Language & Communication, Vol. 12, No. 3/4, pp. 267–292, 1992.
Printed in Great Britain.

0271–5309/92 $5.00 + 0.00

HYPERCORRECTION IN RESPONSE TO THE APPARENT MERGER OF (ɔ) AND (ɑ) IN UTAH ENGLISH

MARIANNA DI PAOLO

Although (ɔ) and (ɑ) are thought to be merged in Utah English, acoustic analysis reveals that there are small, but consistent differences between them unless speakers are made aware of the vowels. Evidence that when the vowels are merged, hypercorrection is involved comes from a matched guise experiment. Listeners rated speakers more favorably on factors associated with Standard English when they produced more of a merger or when they gave no clue about whether they merged than when the speakers more clearly distinguished the vowels.

Introduction

A number of studies in recent years have dispelled the notion that sounds which speakers believe to be merged are always, in fact, merged. In particular, Labov *et al.* (1972), Johnson (1989), and Moonwomon (1987) show that (ɔ) and (ɑ) in central Pennsylvania, Ohio, and California, respectively, are still distinct, although close, in phonetic space in spontaneous speech.[1] Moreover, the first two of these studies gave evidence that the speakers had a different target in their conversational speech than they did when they reflected on their speech. Labov *et al.* reported the case of an elderly man who distinguished these vowels in spontaneous speech with a 'wide margin of security'. However, in a minimal pair task, the man claimed that they were the same although he continued to differentiate them minimally. In a more strictly controlled set of tasks, Johnson recorded speakers reading a list of /hVd/ words before asking them to perform a task called a 'direct estimation of vowel prototypes'. The latter task required them to adjust what were essentially F1/F2 parameters of the neutral vowel of a synthetic speech token of a /hVd/ word to those for the vowel of a target /hVd/ word. In the first task, the participants' (ɔ) and (ɑ) were distinct, but the second task showed that their monitored targets were ones containing overlapping formant values for these vowels.

Both Labov *et al.* and Johnson's experimental conditions caused speakers to shift from an unmerged system in a relatively unmonitored context to a merged system in a monitored context. Style-shifting such as this is expected in cases in which one norm is more prestigious. Two different types of prestige which may trigger a style shift that overshoots the norm have been reported in the literature. (See reviews in Yaeger, 1974, and Trudgill, 1984.) First, 'hypercorrection' or orientation to 'overt prestige' refer to a situation in which speakers' style shift is triggered by their orientation to the standard dialect. Second, hyperaccommodation or orientation to 'covert prestige' is a type of style shift due to speakers' accommodation to a local variety. Because the idea of overshooting the target (either in pronunciation or in self-reports) for a particular norm

Correspondence relating to this paper should be addressed to Dr Marianna Di Paolo, Linguistics Program, Stewart Bldg 213, University of Utah, Salt Lake City, UT 84112, U.S.A.

is more clearly conveyed by the terms hypercorrection and hyperaccommodation, this article will generally make use of these terms. Although Labov *et al.* speculated that the prestige of an emerging norm may have affected the speech of the Pennsylvania man they discuss, neither Labov *et al.* nor Johnson gave systematic consideration to the problem from the point of view of the speakers' orientation or attitudes towards the local variety or towards the standard variety.

The current study systematically examines similar phenomena from Utah English. Three types of evidence will be discussed. First, indirect evidence from a categorization task suggests that Salt Lake Valley natives believe that (ɔ) and (ɑ) are not distinct vowels. Second, acoustic analysis shows that speakers of Utah English produce distinct pronunciations (based on F1/F2, phonation type and duration) of the apparently merged pair (ɔ) and (ɑ) under some conditions and less distinct or merged pronunciations under other conditions. Third a matched guise study employing 'merged' and 'unmerged' pronunciations of (ɔ) and (ɑ) shows that the listeners consider varieties of English which have little or no acoustic contrast between (ɔ) and (ɑ) to be more standard than those with a large contrast.

On the basis of the findings of these results, I argue that the neutralization of the F1/F2 distinction in some highly monitored contexts is due to speakers' perceptions that in Standard English the vowels are not distinct. That is, I argue that speakers hypercorrect because they believe that the more phonetically similar pronunciations of (ɔ) and (ɑ) are more standard, i.e., overtly prestigious, than more distinct pronunciations are. Because an F1/F2 merger is not a feature of their unreflected speech, even in some formal contexts, and because the majority of non-local 'Standard' English that they hear (for example, the English of network news anchors, newsbroadcasterese) generally preserves a distinction between (ɔ) and (ɑ), Utah speakers can be said to be hypercorrecting towards a 'false' standard.

In this study I will be limiting myself to (ɔ) and (ɑ) in environments other than before /r/. The pre- /r/ vowels may be involved in a reversal in some parts of Utah (Labov *et al.*, 1972; Yaeger, 1974), which may be similar to the changes reported by Walsh and Mote (1974) for the Delmarva area; Stanley (1936), Norman (1971), and Rulon (1980) for Texas (discussed in detail in Di Paolo, 1986); and more recently in Thomas and Bailey (1991) also for Texas.

Procedures

This study presents three types of data. The first is participants' remarks made during a vowel categorization task. The second is acoustic measurements of word lists, spontaneous speech, and reading passages. The third set of data is the results of a language attitudes task.

Vowel categorization

The first set of data, which comes from the Intermountain Language Survey, consists of the reactions of 122 of the 149 subjects who completed the vowel categorization task shown in Fig. 1.[2] The subjects were asked to read the words in the boxes, then the words in the word list at the bottom of the page and, finally, they were to write the words at the bottom of the page in the boxes, matching them for vowels sounds. In helping the subjects focus on the matching task, the interviewer gave them directions which included among other things to 'Notice that all three words in any one box have the same vowel sound.' For this present study, I will consider only the speakers' reactions to the words in box 9, a set of words which the interviewers were instructed to ignore unless the subject brought them up. Since the interviewers were all young females who had lived most of their lives in the western United States, it was not difficult for them to be unpreturbed by the box 9 words while reading the statement quoted above.

1. meet, speed, Pete	2. mitt, spit, mid
3. mate, maid, jade	4. met, bed, net
5. mat, lag, bad	6. mood, hoot, food
7. could, book, hood	8. mope, hope, toad
9. stalk, stock, caught	10. mud, buck, but

pad	stain	pull	bowl	tail
cool	sale	span	fill	ten
should	pal	school	dune	Hal
gal	hole	bean	full	
said	skill	chewed	deal	
hid	bell	skull	bead	
bang	pin	pool	heel	
feel	soul	laid	sink	
bone	pale	sell	code	

Fig. 1. Vowel categorization task.

Production data

Data collection. Three word lists were used in this study. The first and second of these, the Balanced Word List, Table 1, and the Homophones Word List, Table 2, were originally constructed for a study on the neutralization of tense/lax vowels before tautosyllabic dark /l/. The first list has only eight words containing either /ɔ/ or /ɑ/ and 32 words with other vowels. The eight relevant words are marked on the table with asterisks. Since two of these words, *Paully* and *Polly*, are polysyllabic, they will not be discussed in this paper. The second list contains only 11 /ɔ/ or /ɑ/ words and 44 words with other vowels. The third list, the Minimal Pair Word List, Table 3, has 20 words with

/ɔ/ or /ɑ/ and only 26 words with other vowels.[3] This word list was constructed to obtain more data on the low back vowels in a wider range of phonetic environments. The words in the three lists were presented in random order.

Table 1. Balanced Word List

heed	heel	peep	peel
hid	hill	pip	pill
hayed	hail	cape	pale
head	hell	pep	pell
had	Hal	pap	pal
			*Polly
			*Paully
*hod		*cock	
*hawed	*haul	*caulk	*Paul
HUD	hull	pup	cull
hood	full	cook	pull
food	fool	poop	pool

Table 2. Homophones Word List

heed	heel	peep	peel
he'd	he'll		peal
	heal		
hid	hill	pip	pill
hayed	hail	cake	pale
	hale		pail
head	hell	pep	pell
had	Hal	pap	pal
*cod	*Col	*pop	*pol
	*Sol		
*hawed	*haul	*talk	*Paul
	*hall		*pall
HUD	hull	pup	cull
hoed	hole	pope	pole
	whole		poll
hood	full	cook	pull
food	fool	poop	pool

Table 3. Minimal Pair Word List

*dawn	WAC	*Paul
*gaul	*hock	cull
*cot	buck	*gol
hood	heed	had
hash	*walk	gull
pal	*hawed	Dan
hag	hug	*long
done	cut	hush
*Tosh	*wok	*cog
cat	*caught	*wash
*hod	food	dung
dang	*hawk	*hog
hack	hid	huck
head	hayed	*pol (as politician)
*Don	*dong	HUD
		gal

Two speakers, BW and NM, read the Balanced Word List and the Minimal Pair Word List six times at a slow rate pausing between the second and third readings and again between the fourth and fifth readings to answer short questions about their early schooling. Both speakers read the Balanced Word List in August of 1989. BW's follow-up reading of the minimal pair list took place about 3 months later; NM read the second word list approximately 1 year later. BW is a 30-year-old male who was raised in a west side suburb of Salt Lake City. NM is a 19-year-old female raised on the east side of Salt Lake City. Both of these speakers were university students. Twelve Utah speakers read the Homophones Word List eight times. The speech of only one of these speakers, who I will call Utah Rural, UR, will be discussed here.[4]

In order to test the hypothesis that speakers distinguished the low vowel word classes in conversation, all of the fully stressed monosyllabic words containing (ɔ) or (ɑ) were measured from the Intermountain Language Survey (ImLS) interviews with subjects 192A and 251.[5] Speaker 192A is a talkative upper middle class, White male from the affluent east side of the Salt Lake Valley. Speaker 251 is a middle class White female from the predominantly blue-collar west side. The speakers were 45 and 16, respectively, at the time of the interview in the summer of 1986 and had both lived in the Valley all of their lives.

Acoustic parameters. Three different acoustic parameters were measured on the (ɔ–ɑ) tokens, formants, phonation type (VQI), and duration. Phonation type (laryngeal differences perceived as relative breathiness or creakiness) and duration, which have been found to be important secondary features of vowels, were also measured in order to test the hypothesis that in Utah English (ɔ) and (ɑ) are in a relationship like that of a tense/lax pair of vowels.[6] In previous studies by McRoberts and Faber (1989), Di Paolo *et al.* (1989), and Di Paolo and Faber (1990) laryngeal differences were shown to be a correlate of the tense/lax distinction in English in both production and perception. These laryngeal differences, also called phonation types, range along a continuum from breathy to creaky voice; tense vowels are more breathy than their lax counterparts. We measure phonation type in terms of Voice Quality Index, calculated by subtracting the amplitude of the strongest harmonic in the region of F1 from the amplitude of the fundamental frequency. The result for a breathy vowel will be a larger (a more positive) number, while that for a creaky vowel will be a smaller (a more negative) number. The emerging consensus is that an absolute value of two dB or larger is a perceptible difference (Di Paolo and Faber, 1990; Marie Huffman and Ken Stevens (Pers. Commun. to Alice Faber)).

In terms of durational differences, the prediction is that the (ɔ) word in the pair is longer than the (ɑ) word, the normal distribution for tense/lax vowel pairs. Duration was measured only for the Minimal Pair Word List and for NM's spontaneous speech. Duration differences of 25 msec or over are referred to as 'just-noticeable differences' or JNDs because in many cases 25 msec duration differences have been found to be just at the threshold of human perception (Klatt, 1976).

Acoustic analysis. Acoustic analysis was performed using MacSpeechLab I at the Department of Communication Disorders of the University of Utah, the Interactive

Laboratory System (ILS) at Haskins Laboratories, or Signalyze. All of the tokens from a given set of words were analyzed on a single system.

All of the relevant tokens from the Minimal Pair Word List and the conversational data of 192A and 251 were analyzed using MacSpeechLab I. With this system, words were digitized at a 10 kHz sampling rate with 8 bits quantization and, for NM, 192A, and 251 amplitude normalization. Formant values for the vowels of NM and BW were taken at two points, one at a position after the transition of the initial consonant and the other before the transition of the final consonant, which generally corresponded to 52 msec from the beginning and 52 msec from the end of the vowel nucleus, respectively.[7] The measurement points were determined from screen displays of wide-band digital spectrograms. Because most of the tokens from the conversational data were too short for both measurements, only the first measurement was made for the majority of these tokens. Measurements of F1 and F2 were made on a screen display of a narrow-band digital spectrogram. Six tokens of six minimal pairs, *caught–cot*, *hawed–hod*, *dawn–don*, *hawk–hock*, *walk–wok*, and *Paul–pol* were analyzed for NM and BW. In addition, six tokens of *Gaul–gol*, were analyzed for BW.

The last five of the six tokens for each word of the Balanced Word List (the first tokens were not measured) and all eight tokens from the Homophones Word List were analyzed using the ILS. Signals destined for analysis with this system were lowpass filtered at 5 kHz and digitized at a sampling rate of 10 kHz with 12 bits quantization and pre-emphasis. Each word was analyzed using linear predictive coding (ILS subprogram API), with frames of 25.6 msec overlapping by 6.4 msec. The default of 14 coefficients was used for most tokens; this number was increased to 16 or 18 if the original analysis did not separate between two nearby formants, or decreased to 10 if the original analysis showed a low-amplitude nasal formant. Spectral peaks were identified with the root-solving subprogram RSO, generally using a 500 Hz maximum bandwidth. On the basis of a screen display of a pseudo-spectrogram, two frames were selected, one near the beginning of the vowel nucleus and one near the end, and F1 and F2 values for those frames were recorded. The measurements at Haskins Laboratories were made by Alice Faber as part of our study on tense and lax vowels before /l/.

An acoustic analysis of NM's spontaneous speech during the Minimal Pair Word List taping and of the (ɔ) and (ɑ) words from the (ɔ–ɑ) reading passages for the two speakers from Utah in the attitude study described below was performed using Signalyze. The measurements for the reading passages were made on the original, unaltered tapes. The words were digitized at an 8 kHz sampling rate with 16 bits quantization using the Audiomedia system. Because the vast majority of tokens from the reading passages were too short for both measurements, only one measurement was made for each token according to the guidelines in Labov *et al.* (1972, p. 29). Decisions concerning the point at which to measure were made on the basis of a screen display of a wide-band digital spectrogram. Measurements of F0 were made on a screen display of narrow-band spectra. Measurements of F1 and F2, and amplitude of F0, F1 and F2 were made on a screen display of narrow-band spectra. A differentiating pre-emphasis was applied prior to spectral analysis.

Attitude study

A matched guise technique was used to assess the relative prestige of Utah, northernized

Texas, and Philadelphia varieties of English by asking listeners to judge stretches of tape-recorded speech on the bipolar adjectives listed in the first two categories and by the questions in the third (adopted from Luhman, 1990) in Table 4. Although the matched guise technique was originally developed to test language attitudes in bilingual situations (Lambert *et al.*, 1960), it was subsequently employed to assess the attitudes of listeners towards various dialects of their language. (See, for example, the recent studies by Edwards and Jacobsen, 1987, and Luhman, 1990.) Matched guise attitude studies are by definition ones in which the listeners are led to believe that each voice they hear belongs to a different speaker. The speakers are told to or induced to produce the speech in two guises, two different language varieties. The listeners are not told that they may hear a speaker more than once because the task is designed to separate particular pronunciations of words from the actual humans that use them in order to understand what characteristics are attributed to the speech itself. The same speaker is heard more than once to control for aspects of his speech such as tone of voice and pitch that cannot be easily manipulated.

Table 4. Bipolar adjectives and factors

Personality:	Dependable
	Friendly
	Good
	Honest
	Sympathetic
	Trustworthy
Success:	Ambitious
	Educated
	Intelligent
	Successful
	Wealthy
Speech:	Change speech*
	Similar speech†

*'Would you be interested in changing your speech to sound more like this person?'
†'How similar is this person's speech to that of your speech and your family's?'

The 13 items listed in Table 4 are intended to ascertain the degree to which the listener judges the voices to have qualities of standard speech as well as local speech. The study tests the following two hypotheses that follow from the claim that the merger of (ɔ) and (ɑ) carries overt prestige and that, therefore, a shift towards the merger is hypercorrection (a shift towards a false standard):

(1) Listener/judges believe that a speaker is more successful, has better speech, and a more favorable personality when he is heard making less of a distinction between (ɔ) and (ɑ) than when he makes more of a distinction.

(2) Listener/judges believe that a speaker is more successful, has better speech, and a more favorable personality when he gives no cue about whether he makes a distinction than when he is heard making a distinction between (ɔ) and (ɑ).

The attitudes study also tests two other related hypotheses which follow from the claim that speakers shift to a more merged pronunciation of (ɔ) and (ɑ) to hyper-accommodate to a local norm.

(3) Listener/judges believe that a speaker has more positive traits, but is not thought to be more successful or have better speech, when he is heard making less of a distinction between (ɔ) and (ɑ) than when he makes more of a distinction.

(4) Listener/judges believe that a speaker has more positive personality traits, but is not thought to be more successful or have better speech, when he gives no cue about whether he makes a distinction than when he is heard making a distinction between (ɔ) and (ɑ).

Although a listener may also attribute positive personality traits to a speaker who is thought to be using a standard variety, in the absence of the listener also believing that the speaker is successful and speaks well, it is more likely that the listener thinks that the speaker is exhibiting local, but not standard, norms.

The listener/judges heard 11 voices read one of two short passages each averaging 1:46 minutes. Four men actually served as speakers on the tapes although the listeners were allowed to believe that there were 11 speakers in all. Two Utah speakers represent the merger area for (ɔ) and (ɑ) variables. The other two speakers were intentionally chosen because they were natives of areas that have an unambiguous distinction in the variables but that differ on the particular phonetic quality of (ɔ).[8] These last two speakers are 'Philadelphia' (Ph) and 'General United States' (GenUS).

The two native Utahns will be referred to as Utah Rural (UR), a native of central Utah, and Utah Urban (UU), a native of the Salt Lake Valley. Impressionistically, the most noticeable differences between their readings is audible creakiness in UR's speech and breathiness in UU's speech. UR also tends to diphthongize some monophthongal vowels in a manner similar to speakers of Southern American English. A number of listener/judges guessed that UR was either a rural speaker or from Texas.

Ph is a native of Philadelphia but had lived in Salt Lake City for over 20 years at the time of the taping. His (ɔ) sounds higher and more rounded than GenUS's. Ph's readings included only one pronunciation that would cue a very knowledgeable listener to his origins, the pronunciation of the word *inherited*. Ph pronounces this word with [ʌ] in the nucleus of the second syllable. Most other linguistic cues about his native area are non-determinant because the reading passages were constructed to avoid salient variables such as (æ), which is noticeably raised in some phonological contexts in Philadelphia English.

GenUS is the most difficult speaker to classify. Impressionistically, his speech on tape contains no segmental features that unambiguously point to his native area. But consideration of his intonation in the story-telling context of the readings may cue a careful listener that he is a native of a variety of Southern American English. Before moving to San Antonio, Texas, at the age of 11, he had lived in other university communities in Liberty, Missouri; Louisville, Kentucky; Durham, North Carolina; and Princeton, New Jersey. His educated parents conscientiously taught him about educated English.

The speakers' ages span a 20-year period from about 45 (Ph) to 25 (UU). In their educational background they range from having completed a doctorate (Ph) to having attended a university for 1 year (GenUS). All four had a background in linguistics.

I did not choose actors to read the passages because the results of the phonetic analysis of the data set examined, coupled with observations of Utahns, led me to believe that the subtle, unmerged distinction could not be consciously controlled. However, three of the four speakers have experience in public speaking. UR and Ph are university faculty members and GenUS had experience as a television news broadcaster.

The four speakers read two narratives, the first containing a large number of (ɔ) and (ɑ) words and the second with the same (ɑ) words but no (ɔ) words. Both versions of the reading related a first-person account told from the point of view of a fashion-conscious businessman who pursued a powerful millionaire to his farm, and weekend retreat, in order to contract a deal with him. Each of the four speakers was asked to read the first version to themselves and then to read the passage out loud twice. These readings were meant to elicit the unmerged, unmonitored pronunciations of the two variables. Then the speakers read the passage without (ɔ) words silently and then out loud twice. This completed the readings for Ph and GenUS. At this point, in order to induce a style shift, the two Utahns were shown the two texts they had read and told that they differed in whether or not they contained (ɔ) words although both contained (ɑ) words. As linguistics students, they focused on the words immediately. Both said that they had noticed the large number of words from these word classes. I then asked them to read the (ɔ–ɑ) passage two more times. The most fluent reading of the passage under each condition described above was presented to the listeners for a total of 10 test readings: three from each of the two Utahns and two from each of the other two. Before hearing the first of the 10 test readings, the listeners were presented with GenUS's, the trained reader's, lesser performance on the (ɔ–ɑ) text. This introductory reading was included to accustom the judges to the 'city-slicker in the country' story and to the rating task, but was not included in any analysis in this study.

The stimulus tapes were constructed from digitized speech using a Macintosh II and Diaxis software in Tracy Peterson's sound laboratory in the Music Department of the University of Utah. The signals were edited to remove non-speech noises and false starts when possible to do so. Two different tapes were made with the 10 'speakers' appearing on each tape in a different random order. The subjects were recruited from introductory Linguistics and Psychology courses, and from the general University of Utah population by means of an advertisement placed in the student newspaper in the Spring of 1991. The subjects recruited from the Psychology course were given course credit; the others were paid $5 at the completion of the tasks. Although there were 74 participants in the study, the data from two of these subjects were excluded from the analysis: one of these was a non-native speaker of American English and the other did not complete the task.

One of the two tapes was played to 45 and the other to 29 participants. Of the individuals figuring into the statistical analyses, 38 were males and 34 were females. Forty-five were native Utahns, 17 were natives of other merger areas, and 10 were natives from non-merger areas (determined by a compilation of data presented in Wolfram, 1991, and Labov, 1991). Because this last group is small, any results based on the responses from these subjects will have to be considered with caution.

Each of the 74 participants was given a packet containing one copy of the response sheet for each of the 11 voices they heard on the tape and a form requesting information concerning personal demographics. After hearing a voice, the subject had 1:45 minutes to respond before hearing the next voice. Questions were solicited from the subjects after they heard the introductory reading. The tapes were not stopped after that.

Results

Vowel categorization task

The results from the analysis of the comments, or lack of comment, concerning whether subjects accepted the claim that *stock*, *stalk*, and *caught* in box 9 of the vowel categorization task have the 'same vowel' show that speakers believe that (ɔ) and (ɑ) are not distinct vowels. Only two out of 122 individuals for whom we have information gave any indication that they did not believe that *stock* and *stalk* have the 'same vowel'. Only six individuals of the 122 commented on the pronunciation of the words in box 9. Two of these six, 161, a 16-year-old female, and 402, a 39-year-old woman, volunteered that they believed the words were homonyms. One other individual, 392, a 49-year-old man, pronounced the vowel sound for each box and gave only one sound for box 9. One other individual, 543, an elderly woman, asked about the pronunciation of 'S-T-A-L-K', suggesting that she was not familiar with the written word. The last two, 132, a 43-year-old female, and 202, a 37-year-old female, did not object to the 'each box has one vowel' direction directly, but did make a point of pronouncing the words with distinct pronunciations for the interviewer. The overwhelming silence on the issue suggests that Salt Lake Valley natives generally believe that *stock* and *stalk* are homonyms.

BW and NM spontaneously commented on their pronunciation of (ɔ) and (ɑ) during the taping session for the minimal pair list. Although BW does not believe that he can tell the two vowels apart, NM, surprisingly, volunteered that *caught* and *cot* were different. In attempting to prove her point, she produced *caught* and *cot* about a dozen times each so differently that I could detect a phonation (creaky–breathy) difference and a duration difference although they did not seem to differ in formant values. The acoustic measurements of these tokens and others containing the vowels in question from her spontaneous speech is discussed below. The results of this analysis confirms my auditory impressions.

Analysis of production data

A two-way ANOVA (Vowel ((ɔ) and (ɑ)) × Voicing (of the following consonant)) was run for the onset and another of the same design for the offset for the Balanced Word List data for NM.[9] The results for these and the other acoustic analyses appear in Tables 5–7. The ANOVA for the onsets was identical in design for BW, but because of a large amount of missing data in the offset measurements, a one-way ANOVA with Vowel only was performed on the offsets of the words that were long enough to be measured twice.

The results for NM show significant main effects for Vowel for the onset of F1 and F2, and the offset of F2 (Table 5). There was also a significant main effect for Voicing for the onset of F1 and the offset of F2 (Table 6). Finally the offset of F1 showed an interaction between Vowel and Voicing (Table 7).[10] The Vowel effects were all in the expected direction. The (ɑ) was lower and more fronted than the (ɔ). The interaction between Vowel and Voicing for the offset of F1 indicates that for this parameter the offsets of the two vowels contrast only before voiceless consonants. Before voiced consonants the offset of F1 for (ɔ) is only 41 Hz higher than for (ɑ), that is (ɔ) is a slightly lower vowel than (ɑ).

Table 5. Summary of significant main effects for Vowel comparing (ɔ) and (α): df, F, *p*

Speaker	Style[a]	df	F1/onset	F2/onset	VQI/onset	df	F1/offset	F2/offset	VQI/offset	Duration
NM	BWL	1,26	8.525**E[b]	30.160***E				18.555***E		—
NM	Spon	1,29			5.376*R		—	—	—	4.239*R
NM	MPWL									
BW	BWL	1,26	25.992***E	52.141***E		1,8		8.691*R		—
BW	MPWL	1,54		3.030 ? E[c]		1,55				2.921 ? E[c]
UR	HWL	1,74		37.696***E	6.246*R[c]	1,74		3.006 ? E		—
UR1	Reading						—	—	—	—
UR3	Reading	1,21	4.678*E	3.746 ? E	4.052 ? R		—	—	—	—
UU1	Reading						—	—	—	—
UU3	Reading						—	—	—	—
192A	Interview	1,95		6.307*E						—
251	Interview	1,74	11.650**R		6.969*R					—

[a]*Key*: BWL = Balanced Word List, HWL = Homophone Word List, MPWL = Minimal Pair Word List, and Spon = Spontaneous speech sample.

[b]Significance levels: ****p*<0.001; ***p*<0.01; **p*<0.05; ?*p*<0.10. E = the difference is in the expected direction, R = the difference is in the reversed direction.

[c]The difference between the means is probably imperceptible.

Table 6. Summary of significant main effects for Voicing/Final Consonant comparing (ɔ) and (α): df, F, p

Speaker	Style[a]	df	F1/onset	F2/onset	VQI/onset	df	F1/offset	F2/offset	VQI/offset	Duration
NM	BWL	1,26	22.508***[b]			1,26		9.903**		—
NM	MPWL	1,67		8.268**		1,66	10.192**		12.659**	32.302***[c]
BW	BWL	1,26	8.279**		10.574**		—	—	—	—
BW	MPWL	1,54		10.973**		1,55	9.911**		50.205***	82.889***
UR	HWL	2,74	5.324**	74.613***	6.856**	2,74	42.339***	16.473***	3.015?	—
UR1	Reading	1,21	4.731*	14.490**			—	—	—	—
UR3	Reading	1,21		4.068?			—	—	—	—
UU1	Reading	1,22	3.266?	5.272*			—	—	—	—
UU3	Reading	1,21	6.485*		4.439*		—	—	—	—
192A	Interview	1,95	18.097***	17.140***		1,37	8.264**	4.512*		—
251	Interview	1,74	14.390***	9.626**	12.460**	1,24	7.309*	6.307*	8.035**	—

[a] *Key*: BWL = Balanced Word List, HWL = Homophone Word List, MPWL = Minimal Pair Word List, and Spon = Spontaneous speech sample.

[b] Significance levels: ***$p<0.001$; **$p<0.01$; *$p<0.05$; ?$p<0.10$.

[c] df = (1,68).

Table 7. Summary of significant differences in Vowel × Voicing/Final Consonant interactions comparing (ɔ) and (α): df, F, p

Speaker	Style[a]	df	F1/onset	F2/onset	VQI/onset	F1/offset	F2/offset	VQI/offset	Duration
NM	BWL	1,26				17.324***[b]			—
NM	MPWL								
BW	BWL	1,26		4.392*	3.119?	—	—	—	—
BW	MPWL								
UR	HWL	2,74	5.163**	3.368*					—
UR1	Reading					—	—	—	—
UR3	Reading	1,21	6.478*			—	—	—	—
UU1	Reading	1,22	3.806?			—	—	—	—
UU3	Reading	1,21		5.958*		—	—	—	—
192A	Interview	1,37					3.230?		—
251	Interview	1,74	3.384?						—

[a]*Key*: BWL = Balanced Word List, HWL = Homophone Word List, MPWL = Minimal Pair Word List, and Spon = Spontaneous speech sample.

[b]Significance levels: ***p<0.001; **p<0.01; *p<0.05; ?p<0.10.

BW's data also show significant main effects for Vowel for the onset of F1 and F2, and the offset of F2. Voicing also produces a main effect for the onset of F1 and VQI. Both the onset of F2 and VQI show an interaction between Vowel and Voicing. The difference between the means for the onset of F1 and F2 show that the contrast is in the expected direction. For F2 the difference between the vowels before the voiceless consonants is especially large. In contrast, the offset of F2 for (ɑ) is lower than the offset for (ɔ), indicating that at the end of the vowels (ɑ) is more backed than (ɔ).

Although the interaction between Vowel and Voicing is at a lower level of significance for VQI, the phonation parameter discussed, the difference between the VQIs of (ɑ) and (ɔ), –3.79 dB, is in the expected direction and is large enough to be perceptible. Furthermore, the VQI data for the vowels of the words of the only minimal pair in the Balanced Word List, *hod* and *hawed*, show a large difference for both BW and NM. The difference is clearer in the offset for BW: (ɑ) had a VQI of –16.21 dB and (ɔ) of –11.59 dB, a difference of –4.62 dB. For NM the difference is clearer in the onset. She had a VQI of –25.06 for (ɑ) and –19.92 for (ɔ), a difference of –5.14 dB. This is well within the 2 dB criterion for perceptibility of VQI contrast. Di Paolo and Faber also found that, for the tense/lax pairs (i–I, e–ɛ, u–ʊ), BW usually has VQI differences in the offset of the vowel and NM usually differentiates in the onset. The results from the Balanced Word List show that both speakers distinguish (ɔ) and (ɑ) in F1/F2 and may also differ in VQI for one minimal pair.

A two-way ANOVA (Vowel ((ɔ) or (ɑ)) × Final Consonant (/p, k/, /d/, or /l/)) was run for the onset and another of the same design for the offset for the Homophones Word List data for UR. The onset of F1 shows a significant effect for Final Consonant, and the onset of F2 for Vowel, and Final Consonant, and the interaction between Vowel and Final Consonant. The means show that (ɑ) has an F2 43 Hz higher than (ɔ). (While the difference between the F2s in *pop* and *talk* is only –2 Hz, *cod* and *hawed* differ by 50 Hz and the vowels before /l/ differ by 56 Hz.) The VQI onset measurements also show significant differences for Vowel, Final Consonant, and the interaction between Vowel and Final Consonant. Although the VQI difference between (ɔ) or (ɑ) is generally too small to be perceptible, the means for the interaction between Vowel and Final Consonant show that *cod* has a VQI that is 2.8 dB higher than that of *hawed*. But the distinction is in the reverse direction of what is expected: the vowel of *cod* is breathier than that of *hawed*. For the offsets, F1 shows an effect for Final Consonant, and F2 for Vowel and Final Consonant, and VQI for Final Consonant. For the offset of F2, the mean of (ɑ) is 67 Hz higher than that of (ɔ). The results of this acoustic analysis indicate that in this word list style UR produces a perceptible but, perhaps, not salient difference in F2 for the vowel pair, and a perceptible VQI difference for one near minimal pair, *cod* and *hawed*.

The question now is whether speakers from this area also distinguish (ɔ) and (ɑ) in conversational speech in spite of their belief that the vowels are the same. Word lists, after all, are notorious for prompting speakers to make distinctions that they do not otherwise make or merge those which their conversational style speech does retain. The results of acoustic analyses on a sample of NM's spontaneous speech, and the spontaneous speech of two other individuals addresses this question.

Thirty-one tokens were analyzed from NM's spontaneous speech during the Minimal Pairs Word List taping session. They consisted of one token each of *wok*, *top*, *rocks*, *Tosh*, *pol*, *lot*, *all*, and *walk*, and 11 tokens of *cot* and 12 of *caught*. The results of a one-way ANOVA with Vowel as the independent variable show an effect for VQI and duration. The difference between the VQIs ((ɑ) – (ɔ)) = 6.22 dB) and the durations ((ɑ) – (ɔ) = 57.8 msec) of the two vowels are sufficiently large enough to be perceptible. However, the distribution for these secondary features is the reverse of what would be expected if (ɔ) was the tense vowel and (ɑ) the lax vowel in a tense/lax pair. As I heard during the session, there was no systematic difference in F1 or F2 although the means of both were slightly higher for (ɑ) than for (ɔ).

Similar ANOVAs to those described for NM's Balanced Word List data were performed on the data from the interviews conducted with ImLS speakers, 192A and 251. For 192A, there were significant main effects for Vowel for the onset of F2, and for Voicing for the onsets and offsets of F1 and F2. In addition, there is also a significant interaction between Vowel and Voicing for the offset of F2. The means show that the onset of (ɑ) has a higher F2 than that of (ɔ) and that, for the offsets, (ɑ) has a higher F2 before voiced consonants and a slightly lower F2 before voiceless consonants.

The results for 251 show main effects for Vowel for the onset of F1 and VQI. Voicing is significant for all the parameters tested: the onsets of F1, F2, and VQI, and the offsets of F1, F2, and VQI. There is also a significant interaction between Vowel and Voicing for the onset of F1. The means for the onset of F1 indicate that 251 reverses the traditional height distinction between (ɔ) and (ɑ). That is, (ɔ) is a lower vowel than (ɑ) at the onset and the difference is greatest for the vowels before voiced consonants. Similarly, the 2.01 dB difference in VQI between the onset of (ɑ) and (ɔ) is the reverse of what is expected, given the prediction that (ɔ) is the tense or breathier vowel and (ɑ) is the lax or creakier vowel of a tense/lax pair. Furthermore, a difference of 2.01 dB is probably large enough to be perceptible. Thus, two of the three acoustic parameters indicate a flip-flop of (ɔ) and (ɑ), and the third, F2 does not distinguish the two vowels.

The results from the acoustic analysis of the speech of the older male, 192A, compared to that of the younger female, 251, suggest that the near merger may be moving towards a flip-flop such as that reported in Labov *et al.* (1972) and Yaeger (1974) for (ɔr) and (ɑr) in rural southern Utah. Although before /r/ the merger and flip-flop are highly stigmatized in Utah and may now be arrested, this sound change is apparently camouflaged in other phonological contexts, and so may continue.

Although the (ɔ) and (ɑ) word classes do not differ on every phonetic parameter tested, they are distinct, contrary to predictions that might be made on the basis of the lack of reaction to box 9. The results so far suggest that a minimal pair word list should elicit small, but systematic, etymologically based differences in F1/F2 along with some differences in the secondary parameters of VQI and duration. However, as I will now show, in the Minimal Pair Word List condition, the distinction between the (ɔ) and (ɑ) word classes weakens or collapses.

The results of the ANOVAs (identical in design to those described earlier for NM's Balanced Word List data) for NM on the Minimal Pair Word List show only significant effects for Voicing. A third ANOVA was run with duration of the vowel as a dependent variable which also only shows an effect for Voicing.[11] The results for BW from similar ANOVAs were somewhat more interesting. Although the analysis of his speech also shows a number of effects for Voicing, there are also two effects for Vowel, for the onset of F2 and for duration, but at a lower level of significance. Also, the differences between the means on both measures are very small, 30 Hz for the onset of F2 and 10 msec for duration, and may, therefore, be dismissed as imperceptible differences.

The results of the acoustic analysis show that Salt Lake Valley speakers in speaking spontaneously or in reading word lists containing relatively few tokens of the target vowels, produce systematic etymologically based differences between (ɔ) and (ɑ), contrary to what they believe they do. However, in reading a word list containing many minimal pairs of these vowels or in discussing these vowels, situations in which the vowel classes should be most salient, the etymologically based differences disappear, become supported only by secondary features, or become imperceptible and, thus, their production matches or approximates their usual self-reports.

Language attitudes task

Analysis of the production data. Two-way ANOVAs (Vowel ((ɔ) or (ɑ)) × Voicing (of the following consonant)) were performed for each reading for each speaker. The results of these analyses are also shown in Tables 5–7. Compared to the other sets of data reported on, the number of tokens in this set was not large: (ɔ) was represented by 17–18 tokens per analysis and (ɑ) by 7–8.

The analysis for UU1 shows significant main effects for Voicing for both F1 and F2 and a significant interaction between Vowel and Voicing for F1. The means corresponding to the interaction show that (ɑ) has a lower F1 (that is, it is a higher vowel) before a voiceless consonant than (ɔ) has ((ɑ) – (ɔ) = –64 Hz) and a higher F1 before a voiced consonant ((ɑ) – (ɔ) = 42 Hz). For UU3, the main effects were Voicing for F1 and VQI. The results for UU3 also show an interaction between Vowel and Voicing for F2. The means indicate that (ɑ) has a lower F2 before voiceless consonants ((ɑ) – (ɔ) = –124 Hz) and a higher F2 before voiced consonants ((ɑ) – (ɔ) = 180 Hz) than (ɔ) does. UU shows evidence of a flip-flop, but, in contrast to speaker 251, F1 (in UU1) and F2 (in UU3) show reversals of the expected values before voiceless consonants. Insofar as differences in F1 may be more salient to listeners than differences in F2, UU1 can be said to have a greater contrast between (ɔ) and (ɑ) than UU2.[12]

For UR1 Voicing is significant for both F1 and F2 but there are no other effects. Thus, there is no indication of an etymological distinction between (ɔ) and (ɑ) based on the phonetic parameters that were measured. On the other hand, the analysis of UR3 shows a main effect for Vowel for F1, F2 and VQI, and for Voicing for F2. It also shows an interaction between Vowel and Voicing for F1. The means indicate that (ɑ) has a higher F1 than (ɔ) and that before a voiced consonant, (ɑ) is a much lower vowel than (ɔ). Likewise, UR3 maintains the conservative distinction by producing (ɑ) with a higher F2 than (ɔ). The VQI difference, on the other hand, is a reversal of the expected distribution. The (ɑ) tokens have a higher VQI (breathier) than the (ɔ) tokens (creakier) and the difference between them, 4.18 dB, is probably large enough to be perceptually salient. UR1 shows less of a contrast and UR3 shows much more of a contrast between (ɔ) and (ɑ) than any other data sets from Utahns examined in this study.

Analysis of the listener/judges ratings. The analysis of the ratings of the listener/judges began with the factor analyses using principal components analysis and varimax (orthogonal) rotation. One factor analysis on the 13 dimensions listed in Table 4 was performed for each of the 10 guises. The results group the dimensions into the three scales shown, Personality, Success and Speech, for all of the guises with the exception of Ph1 and GenUS1. However, reliability tests performed on each of the 10 sets of data (one set for each guise) show that it is reasonable for all of the sets of data to group the dimensions into the three factors developed for the majority of the guises. The α-scores from the reliability tests ranged from 0.6250 to 0.9276. The three factors thus developed were entered into the subsequent MANOVAs.

Two sets of repeated measures MANOVAs were performed for each of the three factors developed by the factor analysis (Personality, Success and Speech). In the first set, the within-subjects variables were Speaker (UR and UU) and Reading (reading 1: with both (ɔ) and (ɑ) words, reading 2: containing only (ɑ) words, and reading 3: with both (ɔ) and (ɑ) words); Sex and the Native Area of the listeners as the between-subjects variables; and tape order as the covariate.

The repeated measures MANOVAs performed on the three conditions for the two Utah speakers only showed a main effect for Speaker, as Table 8 shows. UU was judged to be more successful, and to have a more favorable personality and better speech than UR.

Table 8. Results of MANOVA comparing Utah Rural and Utah Urban

	Success		Personality		Speech	
	df	F	df	F	df	F
Speaker	1,66	20.96000***[a]	1,66	7.05000**	1,65	13.01000**
Sp×Rdng						
(UU1, 3 and UR1, 3)	1,66	7.90034**				
(UU1, 2, 3 and UR1, 2, 3)	1,66	5.67323*	1,66	4.98325*	1,65	4.97067*
Natv×Sex×Sp×Rdng						
(UU1, 2, 3 and UR1, 2, 3)	2,66	3.45444*	2,66	6.54623**		

[a]Significance levels: ***$p<0.001$; **$p<0.01$; *$p<0.05$; ? $p<0.10$.

The analyses also showed two-way interactions between speakers and the three reading conditions for Personality, Speech and Success. For both Personality and Speech, the comparison of the first and third situations for the two speakers showed no difference although the trend was that the third reading was rated more favorably for UU and less favorably for UR. But the addition of the non-contrast, the second, situation produced significant differences. In the non-contrast reading guise, UU was rated as having a much more favorable personality and having much more favorable speech than in UU1 and UU3 and than any of UR's readings. The rating differences for UR1, UR2 and UR3 were minor.

Although the two Utahns were thought to be about equally successful on the basis of their first readings, judging from the third reading, the listeners believed that UU was much more successful and that UR was less successful. The addition of UU2 and UR2 to the analysis also produced a significant difference. UU2 and UR2 seemed to differ from UU1 and UR1 but not from the third readings. Again, UU2 was rated as being more successful than he was on the basis of the corresponding first reading and UR2 was rated as being less successful.

The general trend, then, is that UU is perceived as being more successful and having more favorable speech and personality traits in the non-contrast situation and sometimes in the third reading, the reading showing the most merger. For UR, when the ratings of the second and third readings differ from those of the first, the reading which shows the most merger, UR1 is rated more favorably than UR2 or UR3. Thus, there is evidence that UU's and UR's guises showing the most merger, compared to their guises showing the least merger, are perceived as being more successful and as having more favorable speech and personality traits.

Success and Personality also have a significant four-way interaction among Native Area (of the listeners), Sex, Speaker and Reading. For both factors, the significant differences appear in the comparison between the second reading with the first and third. The largest difference involving the second readings for both the Success and Personality factors occurs in the ratings of the non-merger females for UU2 and UR2. These females perceived UU2 as having much more positive personality traits and as being much more successful than UR2. In fact, the ratings for UU2 and UR2 from the non-merger females represented the most favorable Success and Personality ratings, and the least favorable Personality ratings, respectively, given by any group of females in the sample. These females also rated UR1 as having more positive Success and Personality traits than UR2, and UU2 more positive traits than UU1.

There were a number of other large differences[13] between the readings for various groups of judges which may not be significant differences, but which suggest that there is an overall trend of discriminating between the readings. For the Personality traits, Utah females rate UR2 more favorably than UR3, males of other non-merger areas rate UU2 more favorably than UU3, and females of the other non-merger areas rate UR3 above UR1 and UU3 above UU1. For the Success ratings, Utahns of both sexes rated UR1 as being more successful than UR3, and males of the other merger areas rate UR3 as more successful than UR2. Every group of listeners, except females of other merger areas, rated UU3 as more successful than UU1. Males from Utah and of the other merger areas also thought that UU2 was more successful than UU1 while Utah females believed that UU2 was less successful than UU3.

Another set of repeated measures MANOVAs was performed for Personality, Success and Speech for all four Speakers and the two Readings (readings 1 and 2) which they all read. The design was similar to the first set of MANOVAs with the exception of the values assigned to the within-subjects variables. In this second set, the within-subjects variables were Speaker (UR, UU, Ph, and GenUS) and Reading (reading 1: possible contrast (with both (ɔ) and (ɑ) words), and reading 2: non-contrast (containing only (ɑ) words)); Sex and the Native Area of the listeners as the between-subjects variables; and tape order as the covariate.

As can be seen in Table 9, the analysis shows a main effect for Speakers on all three factors. On the Personality scales, Ph was rated less favorably than the other three and UR was rated significantly less favorably than GenUS. On the Speech traits, UR was judged to have worse speech than GenUS. The Success factor serves to differentiate the four speakers the most. UR differs from GenUS, UU differs from the previous two, and Ph differs from the rest. GenUS was rated as the most successful, Ph was the next most, UU was the next, and UR was rated as the least successful. Although the overall ratings of the speakers from the merger areas are lower than those of the speakers from the non-merger areas, this is not a serious problem for the hypotheses being tested. Since this is a matched guise experiment, the most important comparison for the hypotheses being tested is not that of speaker to speaker, but of the first reading guise to the second reading guise for any given speaker.

Table 9. Results of MANOVA comparing UR, UU, GenUS and Ph

	Success		Personality		Speech	
	df	F	df	F	df	F
Speaker						
UR vs GenUS	1,66	71.73896***[a]	1,66	7.61573**	1,64	44.29210***
UU vs GenUS/UR	1,66	5.41975*				
Ph vs others	1,66	15.94768***	1,66	20.39365***		
Reading			1,66	12.58000**		
Sp×Rdng						
UU1 vs UU2	1,66	11.54675**			1,64	7.02207*
GenUS1 vs GenUS2			1,66	7.13918**		
Sp×Natv						
UR vs GenUS			2,66	3.95013*		
Sp×Rdng×Sex						
UR vs GenUS			1,66	5.04099*		
GenUS1 vs GenUS2					1,64	5.00379*
Sp×Natv×Sex×Rdng			2,66	4.28650*		
Rdng×Sex					1,64	4.12000*

[a]Significance levels: ***$p<0.001$; **$p<0.01$; *$p<0.05$; ? $p<0.10$.

The Personality factor also shows a main effect for Reading. The means indicate that the second, or non-contrast readings, were rated more favorably than the first readings. In general listeners give overall higher scores on Personality on the second guise, the reading in which only (ɑ) occurred, and in which there was no possibility of a contrast with (ɔ).

The results also show a number of interactions. Among them are interactions between speakers and the two reading conditions for Personality, Speech, and Success. The two readings were given similar ratings on the Success and the Speech scales for all except UU. UU was thought to be much more successful and to have much better speech when he spoke in the non-contrast situation, UU2, than in the guise in which he produced the most contrast, UU1. In the Personality scales, GenUS2 was rated as having much more favorable personality traits than GenUS1 was. UU and Ph showed trends similar to GenUS, but the rating differences between their two readings may not have differed from UR1 and UR2, which received almost the same ratings. Again, the general trend is that readings showing a contrast between (ɔ) and (ɑ) are rated less favorably than readings showing no contrast. This trend is also supported by the lack of a rating difference between UR1 and UR2, neither of which showed a contrast between (ɔ) and (ɑ). UR1 was the most merged reading for UR and UR2 only contained (ɑ).

The Personality analysis shows three more interactions, all involving the Speaker variable. First, there is an interaction between the Native Area of the listeners and Speaker, the variable corresponding to the native area of the four speakers. Listener/judges from the non-merger areas gave UR the most disfavorable ratings but GenUS the most favorable ratings. Second, there is a three-way interaction among the Speaker, Reading and Sex of the judges. As in the previous interaction, the difference is between the ratings given to UR and to GenUS. The general trend was for females to rate speakers somewhat lower than males did, but in the case of GenUS2, the females' ratings were much higher than the males'. In fact, they were the most favorable ratings given to any speaker in any guise in the study whereas the females other ratings were well below those of the males for any given speaker in any reading. Also, the males and females did not differ greatly on their ratings of GenUS1. These two interactions show trends that appear in a four-way interaction with Speaker, Native Area, Sex and Reading. In general, the females from the non-merger areas rated the second reading more favorably than the first reading. This was the case for all but UR's first reading guise, the merged reading, in which he was rated as having a more agreeable personality than in his second reading.

The general trend for Ph and GenUS was for the non-contrast reading to be rated higher than the reading with both (ɔ) and (ɑ). For the judgements concerning Ph, the difference in rating between the two readings was large for all listeners from Utah, males from other non-merger areas, and females from the merger areas. For GenUS, the difference was especially large for female Utahns, all listeners from other non-merger areas, and females from the merger areas. The only large difference in the ratings for the first and second readings for UU occurred in the ratings of the non-merger area females, who believed that UU had a more agreeable personality in his non-contrast guise.

The Speech factor shows two more related interactions from the analyses which compared the four Speakers on the two Reading conditions. The first of these is an effect between Reading and Sex. Females judged the speech of the speakers reading the non-contrast reading to be significantly better than when they read the (ɔ)–(ɑ) passage. Males did not differentiate the two readings. The second is a three-way interaction between Speaker, Reading and Sex. The analysis reveals that males and females generally agreed on the ratings they gave except in the case of GenUS, who was thought to have much worse speech in the non-contrast reading by male judges and much better speech in the non-contrast reading by the females.

Discussion

Production

The findings from the acoustic analysis show that every one of the six speakers distinguished (ɔ) and (ɑ) in some context. Thus, it can be said that Utahns have an (ɔ)–(ɑ) distinction. More specifically, under relatively unmonitored conditions (reading a word list with few tokens of the target vowel as in the Balanced Word List and the

Homophones Word List, and in the spontaneous speech from interviews) (ɔ) and (ɑ) for Utahns pattern like a tense/lax pair of vowels in F1/F2 and VQI although the distinguishing parameters vary by speaker and situation. However, when the word classes are made more salient for the speakers, that is, when they read a word list containing a high proportion of pairs from these word classes or talk about the difference (NM's spontaneous speech), they style-shift and produce a vowel distribution with no F1 or F2 difference of more than 65 Hz for a given minimal pair. The particular stylistic differences are reminiscent of those reported by Labov *et al.* (1972) for the Pennsylvania speaker referred to earlier, who had about a 50 Hz difference between the vowels of these word classes in monitored speech. This sort of shifting is probably characteristic of near mergers.

The results of the acoustic analysis on the word list data and on the interview data predicts that UR and UU would produce more of a phonetic contrast between (ɔ) and (ɑ) in the first reading (UR1 and UU1) than in the second reading (UR3 and UU3) of the (ɔ) and (ɑ) passage. The results indicate that UU may have performed the task according to the prediction, but UR clearly did not. UR3 produced a strong and consistent contrast using F1, F2 and VQI while UR1 had no contrast at all. UR's reading of the Homophone Word List, a task comparable to the Balanced Word List reading in having the same proportion of (ɔ) and (ɑ) words to total words, showed an intermediate distinction. In it, UR distinguished the word classes using F2 and VQI, parameters that may be less consciously salient than F1. In UR3, UR may have been orienting himself to a perceived outside norm, triggered by my professorial mini-lecture on the conservative distinction between the word classes. UR's greater experience with different varieties of English, and ESL teaching, both here and abroad, may provide him with a larger verbal repertoire and possible set of norms than those available to UU, whose behavior did not vary as much.

These particular style-shifts raise the question, discussed most notably in Labov's work and in Bell (1984), of whether the underlying explanation of style-shifting is attention paid to speech or to reference group. If 'attention paid to speech' means 'attention paid to what speech should be', that is, speakers' categories, then some of the results suggest that speech, *language* in other words, makes the difference. NM, BW, UU, and UR had the same audience for each of the word lists or reading passages: a tape-recorder and a linguist-acquaintance, or some third party linguist-strangers if they considered the outcome of the recording sessions. The only constant differences between the two word list situations was how many low back vowel words the speakers were confronted with. In other words, the situations differed in features that would be more likely to prime language categories rather than audience categories. On the other hand, the situational difference between the first and third readings for UU and UR was not only how salient the test vowels were, but also how salient the professor–student relationship was. In this case, style-shifting can also be explained in a reference group design framework with linguists as the referee group. When I pointed out that the two sets of words belonged to different word classes, UR may have attempted to accommodate to my expectations as a linguist and an outsider (non-local, non-Mormon) by making a sharp distinction between the vowels. Such a non-local audience has been associated with style-shifting by both Douglas-Cowie (1978), and Rickford and McNair-Knox (1991).

These results also call into question some commonly accepted methodological techniques for inducing style-shifts. Namely, not all word lists are stylistically equivalent. Word lists with a relatively low proportion of the target sound may, in fact, prompt pronunciations that are quite similar to that of spontaneous speech. Likewise, reading passages may result in a range of possible pronunciations depending on how the experimenter presents the situation. These are valuable lessons in constructing and presenting materials to participants. Specific protocols, ones that are produced verbatim, are a must for ensuring task equivalency.

Attitudes

From the general indifference of the ImLS participants, natives of the Salt Lake Valley commonly believe that (ɔ) and (ɑ) are or should be one word class. This indicates that the speaker's perceptions of what is 'correct' or pleasing speech motivates the style-shifting found in the present study. Anecdotal evidence suggests that Westerners from the Rocky Mountain area are not favorably inclined towards raised or very rounded (ɔ). For example, undergraduate students behave as if the (ɔ)–(ɑ) distinction is superfluous and somewhat amusing. In class discussions I have never found any evidence from native speakers of the area that the unambiguous *separation* of the (ɔ)–(ɑ) word classes held any sort of positive value.[14] In other words, the participants of the ImLS behave as if they have one consciously accessible category for both (ɔ) and (ɑ), and in comparable situations residents of the Intermountain area behave as if 'correct' or standard speech is speech in which there is no distinction between the (ɔ) and (ɑ) word classes. The evidence from the language attitudes study confirms my observations.

The results of the study on the attitudes towards texts containing a distinction or lacking a distinction between (ɔ) and (ɑ) show that Utah speakers are rated as being more successful when they are producing less of a contrast than when they more clearly distinguish the two word classes. UU1, producing more of a salient distinction, was rated as more successful than UU3, with less of a contrast. UR1, with no contrasting feature, was thought to be more successful than UR3, with all three phonetic parameters contrasting. These results support Hypothesis 1, and indicate that the merger of (ɔ) and (ɑ) has overt prestige and that a shift towards the merger is hypercorrection.

Hypothesis 1. Listener/judges believe that a speaker is more successful, has better speech, and a more favorable personality when he is heard making loss of a distinction between (ɔ) and (ɑ) than when he makes more of a distinction.

Unexpectedly, the study also shows that females who are natives of non-merger areas (but who now live in Utah) believe that UR is more personable and more successful in his merger guise (UR1) than in his non-contrast guise (UR2). Similarly, they rate UU more favorable on Personality and Success in the non-contrast guise (UU2) than in the guise with a more salient distinction (UU1). Although this group of participants is very small ($N = 5$) and these females are geographically separated from their native speech communities, these findings suggest that females from non-merger areas may begin to lead the near merger of (ɔ) and (ɑ) in their native speech communities.

In general, the results for the analyses comparing all four speakers supported *Hypothesis 2.*

Hypothesis 2. Listener/judges believe that a speaker is more successful, has better speech, and a more favorable personality when he gives no cue about whether he makes a distinction than when he is heard making a distinction between (ɔ) and (ɑ).

The only effect contradicting this generalization is that males believe GenUS's speech was better in his cue guise (GenUS1) than in his non-contrast guise (GenUS2). I have no ready explanation for this finding which would relate to this study.

The finding on the personality scale that females rate UR1, the reading with no difference on the parameters measured, more favorably than UR2, the non-contrast reading, also deserves comment. It suggests that the females, at least, could tell the non-contrast and completely merged reading of UR's apart. That is, in spite of the fact that most of them might consciously say that (ɔ) and (ɑ) are one word class (all but five of them were from a merger area), they must be able to recognize that there were two word classes, (ɔ) and (ɑ), present in the merged reading and only one, (ɑ), in the non-contrast reading. This leads to a number of interesting questions about what word class membership means during a near merger, and which phonetic parameters distinguish speech tokens because the ones measured showed no significant differences for UR1.

This study has presented evidence that the (ɔ) and (ɑ) word classes are still distinct in Utah speech, an area that has been reported to have the word classes merged. However, a speaker may vary in pronunciation from a complete merger to a strong contrast depending on how the speaker assesses the situational context. There are also indications that the vowels may be moving towards a flip-flop or reversal in quality. Attitudinal evidence indicates that this sound change is a prestigious one. Currently, speakers hypercorrect in some situations by merging the vowels, the pronunciation that is the perceived norm.

Acknowledgments—This study was supported in part by a College of Humanities Career Development Grant, the Linguistics Program Development Fund, and the Intermountain Language Survey at the University of Utah. I would like to thank Alice Faber, Susan Andrus Wood, Malcah Yaeger-Dror, Ray Freeze, Judy Florence, Shari Kendall, and Kenneth Ranson for helpful discussion and assistance on the project. David Huth provided me with invaluable advice on the statistical analysis. I also owe Tracy Peterson, Marvin Hanson, Cheryll Lamph, and David Dodd a note of appreciation for allowing me access to their campus facilities. All errors of commission or omission are my own. An earlier work on the phonetics of (ɔ) and (ɑ), co-authored with Susan Andrus Wood, was presented at NWAV 1990.

NOTES

[1]I do not mean to imply that (ɔ) and (ɑ) are never merged. Some investigators, for example, Herold (1989) have, indeed, found true F1/F2 mergers of these vowels.

[2]The Vowel Categorization Task was the final and most difficult one included in the interview protocol. A number of participants, many of them elderly, did not complete the task because of time constraints, illness, or poor vision. In some cases, interviewers erred by turning off the tape-recorder during the instructions for the task and so there is no information for box 9 words.

[3]There is a basic problem with constructing a minimal pair word list for the low back vowels. (See Terrell, 1976, for the historical background of this difficulty.) There are few minimal pairs, especially ones for which both members of the pairs are common words that people are accustomed to seeing in print. For example, although *Paul* is a relatively common word, *pol*, its pair, is not. Likewise, many people feel more comfortable with the word *hawed*, 'the sound a donkey made', than with *hod*; *cod* proved to be more familiar. In the case of *caulk* and *cock*, local pronunciation thwarted our efforts at obtaining minimal pair data. The first word in the Salt Lake Valley is commonly pronounced with a dark /l/ so it is not a minimal pair with *cock*. Because of these difficulties, only one minimal pair appears in the Balanced Word List, *hawed/hod*.

[4]The analysis presented in Faber and Di Paolo (1992) indicates that UR is representative of the others in the homophones Word List sample for the (ɔ) and (ɑ) word classes.

[5]The classification of these words into the (ɔ) and (ɑ) classes was based on Kenyon and Knott (1953). Words that could belong to either class in some dialect of American English were excluded from consideration.

[6]I use the terms 'tense/lax' instead of some other possible designation to refer to the relationship between vowels in pairs of words such as *heed–hid* and *hayed–head* because these terms make fewer claims about the physical reality of a possible contrast. In part, my work is attempting to determine the physical reality, not prejudge it.

[7]Two measurements were made when possible because previous research on Utah English (Di Paolo, 1988; Di Paolo and Faber, 1990) showed that diphthongization is involved in some of the local sound changes. Although the results of the measurements for (ɔ) and (ɑ) show minimal differences in the onsets and offsets relative to other vowels measured in my work on Utah English, many of which consist of a clear nucleus and glide, the two measurement points for these back vowels allow comparisons with that other body of work on Utah English. The measurements were made at 52 msec from the beginning or end of the vowel nucleus (unless these points fell within the initial or final consonant transitions, which they rarely did) in order to establish reliability between my measurements and those of my research assistants, Judy Florence and Shari Kendall.

[8]In this study, the criteria for being a native speaker of an area is that the individual in question had to have lived in an area before the age of 9 and then lived in that area 5 or more years between the ages of 4 and 13. These criteria were used to determine who should count as a native speaker in both the New York City (Labov, 1966) and Philadelphia surveys (Payne, 1980), for a study of cross-dialectal miscommunication (Labov, 1989), and for Di Paolo and Faber (1990).

[9]Consonant voicing and final consonant was included in the ANOVAs because they had been shown to produce significant interactions with Vowel for VQI (McRoberts and Faber, 1989; Di Paolo and Faber, 1990). However, main effects for voicing or final consonants are not of concern for this study since consonant voicing or type does not address the issue of the degree of contrast between the two vowels. Any such main effects are reported when significant but not discussed further.

[10]An interaction between two of the factors tested means that there are statistically significant differences between (ɔ) and (ɑ) word classes when both factors are examined together for the parameter under discussion. In this case, for example, an interaction between Vowel and Voicing for the offset of F1 means that, when the Vowel and Voicing of the following consonant are considered together, there is a significant difference. Specifically, the offset of F1 of (ɔ) is significantly different from the offset of F1 of (ɑ) before voiceless consonants, as is described below.

[11]Vowel duration was run in a separate ANOVA because it did not logically fit with just the onsets or just the offsets of the vowels.

[12]Labov *et al.* (1972), Feagin (1986), and Di Paolo and Faber (1990) report data suggesting that listeners may be more sensitive to smaller changes in F1 than they are to changes in F2. The study of near mergers would benefit greatly from carefully controlled studies in this area.

[13]Since no *post hoc* test could be performed on the data, I arbitrarily took a difference of 0.300 to be a sufficiently large enough difference to report.

[14]The only exception is that one of the (ɔ)–(ɑ) contrast before /r/, which is vehemently stigmatized by those aspiring to join and/or hold the ranks of the college-educated. However, in the urban Salt Lake Valley the preferred pronunciation of pairs such as *born–barn* seems to be [born]–[bɑrn], not [bɔrn]–[bɑrn]. The recent history of (or)–(ɔr)–(ɑr) in Utah seems to be similar to that described in Thomas and Bailey (1991) for Texas.

REFERENCES

BELL, A. 1984 Style as audience design. *Language in Society* **13,** 145–204.

DI PAOLO, M. 1986 A study of double modals in Texas English. Unpublished Ph.D. thesis, University of Texas at Austin.

DI PAOLO, M. 1988 Pronunciation and categorization in sound change. In Ferrara, K., Brown, B., Walters, K. and Baugh, S. (Eds), *Linguistic Change and Contact: Proceedings of the Sixteenth Annual Conference on New Ways of Analyzing Variation in Language*, Vol. 30, pp. 84–92. Texas Linguistic Forum, Austin.

DI PAOLO, M. and FABER, A. 1990 Phonation differences and the phonetic content of the tense–lax contrast in Utah English. *Language Variation and Change* **2,** 155–204.

DI PAOLO, M., FABER, A. and McROBERTS, G. 1989 Phonation differences and sound change in Utah vowels. Presented at the Linguistic Society of America Annual Meeting, Washington, DC, 27–30 December.

DI PAOLO, M. and WOOD, S. 1990 The apparent merger of /ɔ/ and /ɑ/ in the Salt Lake Valley. Paper presented at the Nineteenth Annual Conference on New Ways of Analyzing Variation in Language, University of Pennsylvania, Philadelphia, 19–21 October.

DOUGLAS-COWIE, E. 1978 Linguistic code-switching in a Northern Irish village: social interaction and social ambition. In Trudgill, P. (Ed.), *Sociolinguistics in British English*, pp. 37–51. Arnold, London.

EDWARDS, T. and JACOBSEN, M. 1987 Standard and regional standard speech: distinctions and similarities. *Language in Society* **16,** 369–380.

FABER, A. and DI PAOLO, M. 1992 Homophones and near-homophones and their different roles in sound change. Paper presented at the Linguistic Society of America Annual Meeting, Philadelphia, Pennsylvania, 9–12 January.

FEAGIN, C. 1986 More evidence for major vowel change in the South. In Sankoff, D. (Ed.), *Diversity and Diachrony*, pp. 83–95. Current Issues in Linguistic Theory. John Benjamins, Philadelphia.

HEROLD, R. 1989 Using vowel formant measurements to test a theory of merger. Typescript, University of Pennsylvania.

JOHNSON, K. 1989 On the perceptual representation of vowel categories. Paper presented at the 118th meeting of the Acoustical Society of America, St. Louis, November.

KENYON, J. and KNOTT, T. 1953 *A Pronouncing Dictionary of American English.* Merriam, Springfield.

KLATT, D. 1976 Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *Journal of the Acoustical Society of America* **59,** 1208–1221.

LABOV, W. 1966 *The Social Stratification of English in New York City.* CAL, Washington, DC.

LABOV, W. 1989 The limitations of context: evidence from misunderstandings in Chicago. Typescript, University of Pennsylvania.

LABOV, W. 1991 The three dialects of English. In Eckert, P. (Ed.), *New Ways of Analyzing Sound Change.* Academic Press, New York.

LABOV, W., YAEGER, M. and STEINER, R. 1972 A quantitative study of sound change in progress. Report on National Science Foundation Contract GS-3287. U.S. Regional Survey, Philadelphia.

LAMBERT, W., HODGSON, R., GARDNER, R. and FILLENBAUM, S. 1960 Evaluative reactions to spoken language. *Journal of Abnormal and Social Psychology* **60,** 44–51.

LUHMAN, R. 1990 Appalachian English stereotypes: language attitudes in Kentucky. *Language in Society* **19,** 331–348.

McROBERTS, G. and FABER, A. 1989 Variations in measured voice quality related to linguistic and non-linguistic factors. *Journal of the Acoustical Society of America* **84,** S58.

MOONWOMON, B. 1987 Truly awesome: (ɔ) in California English. In Denning, K., Inkelas, S., McNair-Knox, F. and Rickford, J. (Eds), *Variation in Language: NWAV-XV at Stanford*, pp. 325–336. Proceedings of the Fifteenth Annual Conference on New Ways of Analyzing Variation. Department of Linguistics, Stanford University.

NORMAN, A. 1971 A Southeast Texas dialect survey. In Allen, H. and Underwood, G. (Eds), *Readings in American Dialectology*, pp. 135–151. Appleton-Century-Crofts, New York.

PAYNE, A. 1980 Factors controlling the acquisition of the Philadelphia dialect by out-of-state children. In Labov, W. (Ed.), *Locating Language in Time and Space*, pp. 143–178 Academic Press, New York.

RICKFORD, J. and McNAIR-KNOX, F. 1991 Addressee- and topic-influenced style shift: a study in quantitative sociolinguistics. Paper presented at the Twentieth Annual Conference on New Ways of Analyzing Variation in Language, Georgetown University, Washington, DC, 3–6 October.

RULON, C. 1980 A linguistic outline of Texas dialects: some preliminaries (Preliminary Version). Paper presented at the Annual ADS Meeting, Houston, Texas, 30 December.

STANLEY, O. 1936 The speech of East Texas. *American Speech* **11,** 3–36, 145–166, 232–251, 327–355.

TERRELL, T. 1976 Some theoretical considerations on the merger of the low vowel phonemes in American English. In *Proceedings of the Second Annual Meeting of the Berkeley Linguistics Society*, pp. 350–359. Department of Linguistics, University of California at Berkeley, Berkeley.

THOMAS, E. and BAILEY, G. 1991 Competing mergers and their resolutions. Paper presented at the Twentieth Annual Conference on New Ways of Analyzing Variation in Language, Georgetown University, Washington, DC, 3–6 October.

TRUDGILL, P. 1984 Sex and covert prestige: linguistic change in the urban dialect of Norwich. In Baugh, J. and Sherzer, J. (Eds), *Language in Use: Readings in Sociolinguistics*, pp. 54–66. Prentice-Hall, Englewood Cliffs.

WALSH, H. and MOTE, V. 1974 A Texas dialect feature: origins and distribution. *American Speech* **49,** 40–53.

WOLFRAM, W. 1991 *Dialects and American English.* Prentice-Hall and CAL, Englewood Cliffs.

YAEGER, M. 1974 Speaking style: some phonetic realizations and their significance. *Pennsylvania Working Papers on Linguistic Change and Variation* **1,** 1–60.