonetic-level variation in schwa is perceptually salient, i.e. if that variation should stratify along social factors, whether the listener would be able to perceive this social information in the speech stream. With the question of sub-phonemic discrimination in mind, Alfonso & Baer (1982) asked 12 informants to listen to ca. 25-ms schwa segments excised from speech in which the schwas had preceded certain front and back vowels across an intervening consonant (the schwas preceding front vowels differed from those preceding back vowels by roughly 100 Hz in F1 and by 150 Hz in F2). Alfonso & Baer found that the informants performed above chance level in labeling the schwas on whether they had preceded front or back vowels. These findings imply that even short schwas can carry perceptible information of their segmental context, and thus that they may also carry such information of the speaker's social identity should their quality be socially conditioned.

3. First study

Based on a review of the literature as given in the previous chapter, it seems that further research is warranted to investigate potential social correlates of phonetic variation in schwa; with that research perhaps best orienting itself to natural speech, of which data appears most lacking. This study aims to contribute to those needs by examining whether phonetic schwa in Southern American English has an acoustic target, and whether that target, if any, might be specific to the South (or to particular regions in the South).[7]

The guiding assumption in this study was that the merging of schwa to another, full vowel quality, previously reported as a possibility by Marttila (2015), would implicate a quality target in schwa; the focus of this study was thus pinned on comparing the quality of schwa to that of other vowels, looking for signs of merger. The study was considered exploratory, however, so no specific hypothesis was proposed—promising findings were investigated more stringently in a second study (Chapter 4). In the rest of this chapter, an account is given of the first study, beginning in Section 3.1 with an outline of the materials and methods used, continuing to a description in Section 3.2 of the results obtained, and finishing with a discussion of the results in Section 3.3.

3.1 Material and methods

3.1.1 Corpora and interviews

Southern and Western speech data were obtained for analysis from the Linguistic Atlas Project (*LAP*; LAP 2015), a collection of audio corpora of chiefly American English speech, with each LAP corpus consisting of unscripted field interviews generally from one or more U.S. states. Three LAP corpora were made use of in the present study: the Linguistic Atlas of the Gulf States (*LAGS*), the Augusta Survey (*AUG*), and the Linguistic Atlas of the Western States (*LAWS*). A summary of these corpora can be given as follows (LAP 2015). LAGS in its entirety consisted of 914 speakers' interviews, conducted between 1968 and 1983 in a number of Southern U.S. states. AUG, on the other hand, was a smaller dialect survey of 37 speakers, interviewed in 1976–1977 in and around Augusta, Georgia. Finally, LAWS represented a corpus of Western speech, with then ongoing field interviews having begun in 1988 (the LAWS interviews used in this study were conducted in 1988–1990).

---

[7] The regional partitioning of the U.S. here follows that of Labov et al. (2006).

Speaker*	State	City/town	Locality	LAP code	LAP corpus
Georgia					
AT33	Georgia	Atlanta	Urban	INF104	LAGS
AT45	Georgia	Decatur	Urban	INF097	LAGS
AT65	Georgia	Atlanta	Urban	INF099A	LAGS
AU23	South Carolina	North Augusta	Urban	INF018	AUG
AU33	Georgia	Augusta	Urban	INF017	AUG
AU39	Georgia	Augusta	Urban	INF006	AUG
AU73	Georgia	Augusta	Urban	INF008	AUG
Tennessee					
TN17	Tennessee	Memphis	Urban	INF516	LAGS
TN38	Tennessee	Sweetwater	Urban	INF040	LAGS
TN45	Tennessee	Selmer	Urban	INF499	LAGS
TN49	Tennessee	Paris	Urban	INF487	LAGS
TN51	Tennessee	Maryville	Urban	INF027	LAGS
TN60	Tennessee	Ripley	Urban	INF508	LAGS
TN82	Tennessee	Chattanooga	Urban	INF055	LAGS
West					
CO71	Colorado	Buena Vista	Rural	INFCO022	LAWS
CO80	Colorado	Westcliffe	Rural	INFCO035	LAWS
CO82	Colorado	Walden	Rural	INFCO026	LAWS
WY73	Wyoming	Rock River	Rural	INFWY013	LAWS
WY74	Wyoming	Alva	Rural	INFWY001	LAWS
WY76	Wyoming	Ten Sleep	Rural	INFWY009	LAWS
WY82	Wyoming	Farson	Rural	INFWY014	LAWS

* Age is given in the numeric part of the speaker's name, e.g. AT33 = 33 years old.

Table 1. Speakers in the study.

A subset of the relatively large number of informants in the three corpora was chosen according to the following criteria (these criteria were designed to grant comparability with the results in Marttila 2015, so as to allow building on them in this study). The informants were to be white females who spoke American English natively and who would together represent a wide spectrum of age from young to old. Because the audio quality in the LAP recordings was highly variable from one interview to the next—the worst being unusable for acoustic analysis—each informant's interview recording had to be of sufficient quality to allow for reasonable confidence in formant extraction. In total, 21 speakers from the three corpora satisfied these criteria and were chosen for study. The speakers are listed in Table 1, and their geographical spread is indicated in Figure 1.

In the resulting sample, there were thus seven speakers each from Tennessee and Georgia, together representing Southern speech, and a combined seven speakers from Colorado and Wyoming, representing Western speech. The Georgian subsample was intentionally stratified so that half of the speakers would be from one city (Atlanta, corresponding to Marttila 2015), and the other half from another city (Augusta).[8] This sampling allowed—in theory—for evaluation of potential differences in schwa between Southern and Western speech, between speech in two Southern regions (states), and between two cities in a single Southern state.[9]

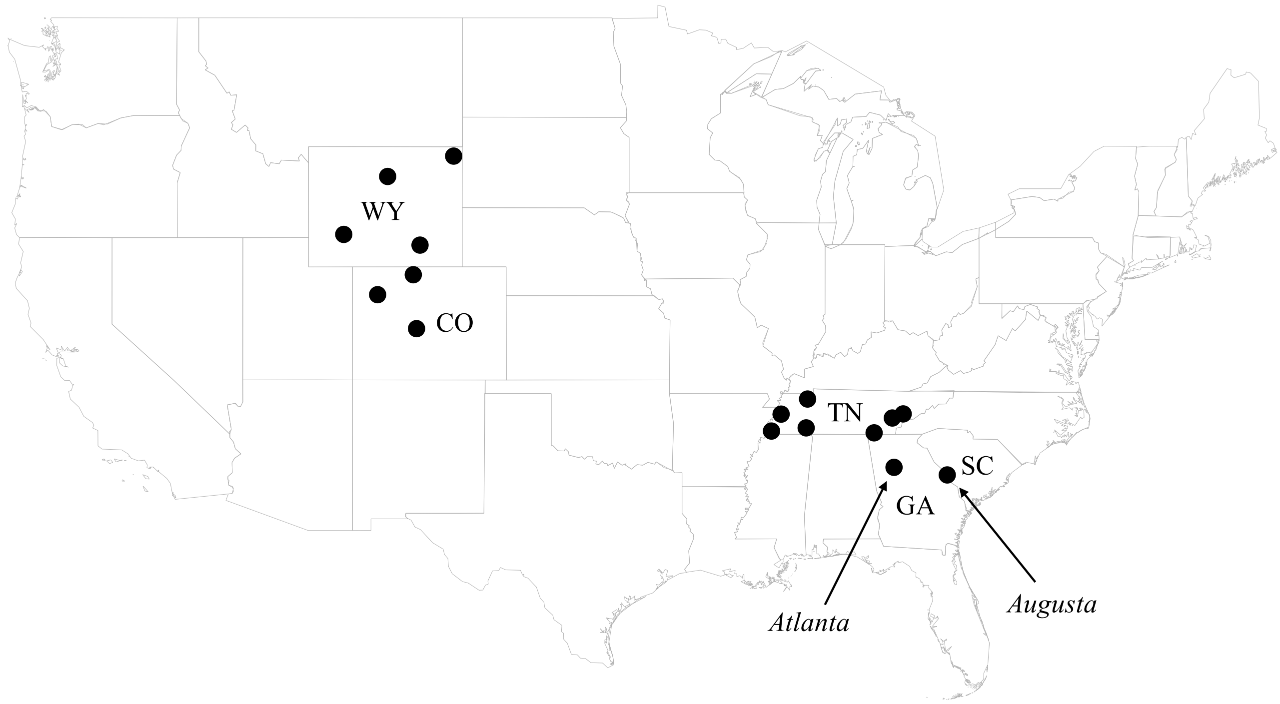

Figure 1. Geographic scatterplot of informants (●). Base map: USGS 2015.

Examination of Table 1 reveals that the Western speakers chosen are relatively older than the Southern speakers, and certainly this Western group does not satisfy the criterion of wide spread in

[8] Originally, the Atlantan subsample consisted of four informants, but one of them was found producing very few instances of schwa and was excluded from study, leaving in the sample three Atlantans and four Augustans.

[9] Speaker AT45 was from Decatur rather than Atlanta; however, Decatur is located alongside Atlanta, as shown in Figure 9, and was thus considered part of the Atlanta region. Likewise, AU23 was located in North Augusta rather than in Augusta—and indeed, in South Carolina rather than in Georgia—but for the fact that these two cities bordered each other, they were considered here a single subregion.

age. This artifact was due to a limitation of available data of Western speech of the LAGS and AUG vintage, rather than by design. A similar effect of this limitation is seen in locality, the Southern speakers being urban whereas the Western informants were all rural. It follows that, should this study find differences between Southern and Western schwa, one could propose that the differences might be due to age or locality rather than regional dialect. This is true; however, one would still grant that the differences might be social in origin.

Each of the three corpora chosen for this study had adopted the LAGS questionnaire format (described in detail in Pederson et al. 1974). That is, the informant was given short questions in oral form by the interviewer—the questions having been designed to elicit pronunciations of specific lexical items—to which the speaker typically responded by either producing the intended item, or initiating a brief conversation with the interviewer on the topic of the question. Below are two excerpts from the interview of WY73 to illustrate these two kinds of question–answer pairs and the manner in which they were conducted (IV = interviewer, SP = informant).

(1) IV: The son of a man's brother is his . . .
SP: Nephew.
IV: Okay. And he, his relationship to that nephew, the, the nephew would call him his . . .
SP: Uncle.

(2) IV: We'll move on to talking about the house here.
SP: Uh-huh.
IV: And, first of all, maybe you could just describe some things as you come in, the furniture of the, uh, the, the things that we see around here. Just describe, give me sort of a list of things in your, in your house.
SP: In my house? Okay, well, let me tell you something about our ranch house, the, the, furniture that we had there. Of course, it was a section house. I think you know what that would be. Where the section foreman lived when, uh, the railroad was going through Wyoming, okay? Uh, it was a two-story house with, uh, upstairs that was, um, rather low roof. And in one corner, or one side of the upstairs, were windows that you could shoot out at, oh, Indians or whatever was necessary to shoot. (laughs)
IV: And, and the windows were made out of . . .
SP: Glass, uh-huh. And, uh, but mother and dad had very little furniture. Our table was made of pine boards, and our chairs were benches. . . .

Although the interviews were conducted in a fairly conversational tone, they were nonetheless scripted in the sense that the interviewer had a list of pronunciation items that he or she was required to elicit from the speaker. In this way, one can see in excerpt (2) that the interviewer interjected with a question about the material of the windows, a question which, while perhaps somewhat incongruous, was presumably intended to elicit the speaker's pronunciation of *glass*. The interviews, despite being technically conversational, thus tended to have a formal air to them; this shortcoming is discussed further in Section 3.3. Potentially mitigating this formality is that the interview appears to have been conducted in the speaker's home ("We'll ... [talk] about the house here. . . . [Give] me sort of a list of things . . . in your house."), rather than in a more formal setting (e.g. a laboratory). This choice of location does not appear to have been specific to WY73, but rather all of the interviews seem to have been conducted in this way. Although the LAP server (LAP 2015) typically made available several hours of interview audio for each speaker, for time economy, only the first hour of each speaker's interview was included in this study.

3.1.2 Extraction of data

The definite article *the* was used as the carrier word for schwa. In its typical, pre-consonantal (cf. e.g. Keating et al. 1994) form, e.g. [ðə], *the* contains a phonetic schwa,[10] occurs with relatively high frequency in unscripted speech, and allows for control of certain confounding factors in vowel reduction.[11] For this study, schwas were extracted for analysis from occurrences of *the* in each speaker's interview audio by the following methods and subject to the following restrictions. Data extraction was initiated by selecting one of the 21 speakers at random, and proceeding to mark by hand onto a Praat[12] (version 5.408) point tier each instance of *the* uttered by the speaker in her interview. This process was repeated for each speaker. Having thus obtained the locations of each occurrence of *the*, their order was randomized and relevant data was extracted from each of them.

The randomization of speakers and instances of *the* was obtained via the Mersenne Twister implementation of the C++ Standard Library (*std::mt19937*, seeded with the system clock). To randomize the order of speakers, each informant was entered into a Standard Library vector (*std::vector*), which was randomly shuffled using the Mersenne Twister generator (via a call to *std::shuffle*). For random-

[10] The schwa in *the* is assumed phonetic following Marttila's (2015) observations that schwa in this position is short in duration and assimilated in F2 trajectory.

[11] Lexical frequency, lexical class, neighborhood density, and—because *the* typically does not occur phrase-finally—phrase position; see e.g. Jurafsky et al. 2001, Bell et al. 2009, Yao 2011, Bell et al. 2003, respectively.

[12] Boersma & Weenink 2016.

izing the order of instances of *the*, a separate program was created by the author (*SLR*, Figure 2). SLR used the same method of randomization as for the speakers, but also provided a user interface designed to facilitate data extraction. That is, SLR would pick a random instance of *the* for the given speaker and copy the time at which that instance occurred in the audio onto the computer's clipboard, from which it could be copied into Praat to locate that portion of the audio.

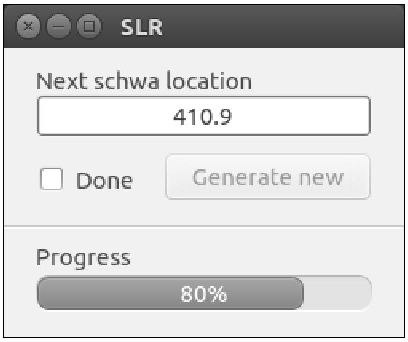

Figure 2. Screenshot of the SLR program.

From each istance of *the*, the following variables were extracted. Note that only the variables in (a) and (b) are reported on in the present study.

a. F1 and F2 of the vowel in *the*.

b. The duration of the vowel in *the*.

c. The consonant in *the*.

d. The word preceding and following *the*.

e. The phoneme following *the*.

Before describing the methods by which these variables were extracted, it is important to note that not all instances of *the* were included in the analysis. That is, *the* in disfluent contexts—segments in which *the* was immediately preceded or followed by e.g. pausing or repetition—were excluded, given that these contexts may occur with deviant patterns of vowel reduction (Fox Tree & Clark 1997, Bell et al. 2003). Instances of *the* uttered in a muted or elevated tone, relative to what this author had evaluated to be the normal tone of speech for that informant, were excluded, as loud speech has been found to impact vowel quality (Schulman 1989), and it was assumed that a similar deviation might occur in muted speech. Contexts where *the* coincided with notable background noise, e.g.

heavy traffic, were excluded on the basis that background noise has been found to influence vowel production (Junqua 1993), and in any case such noise can render formant readings less reliable. In *the*, different qualities of schwa tend to occur when preceding a vowel than when preceding a consonant (e.g. Keating et al. 1994), so the present study included only those instances of *the* which preceded consonants, as done also in Marttila 2015. Additionally, Marttila (2015) found potentially deviant patterns of reduction in pre-consonantal schwa when the consonant was one of [ɹ,j,l,w,h], for which reason the present study, as well as Marttila 2015, excluded schwas preceding one of those consonants. To further facilitate comparability with Marttila 2015, *the* produced with a consonant other than [ð,θ] were likewise excluded (but see further discussion on this matter later in this section).

F1 and F2 were extracted from the vowel in *the* (assumed schwa) using Praat. To do this, the temporal midpoint of the schwa was visually located on Praat's spectrogram, with reference to the waveform and audio where needed, and the formant values suggested by Praat at this midpoint were noted down. If Praat's suggestion did not appear to align with the underlying spectrogram, the 'Number of formants' option (accessed from Praat's spectrogram view via *Formants→Formant settings...*) was adjusted up or down to find better agreement. If such agreement was not found, that instance of *the* was excluded from further analysis. Where the consonant preceding schwa was a voiced dental fricative and thus not always clearly distinct from the following schwa on the spectrogram or waveform, reference was made on the spectrogram to the third formant. The point at which the third formant increased in intensity was taken as the commencing of the vocalic segment. If no such point could be located on the spectrogram, and the waveform did not implicate vocalic changes, that instance of *the* was excluded from analysis on the assumption that its vowel had been deleted. Likewise for consonants other than the voiced dental fricative, if the spectrogram indicated no clear formant structure following the consonant, the vowel was assumed deleted and that instance of *the* was excluded.

Coinciding with the measuring of its formants, the duration of the schwa was also measured, being taken as the time between the schwa's onset and offset as determined when locating the midpoint of the schwa for formant measurements. It should be noted that some echo tended to be present in the interview recordings, making it at times difficult to determine the schwa offset exactly so as not to include echo as part of the schwa segment proper. There is thus a degree of uncertainty, perhaps ca. 5–10 ms, in the measured durations (and in the estimated midpoint). Aspiration following consonantal release immediately prior to the schwa segment was excluded from the estimate of schwa onset location.

The consonant preceding schwa was marked broadly as either [θ] or [ð], such that voiced dental fricatives were categorized as [ð], and cases in which the transition from the consonant to schwa included a break in voicing—e.g. [θ, t, ɾ]—were categorized as [θ]. This categorization was motivated by the author's suspicion that the onset of schwa would be more difficult to determine when bordering a voiced consonant, as mentioned aboved, in which case unvoiced contexts might offer more reliable data of e.g. duration. This distinction in consonant type was not made use of in the analysis, however. Some instances of *the* were produced as [nə], [lə], or [ʒə]; with [nə] in particular being somewhat common;[13] these three variants were excluded from the study to limit variation in schwa's segmental context.

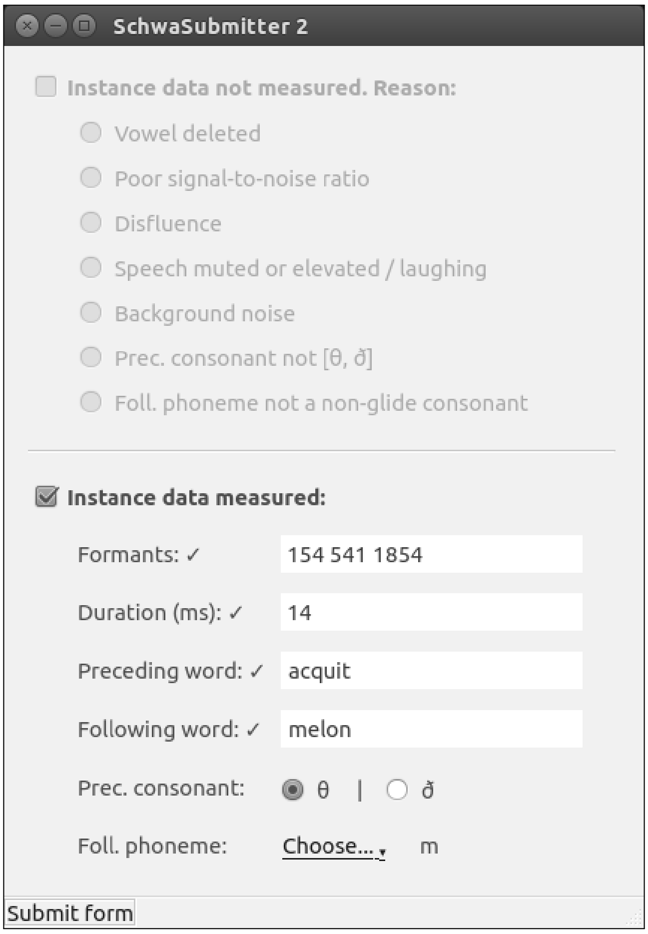

Figure 3. Screenshot of the SchwaSubmitter 2 program. The data shown are fictitious.

[13] In the present study, no count was taken of these variants, but in a separate dataset collected previously for Marttila 2015, [nə] accounted for 21% of the 2 091 instances of *the* where the vowel had not been deleted, while [lə] and [ʒə] together made up 5% of that dataset.

A utility program, SchwaSubmitter 2 (*SS2*, Figure 3), was created by the author to aid in organizing the extraction of data. SS2 was designed to provide a simple interface to either mark an instance of *the* as having been excluded, or to input the relevant data for that instance, the program then exporting that information into a database file for each speaker. Consideration in SS2's design was given to preventing accidentally missing data (e.g. resulting from the user forgetting to enter a certain variable). To that end, SS2 would allow exporting data only once all required fields had been entered. The *Formants* field (cf. Figure 3) accepted a direct copy & paste of text from Praat's formant output, minimizing potential mistakes in entering formant values by hand, while also guarding against missing data by parsing the input to make sure all the required formant values had been provided. The fields for words preceding and following schwa (*Preceding word* and *Following word* in the figure) performed an automatic search in a dictionary for the words entered, alerting the user to potential typographical errors if the words were not found.

F1 and F2 measurements of the full vowels [i,ɑ,æ,u,ʊ,o,ɔ] were also made of each speaker to establish the overall extent of the speaker's vowel space, and in particular to probe the qualities of her back vowels, to which schwa was assumed having potentially merged. For each speaker, four repetitions (but see below for [u]) were measured of each full vowel, submitting to analysis the median formant values of those repetitions. The repetitions were temporally stratified so that two tokens were measured from the first half of the interview and the other two repetitions from the second half; this was done to account for temporally-induced shifts in vowel quality, due to e.g. potential speaker fatigue toward the end of the interview. The full vowel formants were measured at a stable point along the vowel, where the vowel was assumed to have reached its minimum and/or maximum F1 and F2 (e.g. minimum F1 and maximum F2 for the high-front [i]). Anticipating context-dependent allophones of [u] (Labov et al. 2006:152–153), certain restrictions were placed on the contexts from which tokens of [u] were measured. Post-coronal [u] (i.e. [u] preceding a coronal consonant, e.g. in *do* [du]) was given its own category, *uC*, while any [u] preceding [l] were ignored. From other contexts, [u] was measured normally (e.g. from *move* and *food*).

Because of the potential for the segmental context to influence the quality of not just [u] but of any vowel, diverse contexts were favored in obtaining full vowel tokens, thus making an attempt at avoiding biasing the sample toward certain context-dependent forms. Where possible, each of the four repetitions of a given full vowel were measured from different words—e.g. for TN83, the four [i] tokens were extracted from *leading*, *university*, *seems*, and *she*. This diversity was generally not obtainable for [u], however. As seen in Table 6, the word *moved* and its variants tended to be the most common sources from which this phone could be obtained, and it was not uncommon that a speaker did not utter, in the first hour of her interview, four instances of [u] with the segmental limitations mentined above. For example, only two such instances of [u] were recorded of AT45.

Having completed the extraction of data from the 21 speakers, the full dataset to be submitted for analysis consisted of 433 schwas (selected from 3133 instances of *the* by the criteria given earlier in this section) and 652 full vowel tokens. Prior to analysis, the dataset was examined for potential errors and/or outliers. For this, the full vowel formants were plotted on an F1–F2 plane (a scatterplot in which x = F2 and y = F1), and any tokens appearing outlying relative to the other tokens of that vowel were re-investigated in Praat. If found erroneous, the token was re-measured or replaced with a new token. The F1 and F2 of each schwa were graphed as per-speaker box plots in SPSS (version 22), and any formant values indicated by SPSS as outliers on these box plots were re-investigated in Praat. In total, 16 outlying schwa measurements (4% of the dataset) from nine speakers were re-investigated; as a result, two schwa tokens were removed from the dataset, one token's F1 and F2 were re-measured, and the remaining 12 tokens were found to have been measured correctly, despite their outlying, and were kept unchanged

3.1.3 Methods of analysis

The dataset was analyzed with a combination of qualitative and quantitative methods. Quantitatively, formant medians were subjected to statistical testing. As in Marttila 2015, non-parametric tests were favored over parametric ones due to certain non-normal distributions (see also Kretzschmar & Schneider 1996:37, who recommend non-parametric methods for LAP data). For comparing vowel qualities across speakers, Mann–Whitney U tests were used, while paired comparisons—i.e. of sets of data produced by the same speaker—were computed as Wilcoxon signed-rank tests. The Euclidean distance between two vowel categories was also used as a measure of acoustic similarity between the categories (see below for details). Linear and quadratic correlations were investigated using polynomial regression. The threshold for statistical significance was set at $\alpha = .05$, with $p < \alpha$ thus considered significant. As this study was exploratory, α was not adjusted for multiple tests (cf. e.g. Wilcox 2003:407). All statistical tests were performed using SPSS (version 22). Averages were obtained as medians, rather than means, to reduce the influence of outlying values.

Qualitative analysis consisted of plotting formant data on the F1–F2 plane and visually inspecting these plots to infer possible relationships between schwa and the full vowels. To facilitate this analysis, the author's VowelGrapher program, developed in the course of earlier work (e.g. Marttila 2015), was used—in short, VowelGrapher allowed for interactive segmenting of the formant dataset for exploratory visualization. For the present study, VowelGrapher was extended by the author to allow the F1–F2 plots to be exported as vector images, which served as figures in this report.

When comparing formant data across speakers, it is beneficial to normalize the data to control for potential confounding variation resulting from anatomical factors (e.g. Adank 2003, Labov 2001:157–158). In general, formant normalization aims to translate speakers' raw formants into a relative, common frame which reduces or eliminates the anatomical component. In the present study, for plotting formant data, the S-centroid method of normalization (Watt & Fabricius 2002; Fabricius, Watt & Johnson 2009) was used, owing to the method's relative economy in that it requires informantion of the qualities of only four full vowel categories (cf. the relatively more commonly-used *z*-score method [Lobanov 1971], which requires information of all of the speaker's vowel categories).

The S-centroid method works as follows.[14] One first derives the F1 and F2 values of the estimated center of the vowel space (the *centroid*) from four corner vowels which are taken to represent the maximum acoustic extent of the vowel space. One then divides the formants of each vowel to be normalized by the formants of the centroid, translating the vowel's formants to a common, ratio-based frame. In this study, three of the corner vowels were [i,ɑ,æ], with the high-back corner defined in F1 by [u] and in F2 by [ɔ], to account for the extensive fronting of [u] in F2 by Southern speakers (see Section 3.3). The S-centroid-normalized formant values were thus given from $\frac{V}{S} \cdot 10$, where V is the vowel to be normalized, S is the centroid such that $S_{F1} = (i_{F1} + u_{F1} + ɑ_{F1} + æ_{F1}) \div 4$, and $S_{F2} = (i_{F2} + ɔ_{F2} + ɑ_{F2} + æ_{F2}) \div 4$, and the result is multiplied by ten to remove fractions.

In paired statistical tests, no normalization was used, as these comparisons were not between speakers. In unpaired tests, when comparing Euclidean distances, ratios of the individual vowels' raw formants were used. That is, the distances were calculated as two-dimensional Euclidean distances given the formants of the two vowels V1, V2 so that, in Euclidean space, $x = V_{F2}$ and $y = V_{F1}$, but so that V1 was set to lie at point $\{1, 1\}$ and V2 at point $\{V1_{F2} \div V2_{F2}, V1_{F1} \div V2_{F1}\}$.

[14] The method was originally developed for British English, but small modifications were proposed by Bigham (2008:135) to adapt the method to American English. The discussion here assumes Bigham's modifications.

3.2 Results

Figure 4 plots all schwas in the dataset by region. In these data, there appear to be two diverging trends: speakers in the West produced their schwas roughly between [i,u] and roughly coinciding with the median F2 of post-coronal [u] (uC), while schwas in Tennesse appear to cluster around the median qualities of [u,ʊ]. The Georgians do not appear as a whole to conform to one trend or the other; rather, it seems that the Augustans' schwa is somewhat like that in Tennessee, tending toward [u], while in Atlanta one can see clustering perhaps most strongly toward uC in F2.

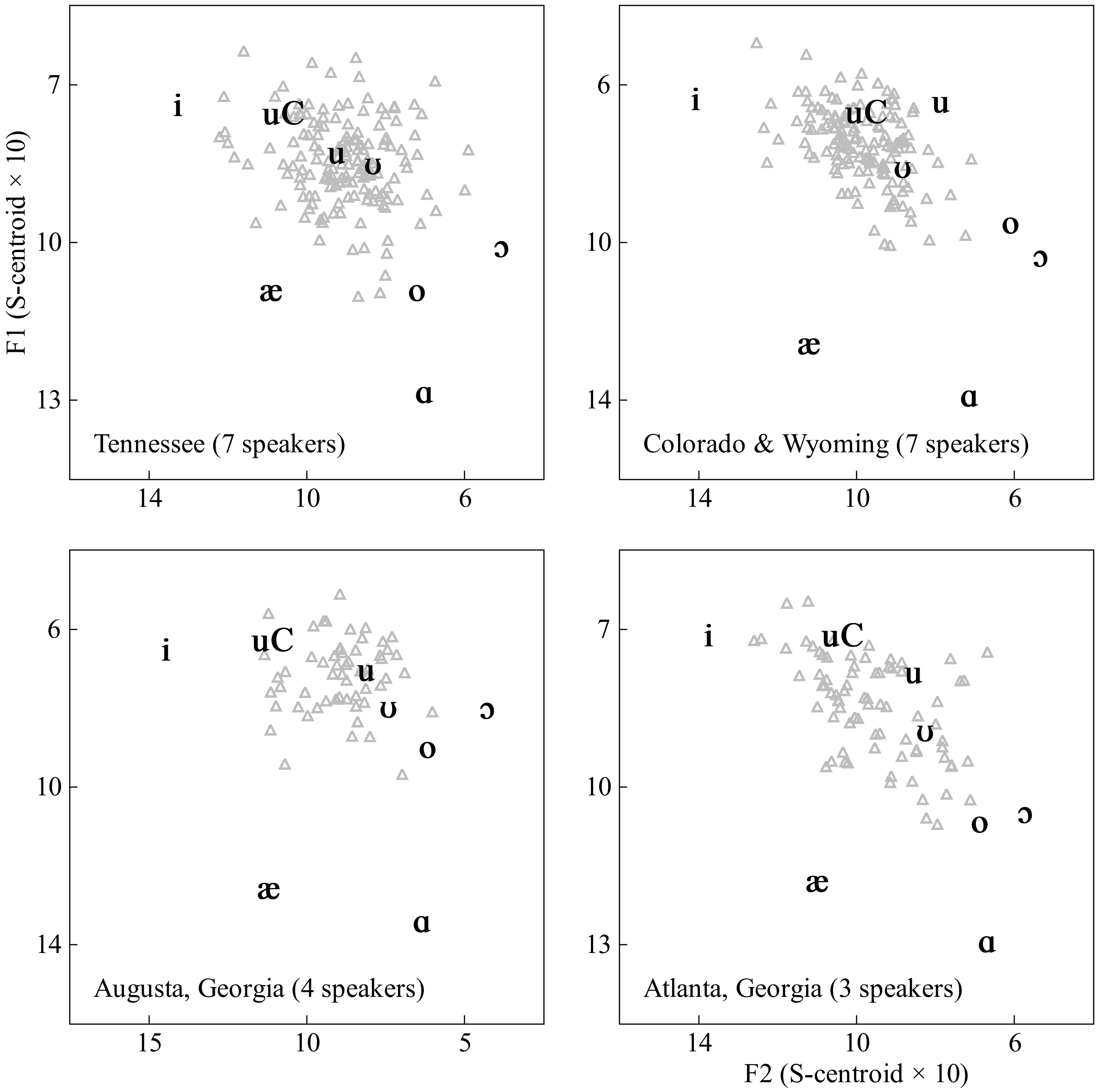

Figure 4. Scatterplots of all schwas (△) in the dataset, divided along the four geographic regions from which the speakers were sampled. Transcriptional symbols indicate the median locations of the corresponding full vowels (uC = post-coronal [u], as in e.g. *two*).

In addition to clustering toward certain individual full vowels, the schwas from each region appear drawn to a medial position in F2 between all full vowels combined. This is shown in Figure 5, which repeats Figure 4 but replaces the median full vowel qualities with individually-plotted full vowel tokens. Descriptive statistics of the acoustic properties of schwa as observed in the South and West are given in Table 2.

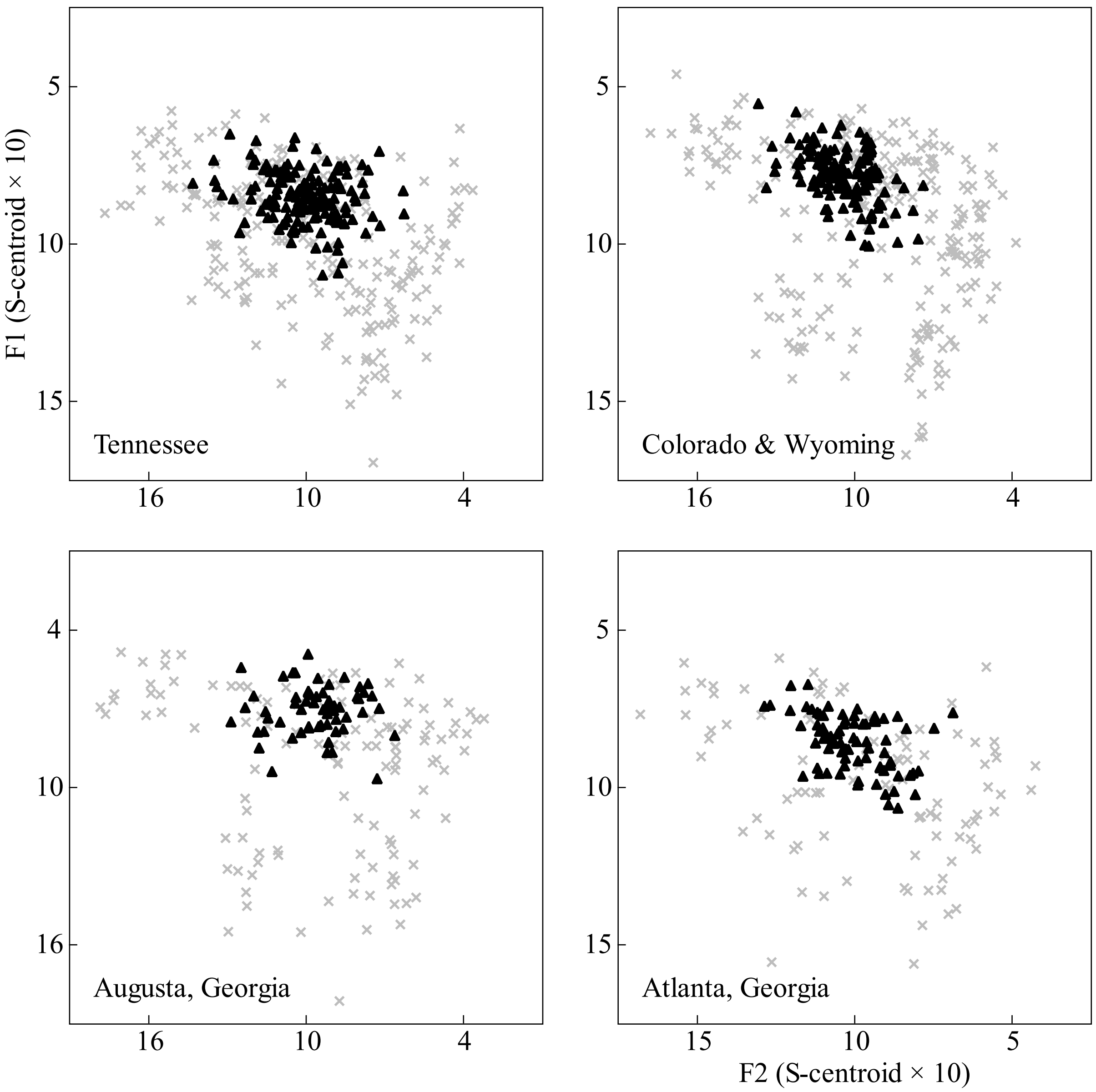

Figure 5. Scatterplots of all schwas (▲) and the full vowels [i,u,ʊ,o,ɔ,ɑ,æ] (×) in the dataset, divided along the four geographic regions from which the speakers were sampled.

The dataset is plotted by speaker in Figure 7, showing for each speaker the median Euclidean distances from schwa to [u,ʊ]. Looking at the left panel in the figure, which shows the distance between

	Median formants		Duration		
Region	F1	F2	Median	Range	Tokens
South	492 Hz	1732 Hz	13 ms	3–39 ms	293
West	456 Hz	1804 Hz	14 ms	3–39 ms	140

Table 2. Acoustic properties of schwa in the Southern and Western regions.

schwa and [u], the seven Tennesseans are seen exhibiting no overlap with the Western speakers. The four Augustans in this panel exhibit a distribution fairly similar to that of the Tennesseans, that is, three out of the four Augustans uttered schwa closer to [u] than did any of the Western informants. In contrast, the three Atlantans do not show as coherent a distribution, one Atlantan producing values similar to those in Tennessee, one similar to those in the West, and the third larger than any of the other speakers in the dataset. In the right panel of Figure 7 are shown the distances between schwa and [ʊ]. Although one again finds schwa produced somewhat closer to [ʊ] in Tennessee than in the West, there is some distributional overlap between these two regions. Neither the Augustan nor Atlantan informants exhibit consistent patterns here, the Augustans in particular showing potential bimodality between speakers whose schwa is relatively close to [ʊ] and those whose schwa is relatively far from it.

Because schwas in Atlanta and Augusta do not appear to exhibit trends as coherent as those in Tennessee and the West, the rest of this discussion will focus on comparing the Western and Tennessean speakers, of whom potentially more reliable inferences could be made. To that end, Mann–Whitney U tests were done to find whether the per-speaker median distances from schwa to [u,ʊ], shown in Figure 7, were significantly different in Tennessee compared to the West. These tests indicated that distances between schwa and [u] were significantly smaller in Tennessee (grand median = 0.07) than in the West (grand median = 0.317), $U = 0$, $p < .001$. The median distances from schwa to [ʊ], on the other hand, were not significantly different in Tennessee (grand median = 0.105) than in the West (grand median = 0.151), $U = 10$, $p = .073$. Paired Wilcoxon signed-rank tests indicated that, in the West, distances from schwa to [u] were significantly larger than the distances between schwa and [ʊ], $Z = -2.366$, $p = .016$. In contrast, the Tennessean schwa was not significantly more distant from [ʊ] than it was from [u], $Z = -0.845$, $p = .469$.

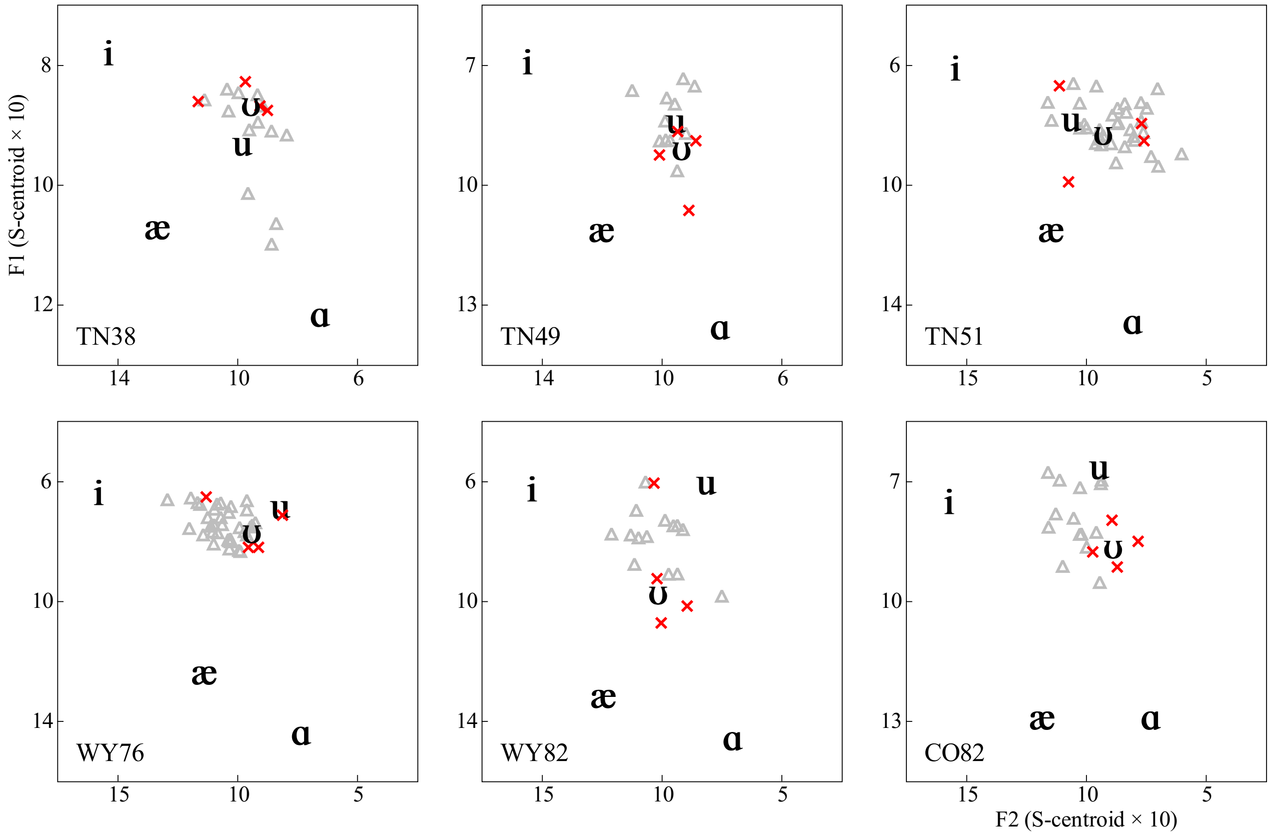

Figure 6. Schwas (△) in Tennessee (top panels) and the West (bottom panels) as uttered by select informants. Also indicated are the individual [ʊ] tokens (×) of each speaker.

It was suggested above that schwas in Tennessee appeared to cluster somewhat consistently toward [u,ʊ]. The three top panels in Figure 6 show the individual vowel plots of three select Tennesseans (see also Figure 18 [Appendix] for all seven Tennessean vowel plots), in which one can indeed find a degree of overlap between the distributions of schwa (△) and the median qualities of [u] and [ʊ]. On the other hand, the select three Western speakers in the three bottom panels of the figure (see also all individual Western vowel plots in Figure 17 [Appendix]) appear to have avoided—whether on purpose or otherwise—notably overlapping schwa with [ʊ] and especially with [u]. For clarity, the panels in Figure 6 indicate individual [ʊ] tokens with the × symbol. Wilcoxon signed-rank tests were done to find whether the median schwa F1 and/or F2 differed from the median formants of [u,ʊ] in either Tennessee or the West; i.e. whether schwa may have merged with these full qualities. Table 3 summarizes the results of those tests. The non-significant Tennessean *p* values indicate that schwa was statistically indistinct from [u,ʊ] in both F1 and F2 in that region, while in the West there was a significant distinction between schwa and both [u] (in F2) and [ʊ] (in F1 and F2).

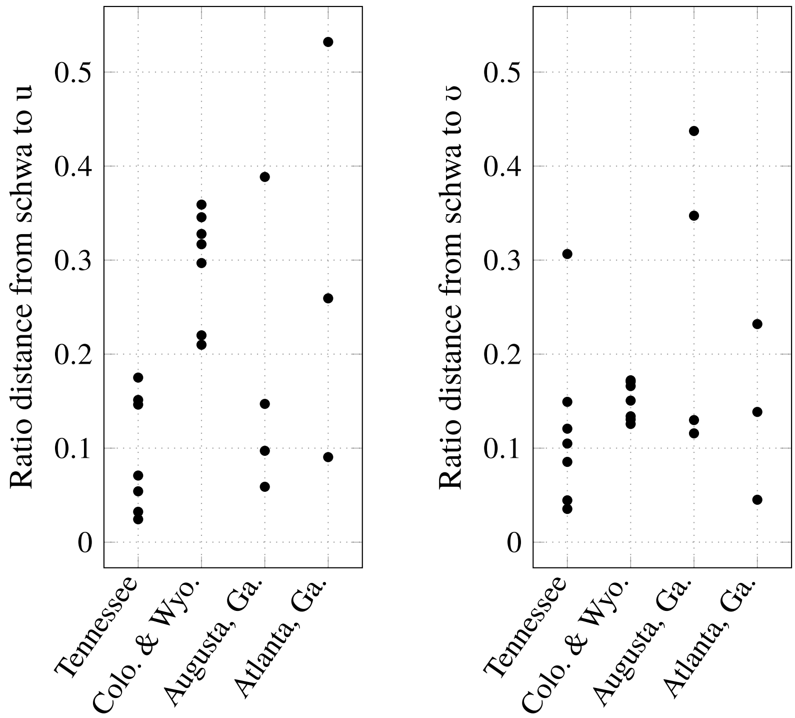

Figure 7. Scatterplots of geographic region vs distance from schwa to [u,ʊ].

	p		*Z*	
Comparison	F1	F2	F1	F2
Tennessee				
ə vs u	>.999	.375	0	−1.014
ə vs ʊ	.688	.078	−0.507	−1.859
Colorado & Wyoming				
ə vs u	.156	.016*	−1.521	−2.366
ə vs ʊ	.016*	.016*	−2.366	−2.366

* Significant difference (α = .05).

Table 3. Results of Wilcoxon signed-rank tests (Z) on formant medians to find whether schwa differed significantly from [u,ʊ].

Finally, one can compare the present data to those reported by Marttila (2015), who found among Southern informants the ratio between schwa F2 and [u] to be age-graded, such that older speakers in particular exhibited ratios close to 1, while middle-aged speakers produced ratios approaching 1.5. Figure 8 plots both Marttila's data (●) and the present observations from Tennessee (◊) and the West (×). Though the Western speakers are not widely enough distributed in age to allow for inference of age-graded effects, one may see that no such effect appears among the Tennesseans, whose ratios maintain a value of roughly 1 regardless of age. A polynomial regression of the Tennessean data (dependent variable = schwa F2 ÷ u F2, independent variables = speaker age, speaker age^2) found neither the linear nor the quadratic term significant.

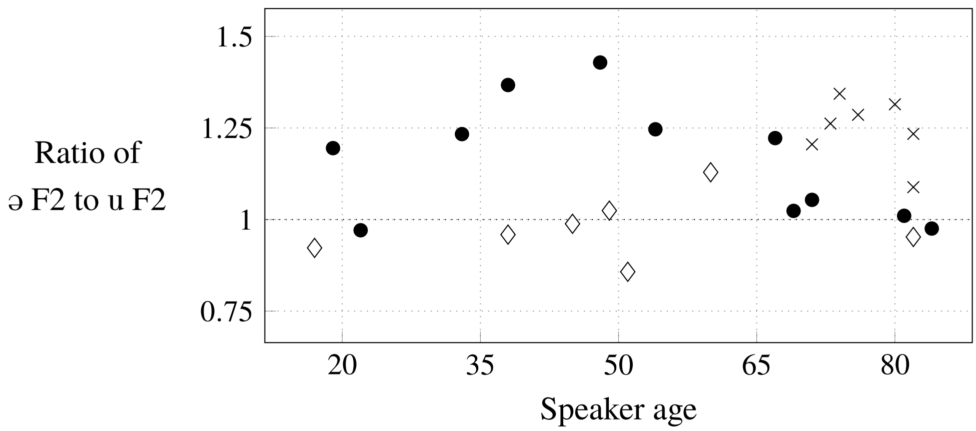

Figure 8. Scatterplot of speaker age vs the ratio of schwa F2 to [u] F2. The figure shows data from Marttila 2015:49 (●) against the present observations from Tennessee (◊) and the West (×).

3.3 Discussion

The 433 schwa tokens submitted to analysis (and plotted in Figure 4) indicate a possible distinction between Western and Tennessean speakers, whereby Western schwas seemingly tend to cluster to an area of the vowel space corresponding to a low F1 and an F2 roughly between [i,u], while Tennessean schwas appear drawn toward the median qualities of [u,ʊ]. The data from Georgia exhibit both Western and Tennessean trends, being divided into schwas that tend toward [u] (in Augusta), and schwas that display clustering toward the F2 midpoint between [i,u] (in Atlanta). The relatively small number of informants from the two Georgian cities casts uncertainty on this inference, however, more so than on the former assessment of the Western and Tennessean speakers. One would also keep in mind when examining the data in Figure 4 that, because the informants did not all produce the same number of schwa tokens, a single speaker with a higher-than-average token count could skew the overall distribution of tokens from her region.

Statistical tests of per-speaker formant medians showed a significant difference between Tennessee and the West in the relative acoustic distance between schwa and [u]. Further testing (cf. Table 3) indicated that while the Western schwa was significantly distinct from [u] in F2, no significant distinction was produced between schwa and [u] in Tennessee in either formant. The Western schwa was likewise significantly distinct from [ʊ], whereas again no such distinction was exhibited in Tennessee. These results suggest—very tentatively—a regionally-conditioned merging of schwa to one of [u,ʊ], occurring in other words in Tennessee but not in the West. However, it is not clear from these numeric tests whether it would be with [u] or [ʊ] that the Tennessean schwa had potentially merged, given that the distances between schwa–[ʊ] and schwa–[u] were not significantly distinct from each other in that region.

On visual inspection of the individual vowel plots (Appendix), one finds a possible source of this statistical ambiguity: the would-be target of the proposed merger may vary by speaker. The schwas of TN45 overlap both [u] and [ʊ], while TN38 and TN51 show potential—albeit not conclusive—overlap with [ʊ] in particular. Extending this inspection to the Georgians, schwas uttered by AU39 appear to overlap [u], while those of AT45 seem clustered about [ʊ]. Though the Western vowel plots also show a degree of overlap between schwa and [ʊ], it is generally not to the extent seen for some of the Southern informants nor to strongly imply a merger.

Assuming, then, for the argument that these observations are valid indications of merger or non-merger, they suggest two things: (1) the merging of phonetic schwa to a full vowel quality is a Southern but not a Western phenomenon; and (2) schwa's counterpart in such a merger is one of [u,ʊ], or potentially both, depending on here-unknown factors. Having said that, in evaluating these vowel plots and statistical tests, the word of caution strongly applies that the median quality of any given speaker's [ʊ] was represented by only four [ʊ] tokens, and the quality of [u] by even fewer tokens for some speakers (cf. Table 6 [Appendix]). These low quantities of data, while they may be suggestive of possible trends, do not allow for the confident inference of mergers or non-mergers.

Beyond such caution, there are three further sources of uncertainty in these results that should be pointed out. First, as also noted in Section 3.1.1, the Western speakers were on average older than the Southern informants, and, unlike the Southerners, were from rural rather than urban localities. It is not impossible that, should the results reflect real differences in schwa quality between the Western and Southern regions, the factors giving rise to those differences could be related to locality or age rather than to regional dialect.[15]

Second, although the present speech material was elicited in unscripted field interviews and can thus be assumed to be closer to natural conversation than speech obtained under experimental conditions (cf. Chapter 2), the way in which these interviews had tended to be conducted is not without criticism. Despite the LAGS field manual advising field workers to keep the interviews "as conversational as possible," (McDavid 1974:45) their results have since been criticized (e.g. Underwood 1974, Dumas 1976, and Baird 1987), typically along the lines that the LAGS interviews "more closely [resemble] a test than a conversation" (Underwood 1974:26). By the present author's evaluation, this criticism is to an extent accurate. A not infrequent exchange in these interviews consisted of (1) the interviewer describing to the speaker a particular thing and asking her to name that thing, followed by (2) the informant either producing the desired lexical item without much further comment or otherwise uttering some variant of 'I don't know,' after which (3) the interviewer proceeded to the next question on his or her list. The AUG and LAWS corpora inherited from LAGS not only its questionnaire format, as mentioned in Section 3.1.1, but also this rigid tone, although AUG perhaps less so than LAWS. Thus, it might best be said that the results in this study reflect unscripted speech in the context of a structurally formal interview.

The third uncertainty in these results is that, as seen Figure 5, both Southern and Western schwas tended to coincide with the middle of the vowel space in F2. This tendency is also evidenced in the individual vowel plots (Appendix) by a clustering of the schwas of most speakers to an F2 S-centroid value of 10, which, recalling from Section 3.1.3, represents a midpoint between [i,ɔ]. In other words, schwa may have had a quality target at the middle of the acoustic vowel space, and this target may have been shared in general by both Southern and Western speakers—indeed, one finds that the median schwa F2 in the South, 1732 Hz, is not very different from the mean F2 of the schwa in *the* as reported of Western speakers in Menezes et al. 2015 (1699 Hz). Whether the data in Menezes et al. are directly comparable with the present results, given e.g. that theirs was an experimental study (cf. Chapter 2), the implication nonetheless is that the seeming schwa mergers discussed earlier in this section might have been coincidental, plausibly resulting from the customary fronting of back

[15] Note also that there are other social variables which the present study did not account for, e.g. education and social status, which might plausibly condition and confound these outcomes.

vowels in the South,[16] which would have pushed at the least [u] toward—and possibly onto—the relatively central schwa, without necessarily involving an intentional merger. This possibility is further discussed in Section 4.3.

It remains to note here that although such a central quality target appears possible, the assumption that it exists specifically between [i,ɔ] is perhaps somewhat arbitrary, given that formants are not direct correlates of physical articulatory boundaries nor do they reflect the status of only a single articulator (e.g. the F2 of [ɔ] can be raised and lowered with changes in both tongue position and lip rounding). Interpolating a midpoint in this acoustic space, it seems, would require the speaker to keep track of the interactions of several articulatory variables while also accounting for the adapting of the vowel space's extrema to different speech situations (e.g. Ferguson & Quené 2014). In other words, a schwa target of this kind seems unnecessarily complex, and it may instead be that such a medial target, if any, would involve some underlying simplification to approximate a central quality, a quality which happened in these data to coincide relatively well with the midpoint between [i,ɔ].[17]

Finally, some comparisons can be drawn between the present results and those reported in this author's previous investigation (Marttila 2015). One recalls from Chapter 2 that Marttila (*ibid.*) found schwa in Atlanta and its edge city, Roswell, to be age-graded. The present findings reveal no such age-grading in Tennessee (Figure 8), and one can infer from Figure 19 (Appendix) that the schwas uttered by AT45, the only middle-aged Atlantan, appear to overlap [ʊ] and thus also do not conform to the anticipated pattern. On the other hand, the nearly-middle-aged Atlantan AT33 does appear to have clustered her schwas between [u,i], while the schwas of the older AT65 appear to coincide roughly with the median F2 of [u]. These observations suggest—as much as they can, having been derived from only three speakers—that there may have been a trend—but not a rule—in Atlanta among speakers closer to middle age to produce schwa between [u,i], and for the older speakers' schwa to approach [u]. As cautioned both in this section and in Marttila 2015, however, this interpretation would not necessarily imply such intent in schwa quality, as it may simply be the case that the quality of schwa remained stationary while [u] moved back or forward in F2, with no interaction between these two vowels necessarily involved.

[16] See e.g. Labov, Yaeger & Steiner 1972; Feagin 1986 for contemporary observations of this Southern back shift.

[17] It could be suggested that a central target in these data might have derived from a neutral tongue position (cf. Browman & Goldstein 1990), perhaps with assimilation in F1 to the stricture of the surrounding consonants to account for the observed low F1 (cf. Flemming 2004). However, it seems that this view would not be compatible with the finding by Menezes et al. (2015) that schwa in *the* has an acoustic target but lacks a speaker-universal articulatory configuration.

4. Second study

The results from the first study, as reported in Chapter 3, implied a potential merging of schwa to [u] or [ʊ] among some of the Southern speakers, but uncertainty in those findings derived in part from the very small number of vowel tokens per speaker. This second study was undertaken to counter that uncertainty. Specifically, a single speaker was selected from those in the first study who had shown relatively strong overlap between schwa and a full vowel (i.e. a potential merger), after which a larger sample of tokens of schwa and the vowel to which schwa was thought to have merged was collected from that speaker's interview data. Thus, whereas the first study analyzed data from only the first hour of each speaker's interview, the aim in this second study was to analyze the complete available interview of a single speaker. Of the speakers in the first study, AT45 exhibited a relatively high degree of overlap between schwa and [ʊ] (Figure 19 [Appendix]), and was thus chosen as the speaker for this study,[18] in which the following hypothesis *H* was thus to be evaluated.

H: Uttered by AT45, schwa and [ʊ] are of the same distribution in F1 and F2.

4.1 Material and methods

4.1.1 Speaker

According to her interview (LAP 2015, LAGS corpus, code INF097), AT45 is a housewife with a high school education. She was born in Chamblee, Georgia, but moved to Atlanta, Georgia as "a tiny baby," living there until age "about four." She relocated to Avondale Estates, Georgia, where she attended elementary and high school. At the time of the LAGS interview, she lived in Decatur, Georgia (Figure 9). Her parents were born in Georgia, her mother a high school graduate and a housewife, and her father, having attended high school, had worked as a salesman.

In all, AT45's interview as available on the LAP server (LAP 2015) runs for roughly four hours, being split into five ca. 47-minute segments, presumably reflecting the capacity of a single side of the kind of audio tape on which the interview was recorded. It appears by this author's evaluation that all of the interview was conducted on the same day, 17 July 1970, in the informant's home. The overall tone in the interview appears formal in structure but relatively casual in speech.

[18] A similarly strong overlap was seen for AU39 between schwa and [u] (Figure 20 [Appendix]); however, concerns were noted in Section 3.1.2 of [u] tokens being relatively rare in these interviews, raising doubt of whether a reasonably large sample would be obtainable of [u].

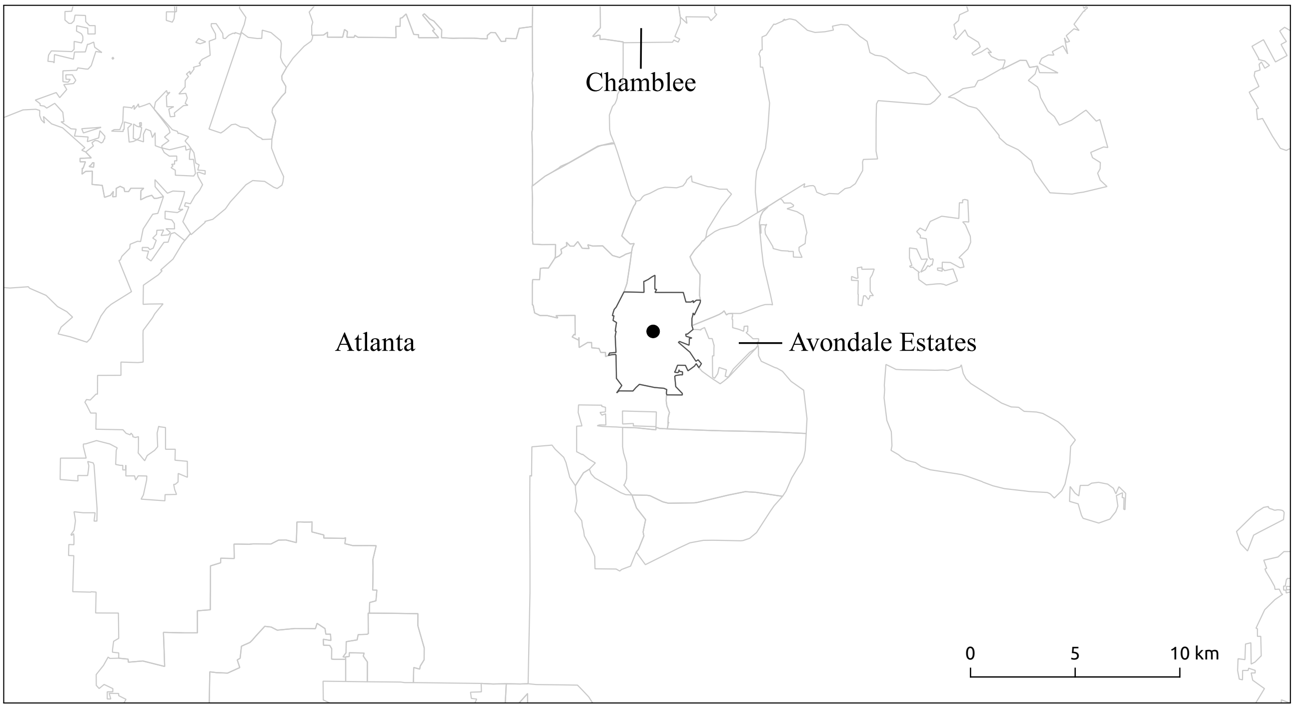

Figure 9. Speaker AT45 (●) in Decatur, Georgia (dark outline); with three other cities of previous known residence also indicated. See Figure 1 for the location of this region within Georgia. Boundaries shown are as of 2013. Base map: USCB 2015.

4.1.2 Extraction of data

As in the first study, the definite article *the* was used as a carrier for schwa, while tokens of [ʊ] were obtained from words in which [ʊ] is typically assumed to occur, e.g. *took*, *good*, and *put*. Neither schwa nor [ʊ] tokens were extracted from disfluent contexts, utterance-initial or utterance-final positions,[19] muted or elevated speech, contexts associated with ambient noises, nor where the token was directly preceded or followed by one of [ɹ,j,l,w,h] or any vowel. Schwa and [ʊ] tokens were also ignored in contexts where they occurred together, e.g. "the good," since it was unclear whether the quality of one would influence that of the other when produced in succession. The analysis did not recycle data from the first study, so e.g. all schwa tokens from the first hour of AT45's interview data were re-measured for this study. However, the full vowel measurements ([i,u,ɑ,æ,ʊ,o,ɔ]) done of AT45 for the first study are used in this chapter in plots of the new data, to provide an idea of the extent of the speaker's vowel space. It is worth noting that these full vowel measurements were obtained only from the first hour of the interview, and may thus not accurately reflect the extent of the vowel space in subsequent hours.

[19] Utterance boundaries were assumed at (short) pauses in the speech stream.

From each utterance of schwa and [ʊ] fitting the above criteria, the following data were extracted. The first two formants (F1 and F2), the vowel's duration, the word in which the vowel occurred, and the phoneme immediately preceding and following the vowel. A description of the methods used was given in Section 3.1.2; though it is to be added here that the formants of [ʊ] were measured from the temporal midpoint if no stable portion was observed on the spectrogram, and otherwise, the formants were obtained from the stable point. The order of interview segments was randomized, but the schwa and [ʊ] tokens were extracted simultaneously. To help manage the entry of data, the SchwaSubmitter 2 utility (described in Section 3.1.2) was adopted in modified form (Figure 10).

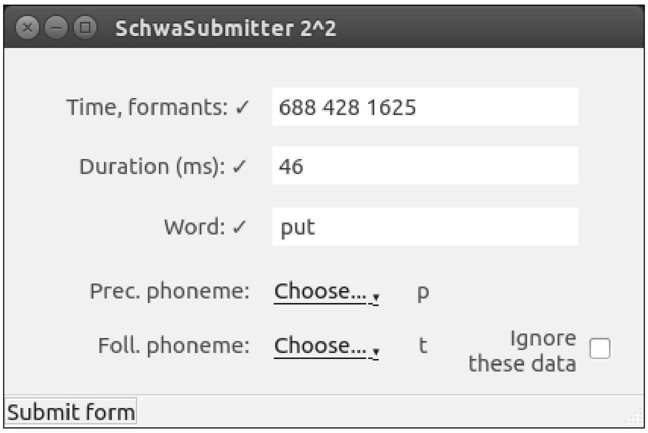

Figure 10. Screenshot of the SchwaSubmitter 2^2 program. The data shown are fictitious.

4.1.3 Full sample and point sample

All told, the four hours of interview audio yielded 111 vowel tokens: 66 [ə] and 45 [ʊ], these tokens together constituting the *full sample*. However, initial inspection of the full sample suggested a need for further partitioning of the dataset prior to analysis, as a potential interaction was found between the durations of [ə,ʊ] and the time at which they occurred in the interview. Figure 11 demonstrates this interaction by plotting the median durations of [ə,ʊ] for each of the five interview segments. In the figure, a general downward trend in duration can be seen with increasing interview time for both [ə,ʊ], with a notable dip for both vowels especially in the third segment.

The cause of this effect was not known; however, it coincided with this author's subjective observation from the interview audio that the speaker may have grown tired as the interview progressed. It is, however, unclear whether such fatigue would involve the shortening of certain vowels, or whether this effect was rather caused e.g. by the speaker having been particularly careful with her pronunciation only at the beginning of the interview. Nonetheless, it appeared that there were potential systematic changes in the speaker's vowel production as a function of interview time, the shortest vowels on average occurring in the middle of the interview, and the longest vowels on average at the beginning. This phenomenon would not be an issue for the analysis had there been a similar number of schwa and [ʊ] tokens from each interview segment, but that was not the case: as seen in Table 4, some recordings yielded notably fewer instances of [ʊ] than [ə], and vice versa. This implies that the full sample may have contained temporal bias.

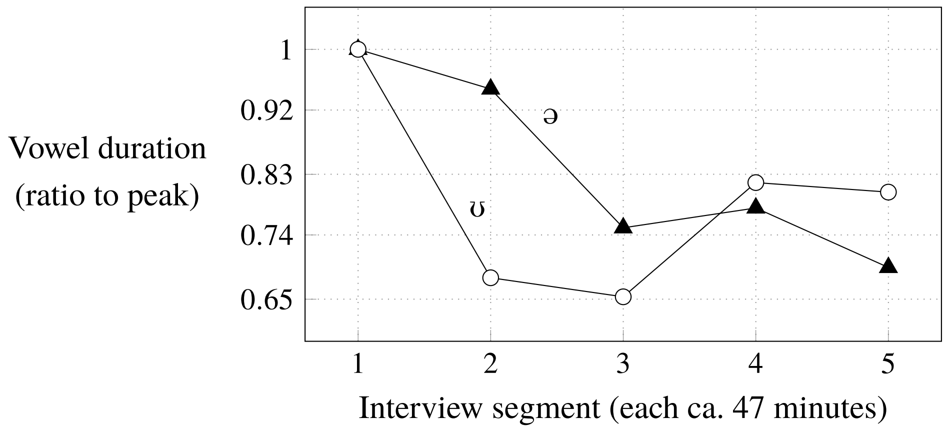

Figure 11. Relationship between the median durations of [ə, ʊ] and the interview segment over which the medians were calculated. Note that the values shown here are ratios to the peak median duration of the given vowel in a segment (peak medians: [ə] = 18 ms, [ʊ] = 75 ms).

	Tokens per segment				
Vowel	1	2	3	4	5
ə	12	20	12	14	8
ʊ	3	9	17	11	5

Table 4. Number of utterances of [ə,ʊ] extracted from each interview segment.

The full sample, despite its potential bias, was submitted to analysis, but the analysis also included a *point sample*, a subset drawn from the 111 tokens of the full sample with the aim to minimize any temporal bias. While the point sample could in theory be formed simply by selecting an equal number of observations of [ə,ʊ] from only one of the interview segments, this would not guarantee that the point sample would contain an ideal number of tokens of each vowel category relative to the period of interview time in which they occur. Instead, the point sample was obtained by treating the interview segments as a continuous stretch of speech, from which such a slice of time was located that contained a reasonable number of [ə,ʊ] tokens in the shortest possible amount of interview time. To locate that time slice, a brute-force algorithm was first used to segment the continuous token stream into all possible slices containing a given number of [ə,ʊ] tokens, this number ranging from 10 to 40 in increments of one. The algorithm is laid out in pseudocode below.

```
for mincount ← 10, mincount ≤ 40 do
    countə ← 0    countʊ ← 0    start ← 0
    while start ≤ {vowels}.n do                                    ▹ loop until reached last vowel
        for end ← start, end ≤ {vowels}.n do
            if {vowels}_end = ə then countə ← countə + 1
            else if {vowels}_end = ʊ then countʊ ← countʊ + 1
            end if
            if countə ≥ mincount and countʊ ≥ mincount then
                {slice} ← Δtime between {vowels}_start and {vowels}_end
                break for
            end if
        end for
        start ← start + 1
    end while
end for
```

The output of the algorithm when applied to the full sample was a list of 1814 time slices, along with the duration of each slice and the number of [ə,ʊ] tokens contained in it. With subjective, a priori judgment, this author selected from that list a time slice which contained a number of [ə,ʊ] tokens thought sufficiently large for statistical inference, but so that it was not so long in duration that notable temporal bias might have been introduced. Ultimately, the slice chosen contained 17 tokens of [ə] and 17 tokens of [ʊ] in a 34-minute period between minutes 123 and 157 of the full interview (i.e. spanning parts of interview segments 3 and 4). Inspection of the token data in this slice, plotted in Figure 12, indicated a fairly even temporal distribution of [ə,ʊ], which suggested that any remaining temporal bias would have been relatively evenly distributed between the vowel categories.

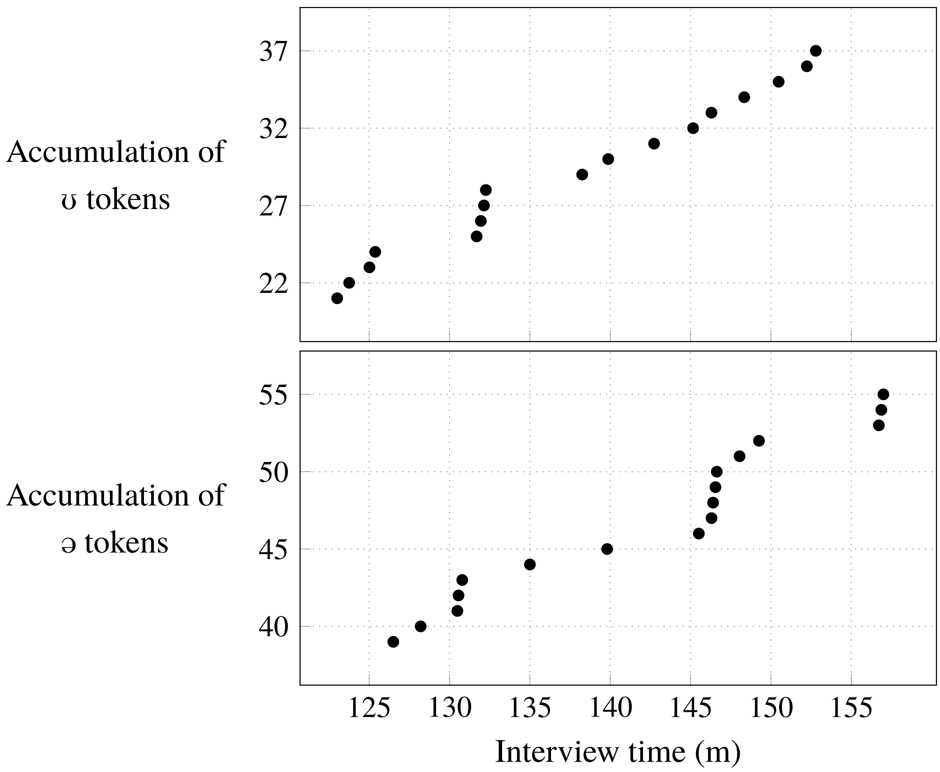

Figure 12. Scatterplots of the accumulation of vowel tokens as a function of interview time in the point sample.

Finally, a brief note must be made of the range of segmental contexts from which the vowel tokens were drawn for analysis. Both the full sample and the point sample were biased such that [ʊ] was more often measured from contexts involving a velar consonant, while schwa more often coincided with a labial (Table 5).

	Coronal	Labial	Velar
Full sample			
ə (*n* = 66)	100%	48%	14%
ʊ (*n* = 45)	80%	29%	67%
Point sample			
ə (*n* = 17)	100%	29%	18%
ʊ (*n* = 17)	82%	18%	76%

Table 5. Segmental contexts of [ə,ʊ] in the full sample and the point sample. The values indicate the percentage of tokens that were either preceded or followed by the given type of consonant.

4.1.4 Methods of analysis

The methods of analysis in this study mirror those in the first study (Section 3.1.3), with some changes. Potential monotonic bivariate correlations were evaluated using Spearman's rho (computed in SPSS). More notably, the *Pillai score* (see Hall-Lew 2010, Nycz & Hall-Lew 2013 and references therein) was used as a quantitative measure of vowel merger. Unlike the Euclidean distance and individual formant comparisons used in the first study, the Pillai score is a multivariate (MANOVA) statistic that can be used to evaluate the degree of combined F1 and F2 overlap between the tokens of two vowel categories. The Pillai score was computed in R (version 3.2.2; as per Hall-Lew 2010:4, 2014), entering F1 and F2 as dependent variables, and vowel category (encoded as a dummy variable) as the sole independent variable.[20]

When interpreting the Pillai score, the following can be kept in mind (cf. Hay, Warren & Drager 2006; Hall-Lew 2010). A low and statistically non-significant Pillai score approaching 0 would imply that vowel category accounted for little of the observed variance in F1 and F2; in other words, that the categories may have merged in F1 and F2. A statistically significant score approaching 1, on the other hand, would imply that the categories are distinct from each other.

To exert some control over the familywise error rate (e.g. Wilcox 2003:407), the Bonferroni correction was applied as a post hoc test. Typically, the Bonferroni correction is defined as $\alpha \div n$, where α is the accepted probability of falsely rejecting the null hypothesis and n is the number of statistical tests conducted, the correction thus aiming to reduce the overall probability that the null hypothesis is falsely rejected in a set of multiple tests. However, in the Pillai test as used here, accepting the null hypothesis (with $p \geq \alpha$) was to imply that the two vowel categories had merged, and so the normal Bonferroni correction would have increased the probability of falsely claiming a merger as one increased the number of Pillai tests conducted. For that reason, the Bonferroni correction was applied as $\alpha \cdot n$, where $\alpha = .05$, as in the first study.

---

[20] The Pillai score accommodates multiple independent variables, allowing one e.g. to control for potential bias in segmental context (Hall-Lew 2010). Although it was noted above (cf. Table 5) that there was indeed segmental bias in the present data, it was also the case that there were no data of schwa preceding anything other than coronal consonants, while the preceding segment was more varied for [ʊ]. For that reason, and for the fact that no data were available of variation in the vocalic context, an attempt at statistical control of context was considered moot. One also notes that a multiple regression with vowel category as the dependent variable and F1 and F2 as independent variables produced the same results as the MANOVA; however, for potential consistency with the literature, the Pillai score was reported here.

4.2 Results

Figure 13 plots the 111 vowel tokens in the dataset, with the left panel showing the full sample and the right panel the point sample, with transcriptional symbols indicating the median full vowel qualities measured of the speaker in the first study. Examining the full sample, some overlap is seen between schwa and [ʊ] in both F1 and F2, though schwa is to some extent drawn toward higher F1 and lower F2 (median F1, F2 = 547 Hz, 1579 Hz) compared to [ʊ] (median F1, F2 = 518 Hz, 1737 Hz). The Pillai score for the full sample is relatively low but nonetheless significant, Pillai = 0.153, F = 9.751, p < .001. Given that labial contexts generally lower vowel F2 (Fowler 1994, Flemming 2009) and that there were more schwa than [ʊ] tokens from labial contexts (Table 5 in Section 4.1.3), the difference in median F2 between schwa and [ʊ] might to an extent result from that segmental bias. Removing at random 18 schwa tokens from the full sample preceding labials, so that 29% of all remaining schwa and [ʊ] tokens were from labial contexts, no notable change is seen in the Pillai score, which remained relatively small but significant, Pillai = 0.136, F = 7.075, p = .001.

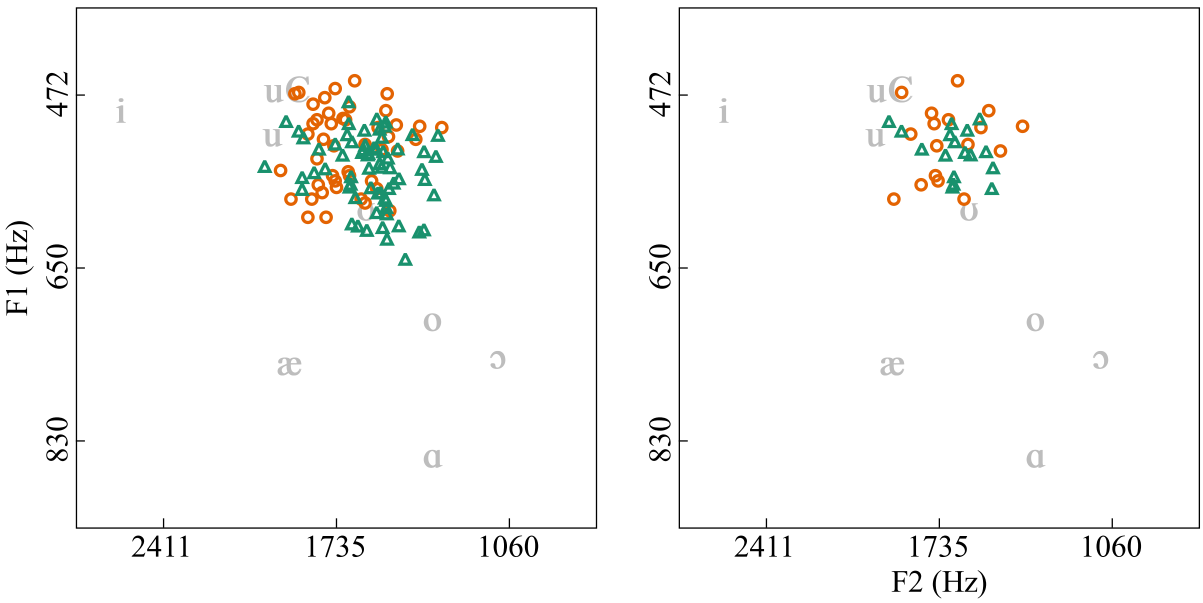

Figure 13. Scatterplots of the distributions of [ʊ] (o) and [ə] (△) as uttered by AT45. The full sample is plotted in the left panel, and the point sample in the right panel. Transcriptional symbols indicate the median locations of the corresponding full vowels produced by the speaker in the first interview segment.

Turning to the right panel in Figure 13, one finds plotted the 17 schwa tokens (median F1, F2 = 530 Hz, 1679 Hz) and 17 [ʊ] tokens (median F1, F2 = 512 Hz, 1743 Hz) of the point sample. Here, the category formant medians are relatively similar, though the distribution of schwa in F1 and F2 is somewhat more tightly clustered than that of [ʊ]; this may reflect the fact that, unlike schwa, [ʊ] was preceded also by consonants other than coronal, being thus obtained from more varied segmental contexts than schwa. Regardless, a low, non-significant Pillai score indicated a merging of schwa and [ʊ] in these data, Pillai = 0.034, $F = 0.547$, $p = .584$. A Bonferroni post hoc test, $3\alpha < .584$, suggested that the Pillai score remained non-signficant when accounting for multiple testing.

On average, [ʊ] was longer in duration in both the full sample (median = 58 ms) and the point sample (median = 52 ms) compared to schwa (15 ms and 14 ms, respectively). Some durational overlap nonetheless existed between the categories in both samples, as shown in Figure 14. Anticipating a discussion in Section 4.3 of the possibility that AT45's [ʊ] may have been reduced to schwa, a comparison is shown in Figure 15 of those [ʊ] in the point sample (left panel) whose duration was greater than 100 ms (range = 157–260 ms, median = 220 ms, $n = 4$) to the rest of the schwa and [ʊ] tokens in the point sample. The long [ʊ] tokens do not appear to deviate notably in F1 and F2 from the shorter [ʊ] nor from schwa. (The words in which the four long [ʊ] occurred are three instances of *good* and one instance of *could*, the token with the lowest F1 being one in *good*.)

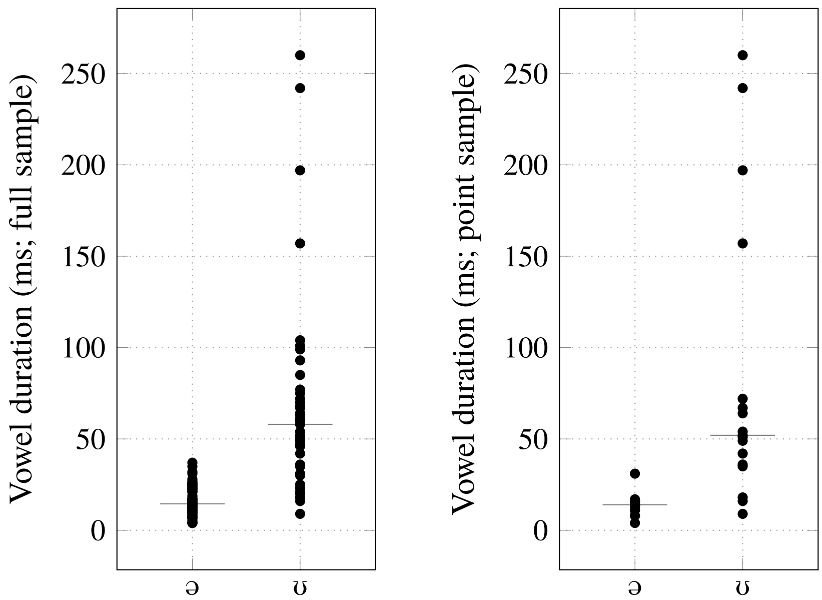

Figure 14. Distributions of the durations of [ə,ʊ] in the full sample and the point sample.

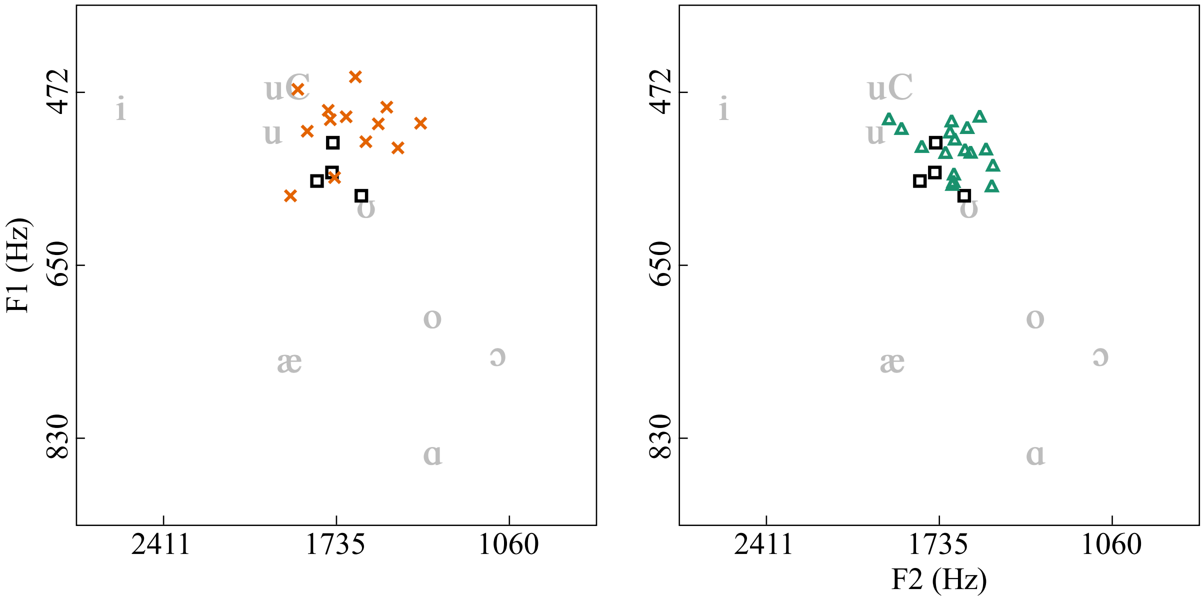

Figure 15. Scatterplots of the distributions of [ʊ] (×) and [ə] (△) as uttered by AT45 in the point sample, with [ʊ] tokens longer than 150 ms (□) indicated separately.

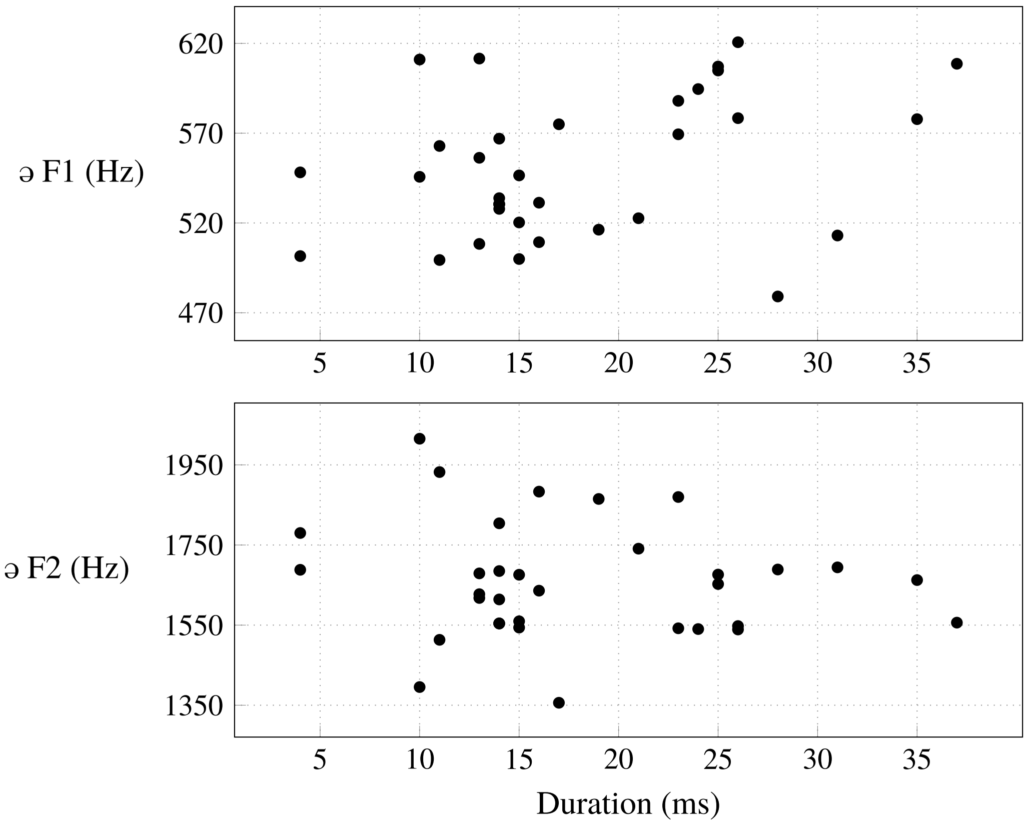

Figure 16. Scatterplots of duration vs F1 and F2 for the schwa tokens in the full sample. These data purposely exclude contexts in which schwa was flanked by a labial consonant.

Recalling from Chapter 2 that Flemming (2009) evoked Lindblom's (1963) undershoot model to explain the speaker's motivation to assimilate short schwa to its segmental context, Figure 16 evaluates whether the formants of schwa and [ʊ] correlate with their durations. The data in the figure are from the full sample but exclude tokens of schwa and [ʊ] that precede or follow labial consonants (as their articulation does not primarily involve the tongue, unlike for coronals and velars). The figure indicates no correlation between duration and F2, but there may be a trend for higher F1 with increasing duration. A Spearman rank correlation test indicated a weak but non-significant correlation between duration and F1, rho = .258, p = .141, n = 34. In the data graphed in Figure 16, median schwa F2 is 1669 Hz, while being 1653 Hz when excluding schwas preceding velars and thus leaving only schwas preceding coronals[21] (n = 25).

4.3 Discussion

A dataset of 66 schwa and 45 [ʊ] tokens as uttered by a speaker of Southern American English (AT45) was analyzed to find whether the speaker had merged schwa and [ʊ] in F1 and F2. In the full sample, the two vowel categories showed a degree of acoustic overlap (left panel in Figure 13), but a statistically significant Pillai score implied that the categories had not fully merged. The Pillai score remained significant when attempting to partially control for segmental bias (cf. Section 4.1.3) by excluding tokens from labial contexts. Though not certain, the full sample may also have been temporally biased, as discussed in Section 4.1.3. Having made an attempt to control for this bias by selecting a smaller number of vowel tokens from a limited section of the interview (the point sample), indication of merger emerged: schwa and [ʊ] in the point sample exhibited considerable acoustic overlap (right panel in Figure 13) and were associated with a low, non-significant Pillai score (0.034), suggesting that the categories had merged in F1 and F2.

It should be emphasized that the results from the point sample do not imply a full acoustic merger (if they imply a merger at all, as discussed below). In the point sample, [ʊ] tokens were generally longer in duration (median = 52 ms) compared to schwa tokens (median = 14 ms), suggesting that if by nothing else, there was a distinction by duration between the categories.[22] The present study also did not evaluate formant transitions, only midpoint formant values; differences in temporal formant

[21] In other words, in these data, preceded by one of [θ,ð] and followed by one of [d,s,n,t,ʃ].

[22] Whether this distinction was intended by the speaker is uncertain, given that function words, e.g. *the*, are more conductive to vowel reduction than the content words from which [ʊ] were extracted (cf. Bell et al. 2009).

dynamics might provide the listener cues for distinguishing between two vowels even should their midpoint formants be similar.

Another line of caution in interpreting these results is with regard to the nature of the [ʊ] tokens. Though the [ʊ] tokens were indeed longer than the schwas on average, there was nonetheless a degree of durational overlap between the categories, as seen in Figure 14. It is also the case that the median duration of [ʊ] in the point sample, 52 ms, is perhaps not considerably different from that of 64 ms observed by Flemming & Johnson (2007, as reported in Flemming 2009) of phonetic schwa. These observations raise, in other words, the suspicion that AT45 may have reduced her [ʊ] to phonetic schwa rather than having merged her phonetic schwa to [ʊ]. As the present study cannot probe the intentions of the speaker, a degree of uncertainty must remain here of whether her [ʊ] had indeed become more like schwa or her schwa more like [ʊ]. One does note, however, that the four longest [ʊ] tokens, > 150 ms, did not appear clearly distinct in F2 from the rest of the category tokens, as shown in Figure 15 (left panel), which may indicate the absense of gradient reduction in [ʊ] toward phonetic schwa with decreasing duration.[23] On the other hand, the speaker may have moved her entire /ʊ/ category toward a more reduced (schwa) quality, which, if she was indeed fatigued as suggested in Section 4.1.3, may not be altogether implausible.

Leaving the discussion of [ʊ] for a moment, it is also useful to consider the quality of schwa by itself, without reference to potential mergers. Since the schwas uttered by AT45 were very short, averaging 15 ms in the full sample, they would in Flemming's (2009) typology (Chapter 2) be considered phonetic schwa, a view supported by Marttila (2015) having observed similar durations of schwa in *the* and having found their F2 trajectories essentially assimilated. Assuming thus that the schwas observed at present were assimilating, one would expect their quality in F2 to reflect a constant influence of the place of articulation of the preceding coronal consonant (e.g. [ð]). The F2 values of coronal loci[24] depend on their segmental context as well as speaker anatomy, but on average, these loci are roughly in the range 1800–2000 Hz (e.g. Fowler 1994; Jongman, Wayland & Wong 2000). With that in mind, and considering that Flemming & Johnson (2007) found phonetic schwa in all-coronal context to have an average F2 of 2042 Hz, it is surprising to note that AT45's schwa averaged only 1653 Hz in F2 when flanked by coronals.

---

[23] Though there may have been greater assimilation in shorter [ʊ] to the stricture of the surrounding consonants, given that the longer [ʊ] tokens appear of somewhat higher F1 compared to the shorter ones.

[24] A locus being a warping in the formant trajectories of surrounding vowels which points toward the place of articulation of the intervening consonant (e.g. Delattre, Liberman & Cooper 1955).

As shown in Figure 16, AT45's schwa tokens do not appear to exhibit correlation between duration and F2 in coronal and velar contexts. This may suggest that her schwas were at all times maximally assimilated to the surrounding articulations (cf. Lindblom 1963, Moon & Lindblom 1994). That in turn might imply that their lower-than-expected average F2 resulted from either the influence of confounders on the consonant loci—e.g. unique speaker anatomy, or a notable overrepresentation of back vowels flanking the consonants (cf. e.g. Fowler 1994)—or the consonants to which schwa assimilated having been articulated as back allophones (cf. Frisch & Wodzinski 2016, who in an ultrasound study found the velar /k/ to have front and back allophones, between which the speakers selected depending on which type of vowel followed). In the case that there is an allophonic inventory of place of consonant articulation available to the speaker, one might grant that the speaker would be able to produce a very short and very assimilated schwa with a specific quality target so long as that target is within reach of the available consonant allophones. In that sense, the present results from AT45 appear acoustically consistent with a schwa assimilating to near-back-articulated consonants.[25]

Besides the question of by which articulations AT45 produced her schwa, which is a question that an acoustic study is not in a position to answer with certainty, another question is, would AT45 have had a particular reason to produce schwa as the somewhat near-back, [ʊ]-coincident quality observed of her here? Although, as noted earlier, this study is not able to probe speaker intent, the proposal can be made that producing schwa as a near-back quality might have been advantageous given the nature of AT45's vowel space. The high-front to high-mid area of her vowel space included not only [i] but the notably fronted [u] and likely [ɪ], as well (cf. vowel plots in Section 4.2). By moving schwa toward a near-back position, AT45 would potentially reduce the number of forms that occur in the high mid-to-front area, thereby possibly increasing contrasts between the vowels that remain there (cf. Manuel 1990, who found an inhibition in vowel coarticulation in the direction of a crowded area of the vowel space). Should the speaker want to limit schwa to a lower F2 to keep it away from the high-front, [ʊ] might provide a stable quality target, whether intentionally or otherwise (with the implication that schwa might not necessarily merge with [ʊ] but perhaps rather be limited to an F2 at most as high as that of [ʊ]).

[25] This kind of targetness in schwa might run counter to the finding by Menezes et al. (2015) that schwa in *the* potentially has no articulatory target. However, Menezes et al. did not posit that no single speaker's schwas had an articulatory target, but rather that the speakers as a group appeared to employ different articulatory strategies to produce what ended up being acoustically fairly similar schwas.

In this regard, there are similarities between AT45 and TN51 (Figure 18 [Appendix]). That is, TN51 produced both a very fronted [u] and a notably backed schwa, suggesting that a backer schwa may indeed be a response to the encroaching of [u] on the high mid-to-front area. On the other hand, AU33 (Figure 20 [Appendix]) and TN82 (Figure 18 [Appendix]) both uttered on average a considerably fronted [u] but not a backed schwa, this implying that there may be no categorical rule regarding the quality of schwa in this situation. It may be, for instance, that schwa is moved back and out of the way of [u] only when the speaker makes an effort to enhance these contrasts, perhaps as a response to the perceived formality of the speech situation, but this remains speculation.

Returning to assimilation and coronal loci, one recalls that the schwas observed by Menezes et al. (2015) of Western speakers in *the* had an average F2 of 1699 Hz. This neither falls within the above-proposed range of 1800–2000 Hz of coronal assimilation, nor agrees with the observation by Flemming & Johnson (2007) of phonetic schwa averaging 2042 Hz in all-coronal contexts when produced by “mainly Western” (2007:86) speakers. Not knowing exactly the segmental contexts from which Menezes et al. obtained their schwas, and thus whether their data may have been biased by such factors, there is nonetheless reason to echo a tentative finding by Gick (2002) that schwa in *the* may be qualitatively different from schwa in content words, such as those observed by Flemming & Johnson. Gick did not expound on this difference, but it may be, as per Section 3.3, that schwa in *the* has an acoustic target—its *underlying form*—at roughly the high-middle of the vowel space.[26]

One can apply for the sake of the argument this hypothetical underlying form to further explain some of the findings made in the first study (Chapter 3). That is, although Section 3.3 noted that the schwas of a number of speakers coincided with the middle of the speaker’s vowel space, and so those schwas may directly reflect the underlying form, it is nonetheless suspicious that some speakers (e.g. TN49 and AU39) exhibited strong overlap between schwa and [u] (see vowel plots in Appendix), and that in general the Tennessean data were clustered about [u,ʊ] (Figure 4). These overlappings may well be coincidental, but one should address the possibility that they are not. To that end, in a hypothetical sense, it does not appear infeasible that as the Southern fronting of back vowels brings [u] closer to schwa, the acoustic similarity of [u] to schwa might act in the way of a magnet, attracting the centrally-located schwa target toward itself and resulting in a merging, either fully or partially, of schwa and [u]. On the other hand, if the underlying form of schwa in *the* is indeed a specific acoustic target, one assumes that the speaker could also intentionally alter its

[26] At least, when preceding consonants (cf. Keating et al. 1994).

quality for various purposes, e.g. to enhance category contrasts, as potentially done by AT45 and TN51, or for reasons of social identification—a schwa interacting with the back vowels being, in other words, a potential marker of southernness; given especially that the Western speakers in the first study appeared to some extent to have avoided such interaction.

A final note about the these results concerns whether trends in data from an interview recorded several decades ago is relevant in the present. The current study was repeated by this author in a more informal manner on an interview of another female Southern speaker, aged 91 years, who had been interviewed in Roswell, Georgia—although having been born in Tennessee—in 2006 as part of the Roswell Voices corpus (LAP 2015, code INF007). On inspection, the speaker was found with a similar arrangement of full vowels as AT45, most notably, having a relatively fronted [u] (median F2 = 1774 Hz, n = 4) and with [ʊ] of somewhat lower F2 (median = 1456 Hz, n = 10) and higher F1 than [u]. Her schwas (median F2 = 1470 Hz, n = 71, in carrier *the*) aligned in F1 and F2 to a similar pattern as those of AT45, i.e. largely coinciding with [ʊ]. The F2 midpoint between her [i,ɔ] was 1738 Hz, notably higher than the F2 of her schwa. Although an informal study and one to be treated cautiosly, it suggests that a near-back schwa, potentially merged to [ʊ], was exhibited in unscripted interview speech still in the near-present. (Though whether solely as a relic in older speech is not clear.)

Having discussed the results of both this (Chapter 4) and the previous study (Chapter 3), and keeping in mind their limitations as discussed, one can address H and Q. H is tentatively accepted on the basis that no statistically significant difference was found between schwa and [ʊ] in the point sample of AT45, thus suggesting a merger of those categories in F1 and F2 in her speech. However, this is indeed a tentative outcome, given that the full sample yielded a significant difference between schwa and [ʊ]. Whether this significant result arose from segmental or some other bias in the full sample and to which the point sample was more resistant is unclear. Further, the phonemic status of [ʊ] was likewise unclear, casting uncertainty on whether a real merger was observed in the point sample or whether [ʊ] had simply reduced to schwa.

Q is likewise answered in the positive, tentatively. It seems that there may be a degree of regional patterning in schwa quality, with Southern informants tending to favor qualities of schwa which coincide with [u] or [ʊ], while such was not observed in general of the Western speakers. The relatively small number of speakers and vowel tokens used make this inference questionable, however. The second study in particular failed to conclusively demonstrate whether the proposed core of this regional patterning, the merging of schwa to certain full qualities in the South, was factual or coincidental.

5. Conclusion

This paper reports on two studies conducted to investigate regionally-based phonetic variation in American English schwa. In total, the two studies analyzed the first two formants of 1196 vowel tokens, among them 499 schwas obtained from the definite article *the*, as produced by 21 speakers from Southern and Western regions of the U.S. The results of the first study suggest that schwa quality may to a degree be dependent on the speaker's regional dialect, at least relative to the rest of the vowel system: the Western informants tended to produce schwa with a quality roughly between [i,u], whereas schwas in the South tended to coincide in quality with [u,ʊ]. The second study, constituting a case investigation of one Southern speaker, found statistical indication of a possible merger between schwa and [ʊ]. However, these results should be considered tentative, as there are a number of uncertainties about their interpretation (as per Sections 3.3 and 4.3).

In general, the results obtained here appear not incompatible with the finding by Menezes et al. (2015) that schwa in *the* may have an acoustic target. The present results extend that hypothesis by suggesting that (a) the acoustic target is located roughly at the middle of the speaker's entire vowel space in F2, and (b) the exact quality of that target may be subject to regional variation. The view expressed by Flemming & Johnson (2007) that very short schwas—such as those observed here—assimilate to their segmental context and thus have no quality target was found potentially moot, given theoretical indication that the speaker could employ allophonic variation in the schwa's segmental context to obtain a target despite assimilation or short duration in the vocalic segment.

This paper has outlined a tentative argument for the existence of regional variation in assimilating schwa, but further research is needed to probe the uncertainties proposed in Sections 3.3 and 4.3 and thereby to solidify or reject this argument. In particular, a perceptual study may be warranted to evaluate regional attitudes to the quality of schwa observed here among the Southerners; that is, whether listeners attach a notion of southernness to schwas which coincide in quality with the back vowels. With that in mind, one also wonders whether schwa in non-Southern but strongly [u]-fronting dialects—e.g. in parts of California—would exhibit patterns of schwa similar to those observed here of Southern informants, given that the Southern fronting of [u] was implicated in some of these observations. The author plans to pursue such and other approaches on this topic.

References

Acoustical Society of America (Eds.) 2010. *Proceedings of meetings on acoustics, Vol. 9*. Baltimore, Maryland: 19–23 April 2010.

Acoustical Society of America (Eds.) 2013. *Proceedings of meetings on acoustics, Vol. 20*. San Francisco, California: 2–6 December 2013.

Adank, P. 2003. Vowel normalization: A perceptual-acoustic study of Dutch vowels. PhD dissertation. Radboud University.

Alfonso, P.J. & Baer T. 1982. Dynamics of vowel articulation. *Language and Speech 25*(2): 151–173.

Baird, S. 1987. LAGS and the 'Southwest' dialect of Texas English. *Annual Meeting of the Linguistic Society of America 1987*. Accessed 19 September 2015. http://eric.ed.gov/?id=ED293374.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M. & Gildea, D. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America 113*(2): 1001–1024.

Bell, A., Jason, M.B., Gregory, M., Girand, C. & Jurafsky, D. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language 60*: 92–111.

Bigham, D. 2008. Dialect contact and accommodation among emerging adults in a university setting. PhD dissertation. University of Texas at Austin.

Boersma, P. & Weenink, D. 2016. Praat: Doing phonetics by computer. Accessed 27 January 2016. http://www.fon.hum.uva.nl/praat/.

Browman, C.P. & Goldstein, L. 1990. 'Targetless' schwa: An articulatory analysis. *Haskins Laboratories Status Report on Speech Research 101/102*: 194–219.

Bybee J. & Hopper P. (Eds.) 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Byrd, D. 1994. Relations of sex and dialect to reduction. *Speech Communication 15*: 39–54.

Crystal, T.H. & House, A.S. 1988. Segmental durations in connected speech signals: Current results. *Journal of the Acoustical Society of America 83*: 1553–1573.

Delattre, P.C., Liberman, A.M. & Cooper, F.S. 1955. Formant transitions and loci as acoustic correlates of place of articulation in American fricatives. *Studia Linguistica 16*: 104–121.

Dumas, B. 1976. Male–female conversational interaction cues: Using data from dialect surveys. *Conference on the Sociology of the Languages of American Women 1976*. Accessed 19 September 2015. http://eric.ed.gov/?id=ED135216.

Ernestus, M. & Warner, N. 2011 (Editorial). An introduction to reduced pronunciation variants. *Journal of Phonetics 39*: 253–260.

Fabricius, A.H., Watt, D. & Johnson, D. 2009. A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics. *Language Variation and Change 21*: 413–435.

Feagin, C. 1986. More evidence for major vowel change in the South. In: G. Sankoff (Ed.) *Diversity and diachrony* 83–96. Amsterdam: John Benjamins.

Ferguson, S.H. & Quené, H. 2014. Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired listeners. *Journal of the Acoustical Society of America 135*(6): 3570–3584.

Flemming, E. 2004. Contrast and perceptual distinctiveness. In: B. Hayes, R. Kirchner & D. Steriade (Eds.) *Phonetically based phonology* 232–276. Basingstoke: Palgrave Macmillan.

Flemming, E. & Johnson, S. 2007. Rosa's roses: Reduced vowels in American English. *Journal of the International Phonetic Association 37*(1): 83–96.

Flemming, E. 2009. The phonetics of schwa vowels. In: D. Minkova (Ed.) *Phonological weakness in English: From old to present-day English* 78–98. Basingstoke: Palgrave Macmillan.

Fowler, C. 1994. Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. *Perception & Psychophysics 55*(6): 597–610.

Fox Tree, J.E. & Clark, H.H. 1997. Pronouncing 'the' as 'thee' to signal problems in speaking. *Cognition 62*: 151–167.

Frisch, S.A. & Wodzinski, S.M. 2016 (In press). Velar–vowel coarticulation in a virtual target model of stop production. *Journal of Phonetics 56*: 52–65.

Gick, B. 2002. An X-ray investigation of pharyngeal constriction in American English schwa. *Phonetica 59*: 38–48.

Hall-Lew, L. 2010. Improved representation of variation in measures of vowel merger. In: Acoustical Society of America (Eds.) *Proceedings of meetings on acoustics, Vol. 9*. Baltimore, Maryland: 19–23 April 2010.

Hall-Lew, L. 2014. CORRECTION: Pillai Scores. Accessed 14 February 2016. http://www.lel.ed.ac.uk/~lhlew/pillai.html.

Hay, J., Warren, P. & Drager, K. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics 34*: 458–484.

Hayes, B., Kirchner, R. & Steriade, D. 2004. *Phonetically based phonology*. Cambridge: Cambridge University Press.

IPA 1999. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press. IPA = International Phonetic Association.

Jongman, A., Wayland, R. & Wong, S. 2000. Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America 108*(3): 1252–1263.

Junqua, J.C. 1993. The Lombard reflex and its role on human listeners and automatic speech recognizers. *Journal of the Acoustical Society of America 93*(1): 510–524.

Jurafsky, D., Bell, A., Gregory, M. & Raymond, W.D. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In: J. Bybee & P. Hopper (Eds.) *Frequency and the emergence of linguistic structure* 229–255. Amsterdam: John Benjamins.

Keating, P.A., Byrd, D., Flemming, E. & Todaka, Y. 1994. Phonetic analyses of word and segment variation using the TIMIT corpus of American English. *Speech Communication 14*: 131–142.

Kenyon J.S. & Knott T.A. 1949. *A pronouncing dictionary of American English*. Springfield, Massachusetts: G. & C. Merriam Company.

Kondo, Y. 1994. Targetless schwa: Is that how we get the impression of stress-timing in English? *Proceedings of the Edinburgh Linguistics Department Conference 1994*: 63–76.

Krapp, G.P. 1919. *The pronunciation of standard English in America*. New York: Oxford University Press.

Kretzschmar, W.A., Jr. & Schneider, E.W. 1996. *Introduction to quantitative analysis of linguistic survey data: An atlas by the numbers*. Empirical Linguistics. Thousand Oaks, NJ: Sage.

Kurath, H. & McDavid, R., Jr. 1961. *The pronunciation of English in the Atlantic States*. Ann Arbor: University of Michigan Press.

Labov, W., Yaeger, M. & Steiner, R. 1972. *A quantitative study of sound change in progress, Vol. 2*. Philadelphia: U.S. Regional Survey.

Labov, W. 2001. *Principles of linguistic change: Social factors*. Language in Society 29. Oxford, Massachusetts: Blackwell.

Labov, W., Ash, S. & Boberg, C. 2006. *The atlas of North American English: Phonetics, phonology and sound change*. Berlin: Mouton de Gruyter.

LAP. 2015. LAP Online – Atlas Projects. Accessed 25 April 2015. http://www.lap.uga.edu/Site/Atlas_Projects.html. LAP = Linguistic Atlas Project.

Lass, R. 2009. On schwa: Synchronic prelude and historical fugue. In: Minkova (Ed.) *Phonological weakness in English: From old to present-day English* 47–77. Basingstoke: Palgrave Macmillan.

Lindblom, B. 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America 35*(11): 1773–1781.

Lobanov, B. 1971. Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America 49*(2B): 606–608.

Manuel, S.Y. 1990. The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Haskins Laboratories Status Report on Speech Research 104*: 1–20.

Marttila, O.P. 2015. Acoustic evidence for age-grading in Southern American English schwa. MA thesis. University of Turku.

McDavid R., Jr. 1974. Field procedures: Instructions for investigators, Linguistic Atlas of the Gulf States. In L. Pederson, R. McDavid, Jr., C.F. Foster & C.E. Billiard (Eds.) *A manual for dialect research in the southern states* 35–60. Tuscaloosa, Alabama: University of Alabama Press.

Menezes, C., Moote, K., Garon, A., Baker, J., Lucarelli, M., Nichols, K. & Pelfrey, B. 2015. Articulatory and acoustic correlates of the mid-central vowel. Accessed 4 February 2015. http://www.icphs2015.info/pdfs/Papers/ICPHS0505.pdf.

Minkova, D. (Ed.) 2009. *Phonological weakness in English: From old to present-day English*. Palgrave Studies in Language History and Language Change. Basingstoke: Palgrave Macmillan.

Moon, S.J. & Lindblom, B. 1994. Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America 96*(1): 40–55.

Nelson, D. (Ed.) 2002. *Leeds working papers in linguistics and phonetics 9*. Leeds: University of Leeds.

Nycz, J. & Hall-Lew, L. 2013. Best practices in measuring vowel merger. In: Acoustical Society of America (Eds.) *Proceedings of meetings on acoustics, Vol. 20*.

Pederson, L., McDavid, R., Jr., Foster C.F. & Billiard C.E. 1974. *A manual for dialect research in the southern states* (2nd ed.). Tuscaloosa, Alabama: University of Alabama Press.

Sankoff, D. (Ed.) 1986. *Diversity and diachrony*. Amsterdam: John Benjamins.

Schulman, R. 1989 (Abstract). Articulatory dynamics of loud and normal speech. *Journal of the Acoustical Society of America 85*(1): 295–312. Accessed 12 October 2015. http://dx.doi.org/10.1121/1.397737.

Underwood, G.N. 1974. American English dialectology: Alternatives for the southwest. *Linguistics 128*(12): 19–40.

USCB, 2015. TIGER/Line® Shapefiles. Accessed 25 February 2016. http://www.census.gov/geo/maps-data/data/tiger-line.html. USCB = United States Census Bureau.

USGS, 2015. One million-scale data download. Accessed 16 April 2015. http://nationalmap.gov/small_scale/atlasftp-1m.html. USGS = United States Geological Survey.

van Bergem, D.R. 1994. A model of coarticulatory effects on the schwa. *Speech Communication 14*: 143–162.

Watt, D. & Fabricius, A. 2002. Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F_1~F_2 plane. In: D. Nelson (Ed.) *Leeds working papers in linguistics and phonetics 9* 159–173. Leeds: University of Leeds.

Wilcox, R.R. 2003. *Applying Contemporary Statistical Techniques*. Los Angeles: Academic Press.

Yao, Y. 2011. The effects of phonological neighborhoods on pronunciation variation in conversational speech. PhD dissertation. University of California, Berkeley.

Appendix

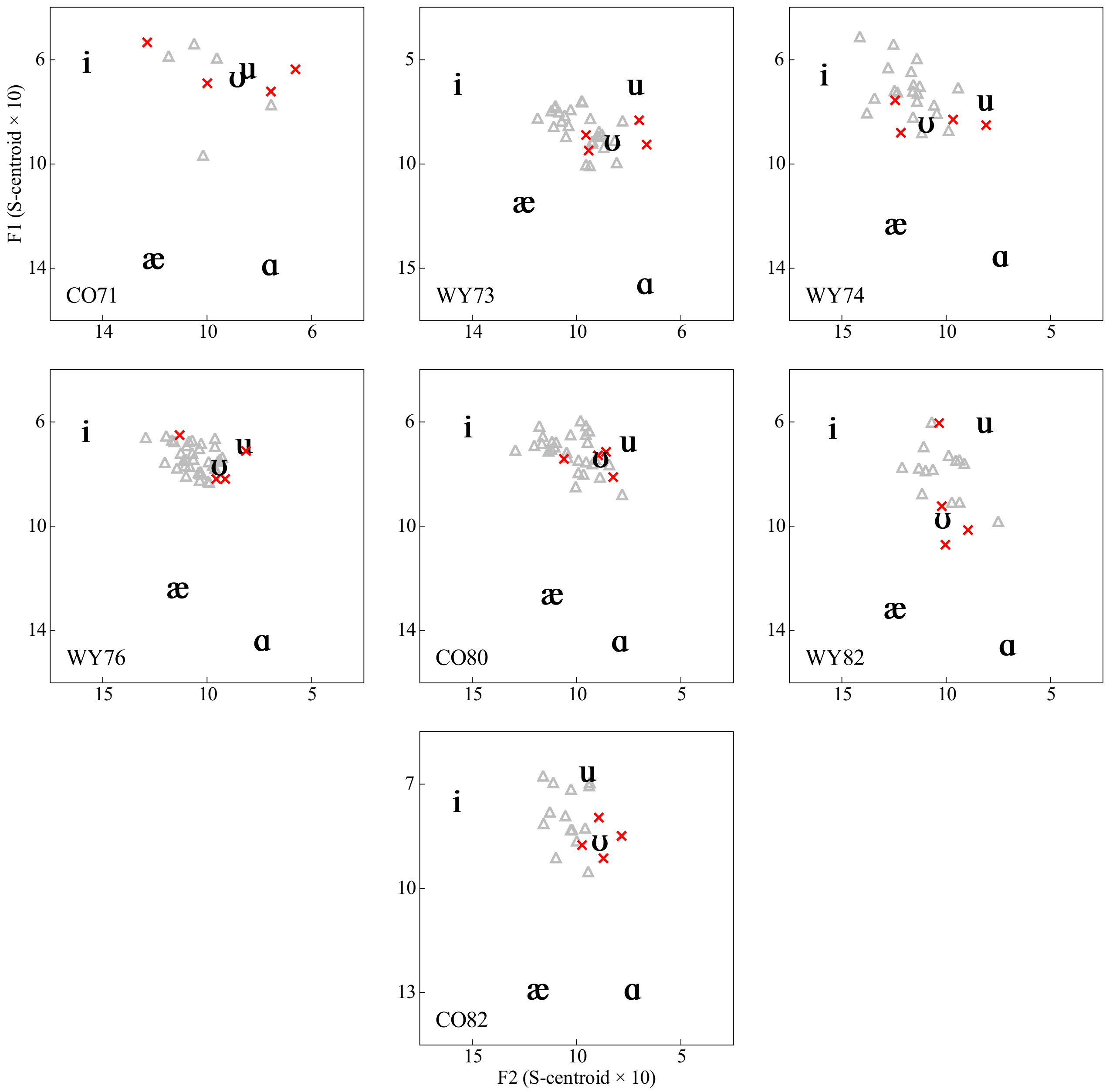

Figure 17. Schwas (△) uttered by the Western informants. Also indicated are the individual [ʊ] tokens (×) of each speaker.

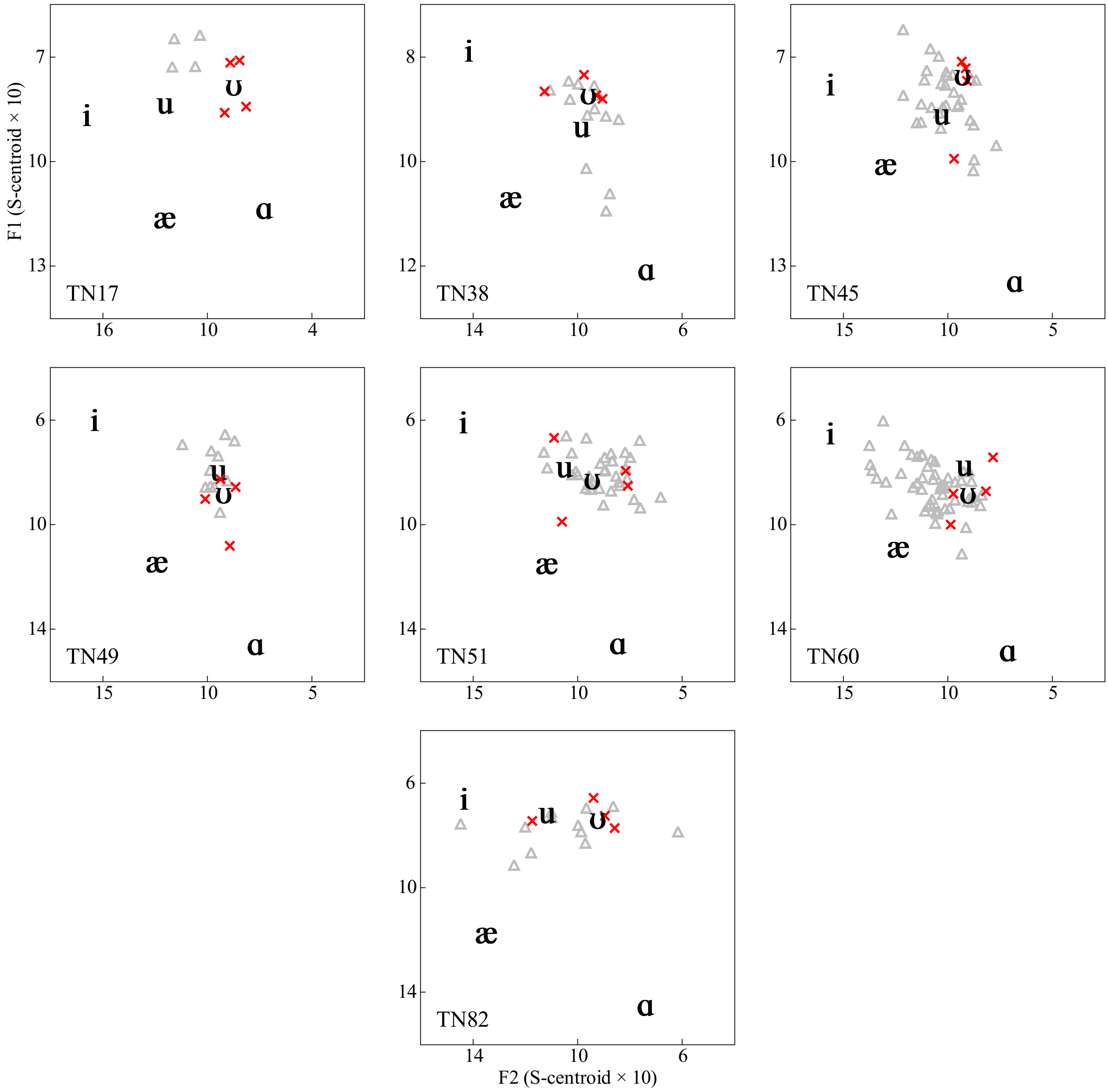

Figure 18. Schwas (△) uttered by the Tennesseans. Also indicated are the individual [ʊ] tokens (×) of each speaker.

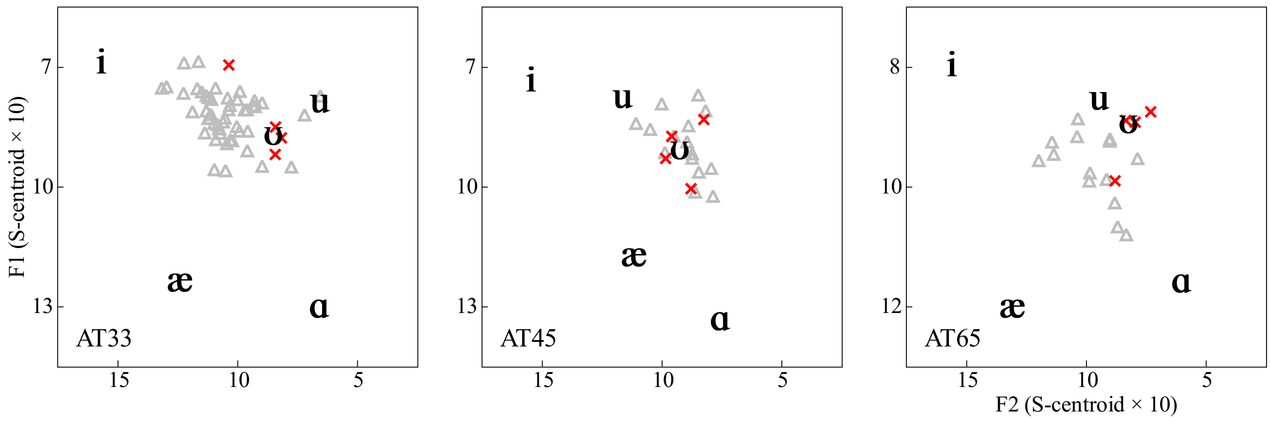

Figure 19. Schwas (△) uttered by the Atlantans. Also indicated are the individual [ʊ] tokens (×) of each speaker.

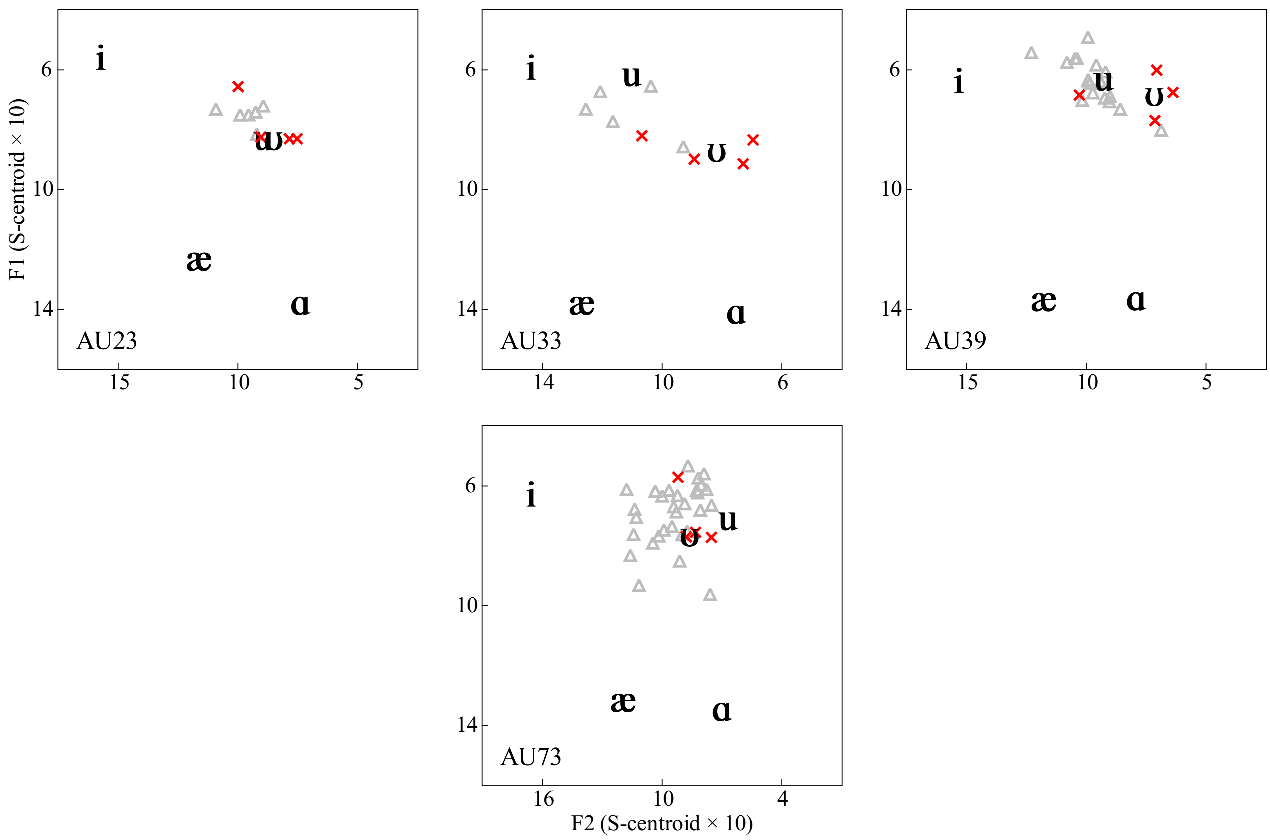

Figure 20. Schwas (△) uttered by the Augustans. Also indicated are the individual [ʊ] tokens (×) of each speaker.

	Words from which [u] tokens were extracted					
Speaker	1	2	3	4	F1	F2
AT45	moved	moved			514	1983
AT65	moved	moved	moved	spoon	403	1479
AT33	moved	who	movable	move	450	1135
AU39	moving	moved			436	1648
AU73	who	move	move		470	1178
AU33	moving	food	goobers	gooseberries	442	1921
AU23	hoop	hooves	hoop	coop	555	1666
TN51	who	moved	coon's	who	469	1770
TN38	moved	who	movable	food	557	1555
TN82	food	moved	moved		511	1900
TN49	hooves	smooching	whooping	whooping	489	1557
TN45	moved	food	spoon	move	577	1744
TN60	moved	food	who		489	1592
TN17	neighborhood	moving			581	2045
CO71	moved	moved			443	1386
CO82	who	moved	smooth	boots	367	1560
CO80	moved	moved	moved	boots	431	1402
WY74	moved	fourche	fourche	moved	449	1407
WY76	beautiful	moved	moved	food	447	1379
WY73	moved	hewes			363	1423
WY82	moose	moved	moved	moved	391	1397

Table 6. Words from which [u] tokens were extracted for each speaker. The median formants of those tokens are also shown.