ACOUSTIC CHARACTERISTICS OF UTAH'S *CARD-CORD* MERGER

DAVID BOWIE
University of Central Florida

ABSTRACT: Several claims have been put forward to explain the character of Utah's CARD-CORD merger, in which /ɔɹ/ is variably produced as [ɑɹ]. Instances of words containing /ɔɹ/ from the running speech of a Utahn who variably exhibits the merger were rated by a panel of speakers on whether they were produced as [ɔɹ] or [ɑɹ], allowing utterances to be classified as merged (into [ɑɹ]), unmerged, or intermediate. Merged and unmerged instances were not found to be separated cleanly when taking just F1 and F2 into consideration, but looking at F1, F2, and F3 simultaneously resulted in a clean split between those categories, with the intermediate cases falling along the border between them. Further, the formant values did not match up with any simple articulatory explanation for the merger and its attendant distinction. Not only does this demonstrate that Utah's CARD-CORD merger is the result of the simultaneous manipulation of multiple articulatory features, it raises the possibility that clear separation between phonetic categories could be found in other cases that might generally be expected to have some overlap.

1. UTAH'S *CARD-CORD* MERGER

ENGLISH SPEAKERS ALONG THE WASATCH FRONT of Utah exhibit a variable linguistic feature that can at least loosely be described as /ɔɹ/ merging into /ɑɹ/, so that the word *cord* is produced as *card*—thus, what is referred to as Utah's CARD-CORD merger. It was first mentioned in any sort of scholarly writing by Pardoe (1935) and has been the subject of a slow but steady stream of study ever since. Even though there is evidence that the feature was extant in the mid- to late nineteenth century among the initial generations of English-speaking natives of the region (Bowie 2003), it does not seem to have become a socially salient feature until the 1950s, possibly the 1940s, with stigmatization of the feature by locals following quickly after that (Cook 1969; Helquist 1970).

In any event, the CARD-CORD merger exists in Utah, with speakers variably producing it. In fact, this would seem to be a completely uncontroversial and uncomplicated position to take, at least as far as the scholarship goes. However, in actual fact, there is complication (and controversy) regarding

American Speech, Vol. 83, No. 1, Spring 2008 DOI 10.1215/00031283-2008-002

Utah's CARD-CORD merger; it simply has not been foregrounded. A close look at the literature on this feature reveals that while everyone who has looked into this merger in Utah is in complete agreement that it exists, there is amazingly little agreement on precisely what it means to say that the merger exists. If one reads through the studies on this feature that have conjectured what the details of the merger might be, one finds the following:

1. Pardoe (1935) described it somewhat vaguely, but in terms that point toward it being a complete merger, with /ɔɹ/ simply collapsing into /ɑɹ/.
2. Cook (1969) developed this further, pointing out that it is a variable behavior. Further, Cook described it as involving a partial overlap of /ɔɹ/ and /ɑɹ/, not a complete collapse of the sounds, which accounted for the perception that the process is a reversal rather than a merger (an opinion that was—and still is—widely held by nonlinguists).
3. Helquist (1970) argued that it is a variable merger of /ɔɹ/ and /ɑɹ/, but that it should be more precisely described as the result of processes of lowering and unrounding affecting /ɔɹ/ rather than one sound simply collapsing into the other.
4. Yaeger (1975), following Labov, Yaeger, and Steiner (1972), described the process as a reversal of the F1-F2 means of /ɔɹ/ and /ɑɹ/, with the phonemes maintaining some overlap and the separation between them seeming to increase over apparent time.
5. Krahnke (1979) claimed that the crucial process involved is a lowering of /ɔɹ/, leading to /ɔɹ/ merging into /ɑɹ/.
6. Lillie (1998) returned somewhat to Pardoe's general description, describing the feature as a merger of /ɔɹ/ into /ɑɹ/, without any further explanation except to recognize its variable character.
7. In an earlier article (Bowie 2003), I claimed that it is a variable merger of /ɔɹ/ into /ɑɹ/, but that it comes about by unrounding of /ɔɹ/.

Obviously, opinion varies regarding the details of /ɔɹ/ and /ɑɹ/ in Utah among those who have studied them. This, then, leads to what should be a central question for anyone looking into this feature of Utah English:[1] given that the CARD-CORD merger exists in Utah, what exactly is going on when it occurs?

2. METHODOLOGY

Since the CARD-CORD merger occurs as a variable merger among speakers of Utah English, this question can be answered only if we can tell the difference between instances of merger and retentions of the distinction. There are, of necessity, two parts to such an investigation: identifying instances of merger

and retention of the distinction, and then analyzing the instances of merger and distinction to discover what the difference between them might be.

To identify instances of merger, a panel of 40 undergraduate college students listened to instances of /ɔɹ/ and rated them as being produced as [ɔɹ] or [ɑɹ]. The stimuli were 100 isolated words selected from speech produced by a male (born in 1922 in northern Utah and a lifelong resident of the Wasatch Front) who, like other speakers from the region, variably exhibits the CARD-CORD merger. All of the tokens were taken from running speech in religiously themed television broadcasts aired between 2000 and 2003. These broadcasts were made under fairly controlled conditions, resulting in very clean audio quality. Audacity version 1.2.3 was used to select words from compact disc recordings produced by the broadcaster; to eliminate possible complications with voicing, no words preceding a pause were used. Twenty percent of these words contained an underlying /ɑɹ/ (as a control), and 80% contained words that could possibly be produced as [ɔɹ] or, if the merger applied, as [ɑɹ].[2] (In fact, the sample was somewhat more complicated than that; the complications are discussed below).

The raters were asked, for each word, whether they heard an [ɑɹ] or an [ɔɹ]. Most of the members of the rating panel were from peninsular Florida; none of them had had any significant exposure to Utah English, and only one of the raters had ever lived in an area where a merger of /ɑɹ/ and /ɔɹ/ has been reported in the literature (she had lived in St. Louis, Missouri, for one year as a teenager). Since the responses of the rater with St. Louis exposure did not differ markedly from those of the rest of the panel (she went against the majority in only two cases, and in those she went against a very slight majority), her responses were included in the analysis.

A note on the nature of the distinction the raters were asked to make is in order here. Since the students did not have an appreciable background in linguistics, they were presented with *are* and *or* as examples of [ɑɹ] and [ɔɹ], respectively. There has, however, historically been a three-way distinction between [ɑɹ], [ɔɹ], and [oɹ] (or, using guide words, *are*, *or*, and *ore*) in English. Previous testing of the rating panel, though, revealed that they participate in the widespread merger of /ɔɹ/ and /oɹ/ (as one would expect, given the findings of Labov, Ash, and Boberg 2006), and so they could not be expected to distinguish between all three possible productions on the part of the Utahn speaker. This raises the possibility, unfortunately, that the rating panel members were actually distinguishing between [ɑɹ] and some sound that was not in their inventory (or even, perhaps, between two sounds that were not in their inventory). An attempt was made to ameliorate any problems this might have caused by not asking the panel to match the sounds they heard to the specific guide words *are* and *or*, but to choose whether each

stimulus "sound[ed] closer to *are* or *or*." Given the fact that people can make reasonably consistent binary distinctions between sounds when asked to rate whether a stimulus is closer to one sound or another (see, for example, Lisker and Abramson 1970 and several related studies since), this was not thought to be a significant problem.

The panelists heard each word three times, and (as already mentioned) rated each word as sounding like it was produced with an [ɑɹ] or an [ɔɹ]. The panelists were explicitly told not to rate what the word was "supposed to sound like," but rather the way it was actually pronounced by the speaker. Their choices between [ɑɹ] and [ɔɹ] were binary and forced, at least as far as a paper and pen survey can have a forced choice. None of the respondents chose both options for any of the words, nor did any provide an alternative to the two choices provided. Individual raters did leave some words unrated, but this was rare, with only a few of the 100 words skipped by any of the 40 raters in the panel (12 words were rated by 39 panelists, 1 by 38, and none by fewer).

Alongside this, the pitch and first through fourth formants were measured for all of the instances of /ɑɹ/ and /ɔɹ/ using the built-in routines in Praat version 4 (Boersma and Weenink 2005).[3] These values were initially obtained for several different points for each instance of /ɑɹ/ and /ɔɹ/: the beginning and end of each vowel, the point of peak intensity of each vowel, and the point of peak intensity of the semivowel /ɹ/. All of these were analyzed separately; the measurements of the vowels all gave similar results, but the most coherent and meaningful results emerged from the point of peak intensity of the vowel, so that is what is reported in this article. (The formant values of /ɹ/ gave completely incoherent results.) In addition, the length of each vowel and its accompanying /ɹ/ were measured, again using Praat version 4, but that data shed no real light on the issue, and so those results are not reported here, either.

All of this means that two types of information were available for investigation in this study: acoustic information on a speaker's production of /ɑɹ/ and /ɔɹ/, and information on listeners' perceptions of those sounds. Having data on both perception and production means that there are two very different things going on in this investigation. Clearly, we need to rely on the perceptual information to know what sounds to place in what categories, so that (for example) items that are actually merged into another sound do not end up grouped with items that are unmerged. Because this is foundational to working with the production data, it requires making some assumptions about what counts as merged or distinct, and those assumptions will need to be checked later—but in any event, once the initial grouping is done, we can investigate the production-related data to determine how the perceptual data relates to it.

Once we have moved on to an investigation of the acoustic data, it becomes clear that an investigation into the acoustic characteristics of the CARD-CORD merger will also, in a very real way, need to investigate the acoustic characteristics of the CARD-CORD distinction. That is, since we start out with sounds divided into various categories such as "X merged into Y" or "X not merged into Y," we will end up having to build a model that simultaneously explains the merger and the distinction.

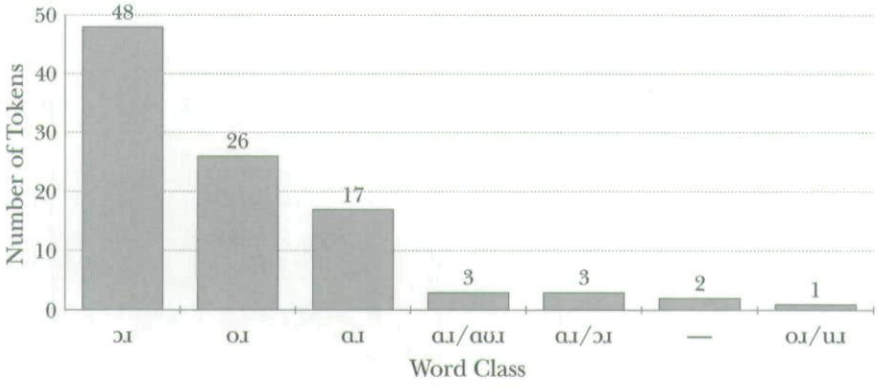
In addition, if we think of the object in terms of building a model to explain what acoustic features lead to Utahn /ɑɹ/ and /ɔɹ/ being perceived as merged or distinct, it seems preferable to try to build a model that shows as bright a line as possible between all categories. Searching for a bright line between variable categories, of course, might go too far, even if only because there's a widespread assumption that some overlap between variable categories is simply to be expected. However, there may well be clear lines between at least some categories, and so it is worth looking for them. Also, even if it turns out to be impossible to consistently draw clear boundaries between categories, the closer we can come to such a model, the closer we can come to making testable predictions about the perception of various productions of Utahn /ɑɹ/ and /ɔɹ/—or, in other words, the closer we will be to having a full-fledged theoretical description of this merger, which could help lead to a better description of mergers generally.

In order to do this, a good deal of exploratory probing is necessary, mainly to determine whether any models that are developed are robust and to make sure that the final model is the best possible model given the data. To this end, this article investigates a number of possibilities for explaining Utah's CARD-CORD merger in the hope that, by the end of this investigation, a model of the CARD-CORD merger will be developed and set up for future testing, while other possible models will have been eliminated from consideration. (Of course, as with any attempt to build a model, some possibilities will be left uninvestigated; some of these are discussed near the end of this article.)

3. THE BREAKDOWN OF THE SPEECH SAMPLE

Before getting to the investigation itself, though, a complication should be noted: as mentioned above, the speech sample the rating panel heard was not simply 20% /ɑɹ/ and 80% /ɔɹ/; the complete breakdown of the tokens was as shown in figure 1. Note that, actually, there were 17 tokens of the historical word class that would only ever be produced with an [ɑɹ] (such as the word *far*), 48 tokens of the historical word class produced only with an [ɔɹ] (such as stressed *for*), and 26 tokens of the historical word class produced only with

FIGURE 1
Number of Tokens in the Sample for Each Word Class Represented



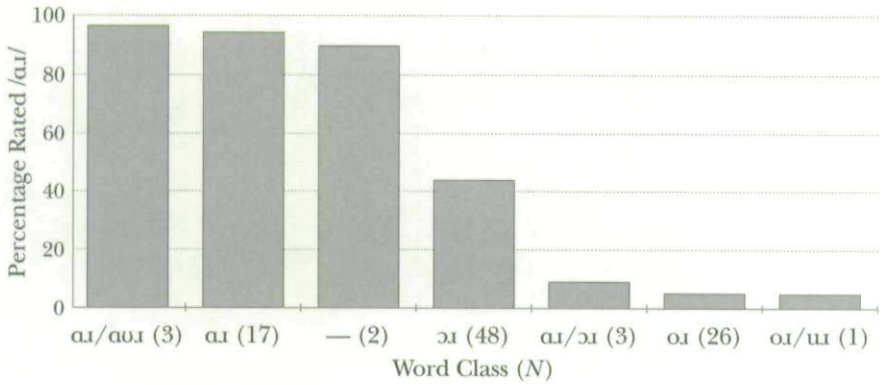
[oɪ] (such as *four*); for simplicity, I will refer to these historical word classes as, respectively, the (ɑɪ) class, the (ɔɪ) class, and the (oɪ) class for the remainder of this paper.[4] All other word classes had three or fewer tokens, meaning that generalizations drawn from them would almost certainly be meaningless.[5] (One class of two words, both of them names for which historical classes could not be determined, is represented in the figure by a dash.)

When considered as groups, these different word classes behaved differently, as shown in figure 2.[6] As you might expect from a study of perception of words containing underlying /ɑɪ/ and /ɔɪ/, some word classes gave rise to a high rate of agreement (90% and above) among the rating panel that the words in them contained an [ɑɪ], while others showed a high rate of agreement (again, above 90%) that the words contained [ɔɪ]. However, one word class—the (ɔɪ) class—resulted in a great deal of disagreement among the raters.

With that as background, we can now weed through the data to simplify the picture somewhat. First of all, as already mentioned, we can leave aside the four very small word classes—the number of words is too small to be able to make any sort of generalization about behavior by class for them. In addition, we can also make another, perhaps unexpected simplification—we can eliminate the (oɪ) class from further investigation. With the exception of one word (*wherefore*), there was 90% and higher agreement among the raters that every one of the words in this class contained [ɔɪ], which means that there's not much to talk about regarding the perception of these words as it relates to the issues at the heart of this paper.[7]

The (ɔɪ) class, on the other hand, is obviously of interest, because nearly all of the disagreement among the raters is found there—the overall rate at which raters believed that (ɔɪ) class words were produced with an [ɔɪ]

FIGURE 2
Percentage of Tokens Rated as [ɑɪ] for Each Word Class Represented



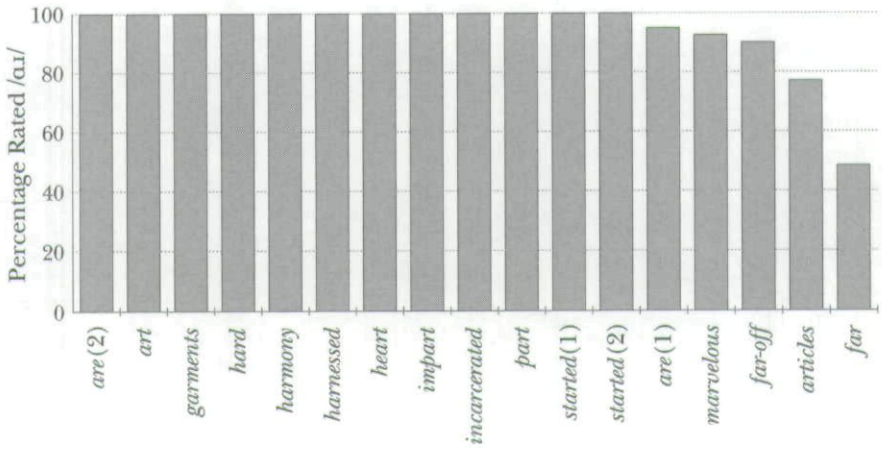
was only 56%, and there was a large amount of variation by word (as will be discussed shortly). However, even though the (ɑɪ) class did not have such variation (raters agreed that words in this class were produced with an [ɑɪ] 94% of the time), it must be included in the analysis. This is because previous studies, despite all their differences in the details, have agreed that Utah English's CARD-CORD merger is a merger of /ɔɪ/ into /ɑɪ/, meaning that we need to have information on the (ɑɪ) class in order to have anything meaningful to say about what's going on with the (ɔɪ) class.

To summarize, at this point we are dealing with two word classes, one of 48 words showing some odd variation and one of 17 words acting as a control to compare the larger class against. For the most part, however, all that has been presented to this point are pooled results for each class. Within the two remaining classes, however, an important question remains: is each class consistent across words, or is there variation among individual words within the classes?

4. PANELISTS' RATING OF (ɑɪ) AND (ɔɪ) CLASS WORDS

For the (ɑɪ) class, as you can see in figure 3, the stability is class-wide—almost.[8] There was a great deal of agreement among the rating panel throughout this class—in fact, 12 of the 17 words had 100% agreement. Of the remaining words, even the one with the second-lowest rate of agreement (*articles*) had 78% agreement on what sound the word contained. The word with the lowest rate of agreement (*far*), however, showed striking disagreement (51% versus 49%), and for that one the bare majority believed that it contained [ɔɪ], not the expected [ɑɪ]. Since there was such a high level of agreement

FIGURE 3
Ratings of (ɑɪ) Class Words as Containing [ɑɪ], by Word



about all of the words in the (ɑɪ) class except for *far*, the rest of this article will treat the (ɑɪ) class as if the word *far* was not a part of it. (I will, however, bring *far* back into the discussion later.)

The (ɔɪ) class of words is larger, and the panelists' ratings of it are quite a bit more complex, as you can see in figure 4. Responses from the rating panel range from two words with 98% agreement that the word was produced with an [ɑɪ] (*cornerstone* and *organization*) to six words with 100% agreement that they contained [ɔɪ] (*born*, the third instance of *for*, *importance*, the first instance of *important*, the second of *lord's*, and the first of *morning*).[9] The 40 words in between are fairly evenly scattered over nearly the entire range of possible results.

To simplify the remainder of the analysis somewhat, I focus first on the (ɔɪ) class words with more than 75% agreement one way or the other from the rating panel.[10] Those with more than 75% of the raters perceiving an [ɑɪ] are listed in table 1; those with the same level of agreement that they were produced with [ɔɪ] are given in table 2. While it is obvious that there are some interesting things going on in the middle 50%, focusing on the edges for the moment allows us to draw clearer distinctions.

At this point we really have three sets of words involved in our analysis: the 14 out of 48 tested words in the (ɔɪ) class that the rating panel generally agreed were produced with [ɑɪ], the 19 words in the (ɔɪ) class that the panelists generally agreed were produced with [ɔɪ], and 16 of the 17 tested words from the (ɑɪ) class.[11] For ease of reference, I will refer to the (ɔɪ) class words that the rating panel generally agreed were produced with [ɑɪ] as

FIGURE 4
Ratings of (ɔɹ) Class Words as Containing [ɑɹ], by Word

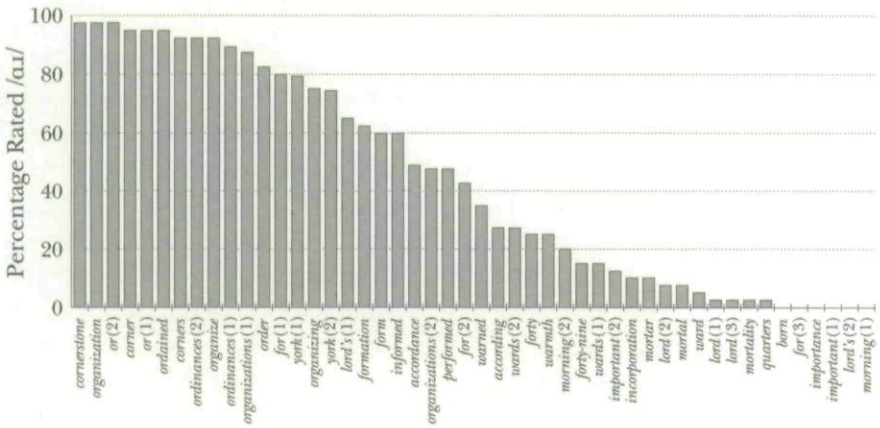


TABLE 1
(ɔɹ) Class Words Produced with [ɑɹ] According to More Than 75% of the Rating Panel

cornerstone	98%	*ordained*	95%	*organizations* (1)	88%
organization	98%	*corners*	93%	*order*	83%
or(2)	97%	*ordinances*(2)	93%	*for*(1)	80%
corner	95%	*organize*	93%	*york*(1)	79%
or(1)	95%	*ordinances*(1)	89%		

TABLE 2
(ɔɹ) Class Words Produced with [ɔɹ] According to More Than 75% of the Rating Panel

morning(1)	100%	*mortality*	98%	*mortar*	90%
lord's(2)	100%	*lord*(3)	98%	*incorporation*	90%
important(1)	100%	*lord*(1)	98%	*important*(2)	88%
importance	100%	*ward*	95%	*wards*(1)	85%
for(3)	100%	*mortal*	93%	*forty-nine*	85%
born	100%	*lord*(2)	93%	*morning*(2)	80%
quarters	98%				

the "merged (ɔɹ)" class (since, in the collective opinion of the rating panel, these words involve an application of the merger of /ɔɹ/ into /ɑɹ/), and the (ɔɹ) class words rated as produced with [ɔɹ] as the "unmerged (ɔɹ)" class (since there is no such merger in these cases).

5. INITIAL TESTS OF THE MERGER'S ACOUSTIC CHARACTERISTICS

Since the point of this study is to determine what exactly the acoustic characteristics of the CARD-CORD merger are, and since we now have classes of merged and unmerged words containing /ɔɹ/ (along with a control group of words containing /ɑɹ/), it seems worthwhile to look into the acoustic characteristics of these three groups. To this end, an F1-F2 plot of these three groups is given in figure 5, which shows more or less the mid back portion of this speaker's vowel space. At first glance, it certainly seems as if this plot shows real results that we can use for what follows. To begin with, the overlap between the (ɑɹ) class and the merged (ɔɹ) class vowels is nearly complete. (In fact, the means of the two sets of vowels are nearly identical for both formants.) On the other hand, there is a clearly visible separation between the (ɑɹ) class and the unmerged (ɔɹ) class vowels, making it seem that Utah's CARD-CORD merger is simply one of articulation that can be measured along these two axes.

That conclusion is supported by statistical testing; *p*-values from *t*-tests on the F1 and F2 values of each pair of vowel classes are given in table 3. Basically, there is clear statistical evidence that the unmerged (ɔɹ) class vowels are different from both the (ɑɹ) class and the merged (ɔɹ) class vowels, while the merged (ɔɹ) class vowels are not statistically different from the (ɑɹ) class ones.[12] Further, the fact that the unmerged (ɔɹ) class vowels generally have

FIGURE 5
F1-F2 Plot of Vowel Peaks by Word Class

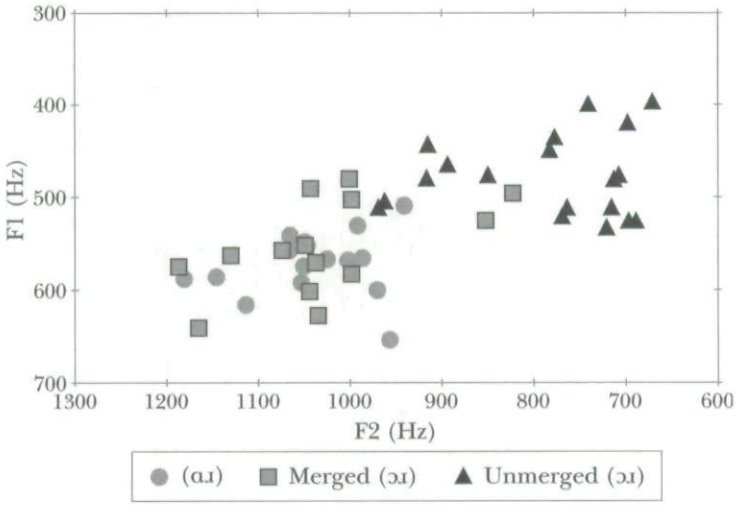


TABLE 3
P Values for Peak F1 and F2 Values for Each Word Class

	Merged (ɔɪ)	*Unmerged* (ɔɪ)
(ɑɪ)		
F1	.34116	< .00001
F2	.92296	< .00001
Unmerged (ɔɪ)		
F1	.00004	—
F2	< .00001	—

lower F1 and F2 values is consistent with Helquist's (1970) and my earlier views (Bowie 2003) that unrounding is involved in the merger.

This finding—that there appears to be a clear articulatory difference between the merged and unmerged (ɔɪ) class vowels—is extremely encouraging. Not only do we have some idea of what is going on with Utah's CARD-CORD merger—at this point it seems likely to be a fairly straightforward articulation issue, quite possibly unrounding—but it seems that we possibly even have some of the parameters of it. That is, if we were to think in terms related to the old idea of variable rules (see Paolillo 2002 for a discussion of this approach), we could posit that people run some kind of fuzzy logic in their heads as they perceive language, and so we could therefore at least partially describe this process statistically. This would give us the beginnings of a model of the perception of this merger (not a complete model, of course, but it would be a beginning). It would look something like this: there is a reference set of (ɑɪ) class vowels, and if an (ɔɪ) class vowel is produced that falls within the realm defined by first formant separation from the (ɑɪ) class measurable as $p \geq .34116$ and (or possibly or) second formant separation measurable as $p \geq .92296$, an instance of the merger is perceived. The actual cutoff levels in this possible model are tentative and may end up being different, of course, but that is what future study is for.

Unfortunately, if you look at this model more closely, it runs into trouble on two levels, both of which you can see in figure 6. This is essentially the vowel chart in figure 5, but with two additional bits of information. The first is that minimal ovals have been placed around each set of vowels, showing the range of each class. As we would expect, there is a lot of overlap between the (ɑɪ) class vowels and the merged (ɔɪ) class vowels—these classes are merged together, after all, and therefore it stands to reason that there would be nearly complete overlap. There is, however, also overlap between the unmerged (ɔɪ) class vowels and both of the other sets—in fact, at least two of the unmerged (ɔɪ) vowels look like they are more "merged" than some of the merged ones.[13]

The second new bit of information in figure 6 is the position of the vowel in the word *far*, which was mentioned earlier as the only (ɑɪ) class word not to achieve general agreement from the rating panel that it was produced with an [ɑɪ]. Its first and second formant values are shown by the ✖ on this chart—and, oddly enough, while it is near the edge of the set of vowels that were rated as being produced with [ɑɪ], it clearly falls within the boundaries of that set. These two issues are not good news for the model discussed above, because now, even though *t*-test results may have seemed to say that there was or was not a merger in any given circumstance, we do not have a situation where we can claim that the proposed model would work in a consistent manner—there is no clean line that individuals could use to differentiate what they perceive as [ɑɪ] or [ɔɪ]. Therefore, we have to cast our net more widely.

6. A CLOSER LOOK AT FORMANT VALUES

Since simple F1-F2 differences do not tell the whole story, we have to look at more formant values—basically, extending what has already been attempted. Unfortunately, this does not immediately give clearer results. Figure 7 shows the pitch plus each formant from the first through the fourth graphed against each other. (This chart includes, of course, the information from

FIGURE 6
F1-F2 Plot of Vowel Peaks by Word Class

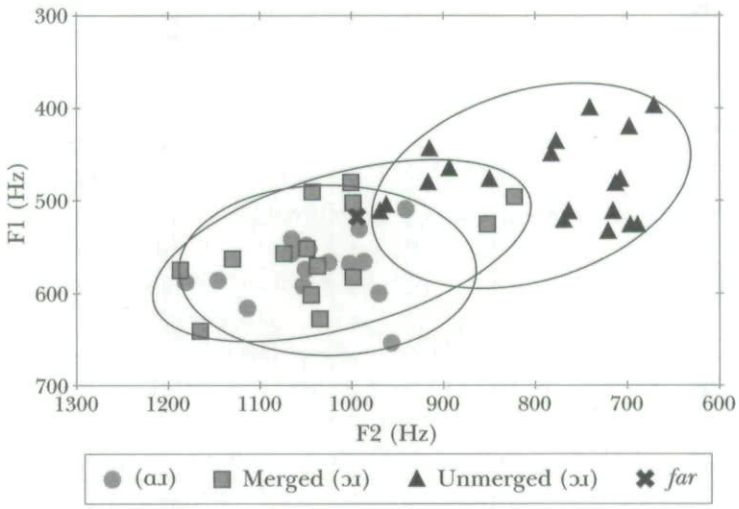
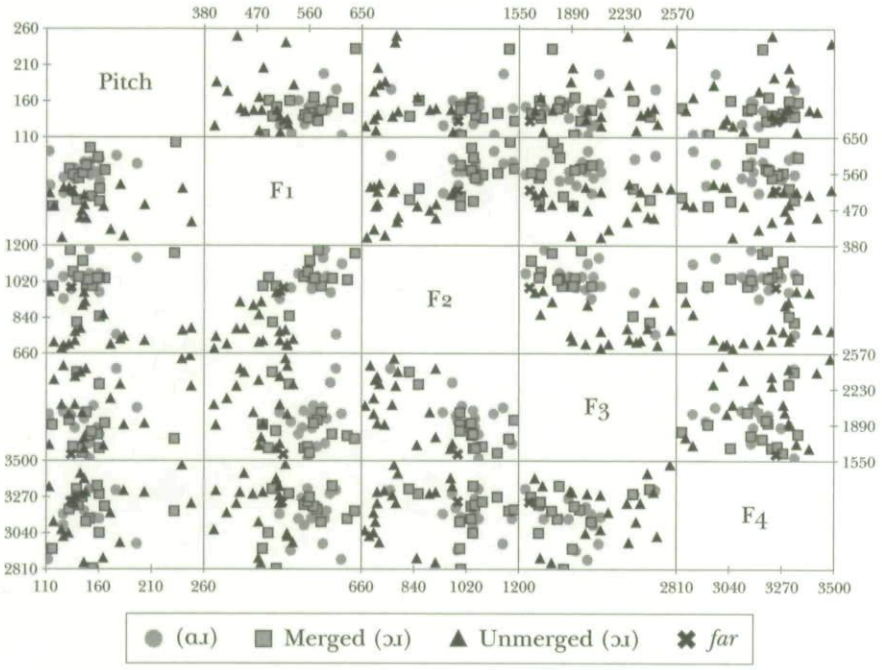


figure 5, since F1-F2 graphs are included as part of the matrix.) This is a rather complicated set of graphs, because it attempts to communicate all of the pitch and formant information analyzed here all at once, but it actually has a rather simple purpose: to allow the data to be eyeballed in the hope that we can come up with a formant pair that would allow us to draw a clean line between what the rating panel perceived as [ɑɪ] and [ɔɪ]. With that in mind, the message of figure 7 is simple: there is no formant pair that gives a clear separation between the unmerged (ɔɪ) class vowels on the one hand and anything else on the other—nor, for that matter, are there very many formant pairs that even manage to differentiate the word *far* from the (ɑɪ) class vowels.

It is worth noting, however, that a few formant pairs at least come close to differentiating these vowel classes completely—F1 versus F3, and to some extent F2 versus F3. Given that there are three formants in this list, it may be that part of the problem up to this point has been restricting our view of what is going on with the CARD-CORD merger to two dimensions—that is, so far we have been looking at the relationship between two variables (such

FIGURE 7
Matrix Plot of Pitch and F1 to F4 by Word Class

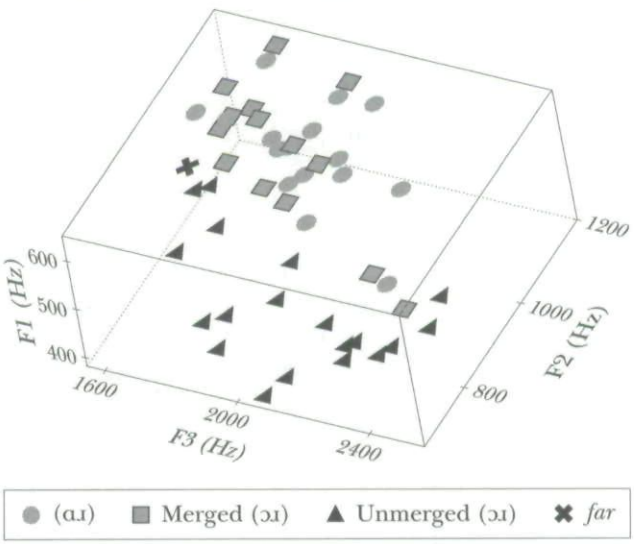


as F1 and F2, or F1 and F4, or whatever other pair). Therefore, it seems worthwhile to look at more than two things at a time.

Once we expand our view to three variables and look at the relationships among F1, F2, and F3 all at once, we get a much clearer picture, which you can see in figure 8. At this particular rotation of the axes in the three-dimensional space, and only at this rotation (which includes the one completely opposite it, of course), the (ɑɪ) and merged (ɔɪ) classes overlap rather completely, and yet are completely separate from the unmerged (ɔɪ) class. The unmerged (ɔɪ) class wraps around the others a bit at one end, so the separation is not necessarily a simple plane but a curved surface, but it would in any event be fairly simple to create a model of the surface delimiting the separation.[14] In addition, this plot shows the word *far* essentially at the boundary between those sounds perceived as [ɑɪ] and [ɔɪ]—a position that might nicely explain the rating panel's near fifty-fifty split on what the word sounded like.

Given that F3 turns out to be important for the description of the merger, though, it turns out that we cannot describe the merger as due simply to unrounding, as was claimed earlier: rounding results in the lowering of all formants (especially F2 and F3), and while F3 was generally lowered in the merged (ɔɪ) tokens when compared to the unmerged ones, F1 and F2 were both higher for the merged tokens. Thus, in articulatory terms, something more complicated is going on.[15]

FIGURE 8
F1-F2-F3 Plot of Vowel Peaks by Word Class



Focusing on the results for F2 and F3 illustrates the situation's complexity. While F3 is generally lower for the merged (ɔɹ) tokens than the (ɑɹ) and unmerged (ɔɹ) tokens, this is not consistently the case—the speaker seems to manipulate F3 across a wide range so that the positions of the tokens are spread out in the F1-F2-F3 acoustic space, allowing the separation of (ɔɹ) from (ɑɹ) to be maintained or lost (as the case may be for any particular token). As a result, a token of (ɔɹ) with a comparatively high F3 value might still end up merged, as long as it is also produced with a low F2 value; conversely, a token of (ɔɹ) with a comparatively low F3 value might remain unmerged, as long as it has a fairly high F2 value. Clearly, the phenomenon cannot be summed up in just one or two articulatory features (such as "rounding" or "lowering and unrounding"), as previous studies have tended to do.

One possibility is that there is some sort of persevatory coarticulation effect, since preceding labials and velarized /l/ correlate with lowered F1 and F2, and a glance through table 1 and table 2 makes it appear that there is an interaction between a preceding lateral and the merger process, in that a preceding /l/ seems to correlate with the absence of the merged variant. This is particularly interesting given that my previous work has found that the preceding sound strongly affected the application of the CARD-CORD merger during its early history (Bowie 2003). Unfortunately, not much is known about the effect of a consonant on a following vowel's F3, so it is impossible to say whether that is what is going on here; however, once more is known about this phenomenon, it may be worth looking at this study's data again to see whether persevatory coarticulation merits a closer look.

So, to summarize to this point, after a fair amount of testing, one method has been found (F1, F2, and F3 at the peak intensity of the vowel, all considered at once) that can be used to draw a distinction between the (ɑɹ) and merged (ɔɹ) classes on the one hand and the unmerged (ɔɹ) class on the other. This method has the additional advantage of plausibly explaining the word *far*'s anomalous behavior.

7. THE PROBLEM OF MERGER-INDUCED HOMOPHONY

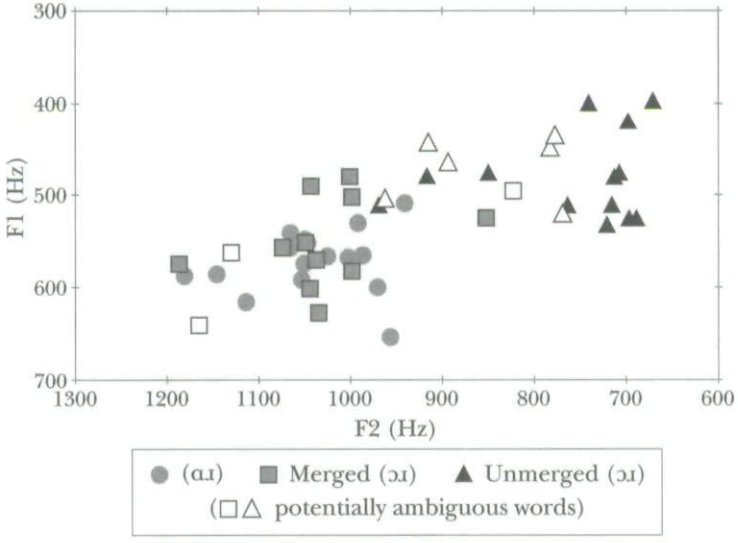
Although we now have a promising solution to the problem of detailing the phonetic characteristics of the Utahn CARD-CORD merger, it is worth stepping back to deal with a possible complication. As already discussed, there was no attempt to get a representative sample of words uttered by the speaker—words with stressed /ɔɹ/ or /ɑɹ/ were simply selected as they came up, with very few limits. However, among words in these classes, there are words that the CARD-CORD merger could turn into homonyms—for example, the word *born* would sound, when merged, like the word *barn*. On the other

hand, the word *organize* would, if merged, sound like *arganize*, which is not a word of English. It would be reasonable to expect that this would work to block words like *organize* from being perceived as merged, since listeners would know that the sound in the word would have to be [ɔɹ], while words such as *born* would have no such block from being perceived as merged.

Figure 9, then, is provided to give insight into whether merger-induced homophony is part of what is driving the rating panel's perceptions of merger. Figure 9 is the same as the F1-F2 plot in figure 5, except that the (ɔɹ) class words that would still make English words if the merger applied to them are represented by unfilled shapes. What we see is that these potentially ambiguous words are scattered throughout the space of the (ɔɹ) class words, and that they are perceived by the rating panel as both merged and unmerged. In fact, twice as many are perceived as unmerged, which gives the impression that this is not simply a case of the rating panel being confused by some words being in some way ambiguous.

Similarly, merger-induced homophony does not work as an explanation when taking F1, F2, and F3 values into consideration all at once, as you can see in figure 10, which is a version of figure 8 with information added to show which words could be subject to merger-induced homophony. In figure 10, those (ɔɹ) class words that are potentially subject to merger-induced

FIGURE 9
F1-F2 Plot of Vowel Peaks by Word Class
with Potentially Ambiguous Words Highlighted

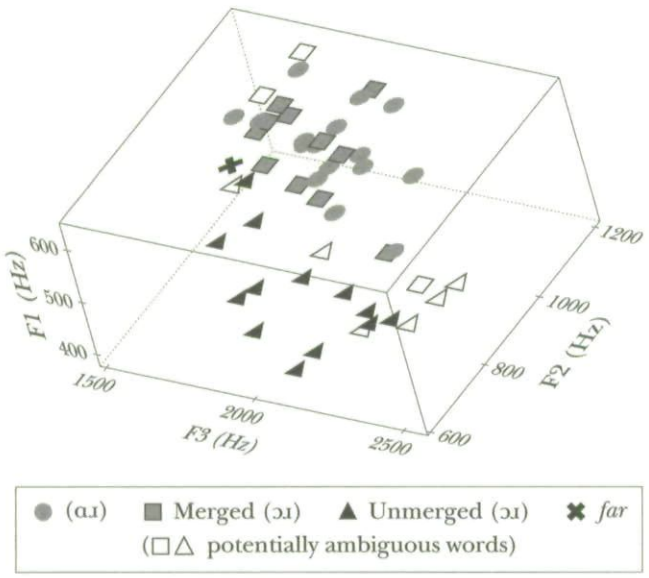


homophony are highlighted by being shown as hollow shapes. Perhaps oddly, the unmerged (ɔɪ) class words potentially subject to merger-induced homophony appear startlingly close to the border between those words generally perceived as [ɔɪ] and those perceived as [ɑɪ], while the few such words in the merged (ɔɪ) class appear at both ends of the space taken up by the words perceived as [ɑɪ]. However, no matter the relative distributions within each set, the result is clear: once again, merger-induced homophony cannot be used to clarify any of the earlier findings. Essentially, it appears that an analysis of this merger based on assumptions that functional concerns are driving the perception of the merger simply will not work.

8. TESTING THE "INTERMEDIATE" CASES

At this point there is a good deal of evidence that Utah's CARD-CORD merger can best be described as a process involving speakers' simultaneous manipulation of F1, F2, and F3 (or, to put it probably more correctly, it is speakers' manipulations of those articulatory processes that are acoustically reflected by differences in F1, F2, and F3). However, before making such a claim

FIGURE 10
F1-F2-F3 Plot of Vowel Peaks by Word Class
with Potentially Ambiguous Words Highlighted



definitively, we must return to a set of words that I temporarily excluded from analysis earlier in this article—those (ɔɹ) class words that the rating panel did not reach a high level of agreement on, with 25–75% of the panel rating them as containing an [ɑɹ] or [ɔɹ]. For convenience, I call these the "intermediate (ɔɹ)" class; they are listed, along with the ratings the panelists gave them, in table 4,[16] and their formant values are shown in figure 11 (which is the same chart as figure 8, but with the intermediate (ɔɹ) class now included).

In a perfect world, figure 11 would have had all of the intermediate (ɔɹ) cases laid out along the boundary separating the merged (ɔɹ) and (ɑɹ) classes on the one side and the unmerged (ɔɹ) class on the other. This is not what we have, though—the intermediate (ɔɹ) class does cluster along that border, but there is some visible scattering. However, there is one promising facet of the behavior of the intermediate (ɔɹ) cases that is immediately visible: they are completely separate (in the F1-F2-F3 space) from the unmerged (ɔɹ) class. This means that, if we are to find what separates the intermediate (ɔɹ) class from the other classes, it would appear that we only need to find what separates it from the (ɑɹ) and merged (ɔɹ) classes, since we already know what separates it from the unmerged (ɔɹ) class.

Unfortunately, no tests comparing formant values provided any worthwhile results when they were used to compare the (ɑɹ), merged (ɔɹ), and intermediate (ɔɹ) classes. We can see this graphically in figure 12, which is a chart showing the values for the pitch and the first through fourth formants for each of these classes; as it turns out, unlike the situation seen in figure 7, there is no formant or set of formants that even appears to begin to distinguish between the classes. Basically, it seems like the addition of the intermediate (ɔɹ) class paints the entire analysis into a corner that it is impossible to escape—it seems that there is no way to come up with a model that allows us to understand exactly what listeners are relying on to classify words involved in Utah's CARD-CORD merger.

However, at this point it's worth stepping back a bit and taking a closer look at some of the individual words in the sample. As previously mentioned,

TABLE 4
(ɔɹ) Class Words Produced with [ɑɹ] According to 25–75% of the Rating Panel

organizing	75%	*informed*	60%	*warned*	35%
york(2)	74%	*accordance*	49%	*according*	28%
lord's	65%	*organizations*(2)	48%	*wards*(2)	28%
formation	63%	*performed*	48%	*forty*	25%
form	60%	*for*(2)	43%	*warmth*	25%

FIGURE 11

F1-F2-F3 Plot of Vowel Peaks by Word Class, Including Intermediate (ɔr)

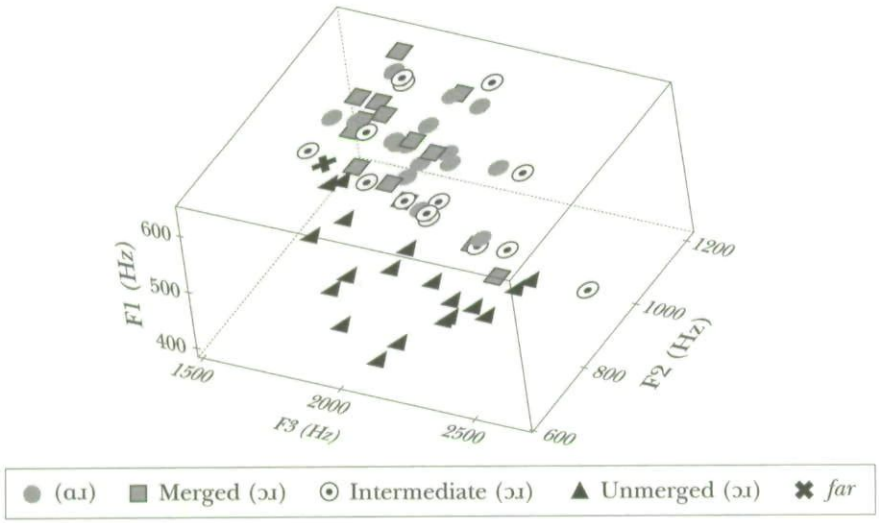


FIGURE 12

Matrix Plot of Pitch and F1 to F4 by Word Class, Including Intermediate (ɔɹ)

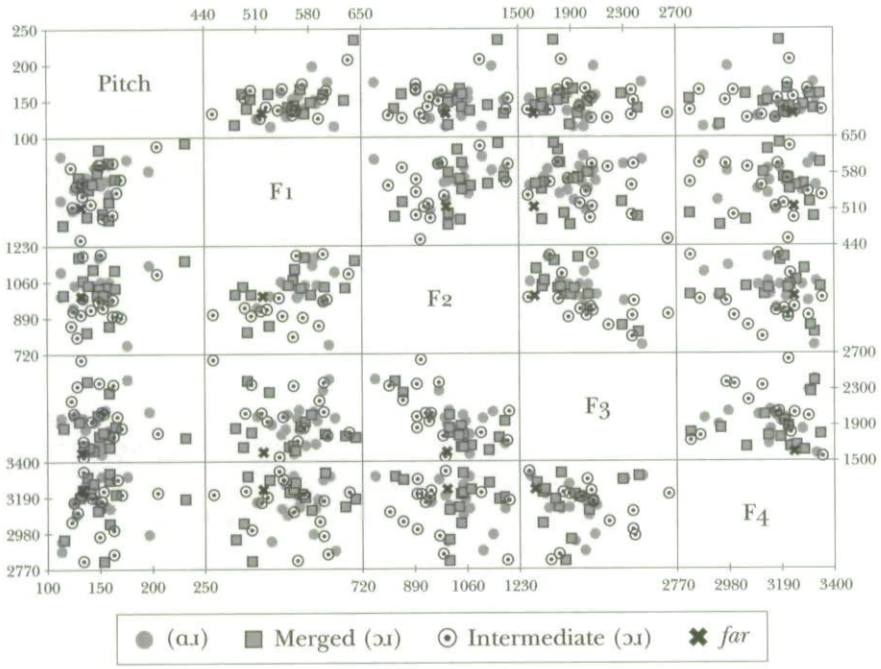


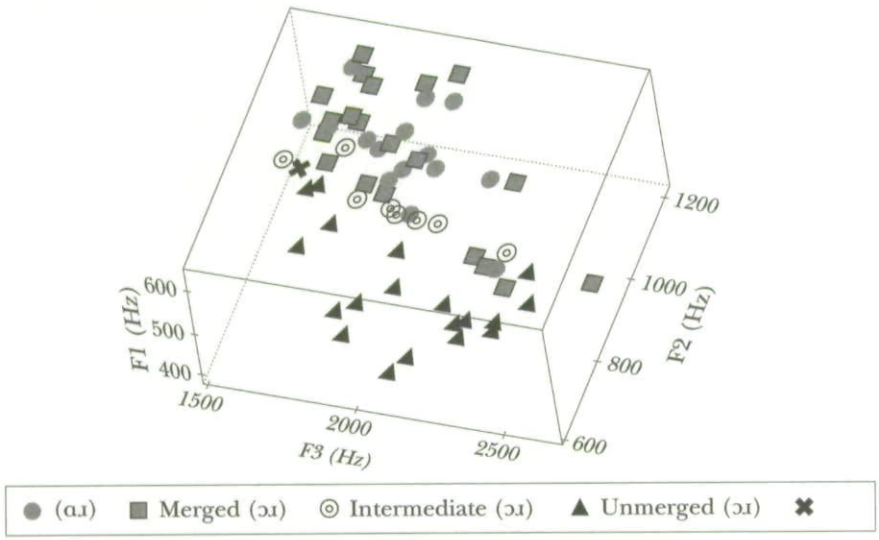
figure 11 shows that most of the vowels in the intermediate (ɔɪ) class words cluster along the border between the (ɑɪ) and merged (ɔɪ) class vowels on the one hand and the unmerged (ɔɪ) class vowels on the other. If all of the intermediate (ɔɪ) class words fell along that border, we could conclude that (ɔɪ) class vowels produced along the border are perceived as merged or intermediate. However, there are five intermediate (ɔɪ) class words that fall away from that border—and the actual identity of those five words provides us with a possible solution to our conundrum.

The five intermediate (ɔɪ) words that do not fall neatly along the border between the words generally perceived as [ɑɪ] and those generally perceived as [ɔɪ] are, in alphabetical order: *form, formation, lord's, organizing,* and the second instance of the word *york.* A look at table 4 gives a startling result: These five words make up nearly the entire list (the exception is *informed*) of intermediate (ɔɪ) class words that more than half of the rating panel perceived as being produced with [ɑɪ]!

This finding allows us to try something a bit more complicated, but that ends up leading in positive directions—we can play for a moment with the idea that we might actually have two subclasses of intermediate (ɔɪ) class words, based on the rating panel's perceptions: an intermediate (ɔɪ) class made up of those words rated as containing [ɑɪ] by 50–75% of the rating panel,[17] which nearly all fall into precisely the same space as the merged (ɔɪ) vowels; and an intermediate (ɔɪ) class made up of words rated as containing [ɑɪ] by 25–50% of the rating panel, which fall along the border between the unmerged (ɔɪ) and (ɑɪ)/merged (ɔɪ) vowels. This results in four subclasses of (ɔɪ) class words, giving a total of five classes when we include the (ɑɪ) class, plus the outlying word *far.*

Upon further reflection, however, this turns out to be an unnecessary complication, since we can actually collapse categories so that we only need to deal with a total of four (though divided differently than those we had previously), not five, plus the word *far.* If we classify the intermediate (ɔɪ) class cases where more than half of the rating panel rated them as containing [ɑɪ] as not, in fact, part of an intermediate (ɔɪ) class at all, but as cases of merged (ɔɪ), we have three classes that behave nearly precisely as we would like to see—the newly redefined merged (ɔɪ) class overlaps entirely with the (ɑɪ) class, the unmerged (ɔɪ) class is completely separate from the other classes, and the newly defined intermediate (ɔɪ) class sits on the border between the other two. A revised F1-F2-F3 chart showing this distribution is given in figure 13.

FIGURE 13
F1-F2-F3 Plot of Vowel Peaks by Word Class, with Revised Intermediate (ɔɹ) Class



9. DISCUSSION AND CONCLUSIONS

After all of these convolutions, what is the point of it all?

The initial question posed in this article about the CARD-CORD merger was: given that the CARD-CORD merger exists in Utah, what exactly is going on when it occurs? From this study, it appears that what is going on is the articulatory manipulation of certain tokens of /ɔɹ/ such that they fall together with /ɑɹ/ in a way that is reflected acoustically in a relatively higher F1 and F2 and lower F3 in the merged tokens. The exact process that results in these formant values is unclear, but it does not seem that it is as simple as previous studies have made it out to be—in all likelihood, there are multiple articulatory processes going on at once (possibly involving the simultaneous manipulation of articulations as different as lip rounding and palatal constriction). In addition, there is a possibility (pending future research) that persevatory coarticulation, particularly when there is a preceding /l/, plays a role in at least limiting the application of the merger.

To come at that from a somewhat different direction, this study clarifies what features are and are not useful in determining what exactly, in phonetic terms, is going on with Utah's CARD-CORD merger. As we might have expected, vowel formants measured at the point of peak intensity are useful for describing the differences between the (ɑɹ) and merged (ɔɹ) classes on

the one hand and the unmerged (ɔɹ) class on the other, as well as explaining the behavior of the intermediate (ɔɹ) class vowels. However, as it turns out, the first three formants are needed to draw a clear distinction, rather than the more commonly considered first two. Once those three formants are taken into consideration, the clearly opposing groups are separated by a nearly planar surface, with the interesting effect that an (ɑɹ) class word at the boundary might be widely (though not generally) perceived as containing [ɔɹ], and an (ɔɹ) class word at the boundary might be relatively often (though not widely) perceived as having an [ɑɹ]. Meanwhile, potential merger-induced homophony, which common sense would lead one to believe would have an effect on the perception of the merger, does not appear to affect perception of any of the groups in any meaningful way.

Of course, with these conclusions come some caveats. First of all, and perhaps most importantly, this study looked at the production of only one speaker, so it is unclear if all speakers from that speech community treat the merger the same. In addition, the panelists had had no contact with Utah English, and it would be worthwhile to find out if Utahns, particularly those from along the Wasatch Front (where the CARD-CORD merger has been most clearly documented in Utah—see in particular Cook 1969; Helquist 1970), perceive the merger in the same way as those from elsewhere.

In addition, this study was concerned primarily with acoustic correlates of articulatory realities and how they related to perception. In doing so, however, it looked only at pitch and F1 through F4, even though there are more articulatory processes than that,[18] including voicing processes (e.g., creakiness or breathiness, which were not considered at all in this study) and pitch contours (pitch was analyzed, but only at individual points, not across the entire segment). Also, the sample of the speaker's speech was not representative, and so the question of whether all instances of merged (ɔɹ) are treated the same way needs to be investigated more deeply, even though they were for all of the words in that class in this study.

Further, it is necessary to keep in mind that the speaker whose production was analyzed here is part of one particular speech community where the CARD-CORD merger is clearly (though variably) present. Therefore, this speaker's method of distinguishing those instances of /ɔɹ/ that remained unmerged from the other sounds may not be the same method that someone from a different speech community with the same merger, or for that matter from a speech community without the merger, would use. For example, whereas the speaker analyzed here manipulated the articulatory correlates of F1, F2, and F3 to produce the various sounds related to the CARD-CORD merger, a speaker from another speech community might well articulate the sounds involved in the merger in such a way that the important acoustic features

are just F1 and F2, or even just F1, or some other constellation of features. Therefore, what is presented here is not a universal model for everywhere the CARD-CORD merger is found, but rather a model highlighting a particular way in which mergers and their attendant distinctions can occur.

Despite these caveats and questions, however, I would like to stress that this study does clearly offer a methodological caution for those studying linguistic variation. When we are trying to get to the phonetic and articulatory facts to answer a "What exactly is going on?" type of question, we often immediately resort to simple measures of F1 and F2, possibly enriched by statistical analysis, to get an answer. However, we need to reach beyond that sort of a solution, even when (as was the case in this study) the initial statistical testing seems to support the claim that the first two formants are all that is needed to resolve the issue.

Going further, this study gives some insight into profitable future directions for investigating processes of merger generally. At least for the merger discussed here, clear boundaries could be drawn between the [ɑɹ] and [ɔɹ] realizations of the phonemes under study. At some level, though, it could be seen as surprising that F1, F2, and F3, when taken together, would allow clear boundaries to be drawn between tokens generally perceived by the rating panel as [ɑɹ] and [ɔɹ]. After all, most analyses of vowel articulation that look at formant values accept some overlap among phonemes as a matter of course—the crucial thing is whether the target of each phoneme is the same or different.

In fact, this is what seemed to be happening at the beginning of this analysis—when we were looking at only F1 and F2, there was overlap between the (ɑɹ)/merged (ɔɹ) and unmerged (ɔɹ) classes. However, given that the rating panel heard the two classes differently and given that the panel was not relying on contextual clues such as merger-induced homophony to distinguish the classes, it makes sense that there really was a clear (and audible) boundary between them. This is particularly supported by the fact that there was such disagreement among the panel members about what to do with the word *far* and the intermediate (ɔɹ) class—since they lay on the boundary between those perceived as [ɑɹ] and those perceived as [ɔɹ], it makes sense that they would be perceived as neither clearly one nor the other.

I do not claim, of course, that all mergers behave in the same way as the merger analyzed in this study. This statement applies on two levels. The first, and more trivial, is that I do not claim that all mergers are the result of sounds crossing a boundary in F1-F2-F3 space—in fact, each merger is likely to be the result of a different set of articulatory factors working together in different ways. Second, I do not claim that all mergers are the result of crossing some sort of bright-line articulatory boundary, however complex—but

given that something like that seems to be at work in this case, it does seem worth investigating the possibility more closely.

In general, some overlap between categories in studies of linguistic variation has traditionally been seen as simply part of the system. However, those of us working with such data might do well to look more closely to see if clear boundaries can actually be drawn between categories. If, in the end, it turns out that such clear boundaries actually cannot be consistently drawn, then we will simply have to go on living with the fuzziness that that entails. However, if we find that individuals actually do draw clear distinctions between categories, we will be able to develop more reliable and replicable models of linguistic variation. In any event, though, whether we end up confirming or refuting the possibility that linguistic categories are clearly bounded, investigating the possibility would lead us to a more complete understanding of the processes underlying variation in language generally, and so it merits wider consideration.

NOTES

There are a number of people who have helped in some way with this article. However, I would particularly like to thank Melody Bowdon, Mariana Chao, and Stephanie Colombo for their input into very early versions of this paper, as well as the anonymous reviewers for a number of excellent suggestions that have resulted in this article being much improved over what it otherwise would have been.

1. The use of the term "Utah English" sometimes creates some controversy, since trying to describe what is spoken in all of Utah as a single monolithic and separate variety is doomed to failure, as Utah's borders do not coincide with barriers to linguistic diffusion, and Utah certainly contains multiple varieties (see Lillie 1998; Carver 1987). I am, however, using "Utah English" here the same way it is usually used in the literature: as a shorthand for English as spoken along the Wasatch Front and immediately adjacent areas of northern Utah, which is and consistently has been home to the overwhelming (i.e., greater than 80%) majority of Utah's English-speaking population.
2. No attempt was made to get a representative sample of words containing /ɔɪ/ or /ɑɪ/ in this individual's speech—words with stressed /ɔɪ/ or /ɑɪ/ were taken as they came up. However, no more than three tokens of any single word were selected for the sample. Note that this restriction was made by word, not by morpheme, since the implications for merger-induced homophony are different for different words. For example, *forty* with /ɔɪ/ merged into [ɑɪ] would presumably be easily perceived as the impolite but real word *farty*, but *forty-nine* with the same merger could reasonably be expected to be less easily perceived as the nonword *farty-nine*.
3. Note that pitch, not Fo, was measured for these segments, and was obtained using Praat's built-in routine for measuring pitch.

4. For further discussion of historical word classes, see the work of Labov, Yaeger, and Steiner (1972), which covers the topic at some length. The present study does not follow their methods precisely (mainly because this study divides classes based on both historical classification and perceived differences within such historical classes), but the parallels should be clear.
5. For completeness, the words in the small classes are listed here: the (ɑɪ/ɑʊɪ) class contained *ourselves* and two instances of *our*; the (ɑɪ/ɔɪ) class contained *authority*, *foreign*, and *moral*; the (—) class contained *McCord* and *Orson*; and the (oɪ/uɪ) class contained *course*.
6. For reasons of completeness and consistency, this chart includes the word classes that are small enough to prohibit meaningful generalizations. The larger word classes—words with sequences historically produced as [ɑɪ], [ɔɪ], and [oɪ]—are the only ones you need to pay close attention to, however.
7. The issue of exactly why and how the (oɪ) class is perceived differently than the (ɔɪ) class is certainly of interest, since it raises the possibility that the HOARSE-HORSE distinction survives in Utah and is being perceived as such by a rating panel that does not maintain the distinction. This falls well beyond the bounds of the current study, however, but as it is both intriguing and uninvestigated (see, for example, Lillie 1998, where it is implicitly assumed that the distinction has collapsed, but no evidence is provided for the assumption), it is something I plan to investigate more deeply in the future.
8. In figure 3 and similar charts, words with a number in parentheses are words that were in the speech sample more than once, with the number giving the order the words occurred in. Therefore, "*started*(1)" is the first instance of *started*, and "*started*(2)" is the second instance of that word.
9. It is worth noting that *cornerstone* and *organization*, both generally perceived as produced with [ɑɪ], do not have parallel words containing [ɑɪ], while *born*, *for*, and *lord's*, which were universally perceived as produced with [ɔɪ], do (*barn*, *far*, and *lard's*). This provides some evidence that the members of the rating panel were in fact rating according to whether they heard [ɔɪ] or [ɑɪ], not according to expectations given the content of the rest of the word. This is discussed in more detail elsewhere in the article.
10. I readily admit that 75% is an arbitrary line (and one that will be altered somewhat later in the paper), but it is a good one to use as a starting point. First of all, if more than three-quarters of the rating panel agree that a particular word was produced with a particular sound, it is intuitively clear that the panel's preference is not merely accidental. Also, a 75% threshold divides the (ɔɪ) class into three similar-sized chunks: 14 generally perceived as produced with [ɑɪ], 19 generally perceived as produced with [ɔɪ] (including 6 universally perceived that way), and 15 with a more mixed perception.
11. The 17th (ɑɪ) class word (*far*) and the 15 (ɔɪ) class words with less than 75% agreement might be thought of as fourth and fifth sets, since they will be reintroduced later.
12. This assumes the arbitrary but conventional limit of $p < .05$ as the threshold for determining statistical significance.

13. I should note here that I recognize that the relationship between formant frequencies and the perception of sounds is complicated, even though I am treating them here as fairly straightforward. However, if it is possible to find a simple, straightforward correlation between formants and perception, it seems worthwhile to pursue it. Therefore, in the interest of developing a simple model, I will continue to act as if the relationship is simpler than it might be, while recognizing that this means that the model may well need to be refined later on.
14. There are other possibilities involved in how to describe this model—for example, one could imagine the possibility of reducing this again to two dimensions, where (to choose at random) F3 minus F2 against F3 minus F1 provided a clear separation. However, testing of these possibilities revealed that that sort of simple transformation does not provide the separation that is visible at three dimensions.
15. F1 is widely recognized as being most closely correlated with vowel height, and F2 with backness. The articulatory correlates of F3 are less completely understood, but studies have found a relationship between F3 and both anterior tongue position and palatal constriction (Ong and Stone 1998). The precise nature of the sampled speaker's articulation is less important for this study than the formant measurements, however.
16. Table 4 only lists the percentage of the panel that rated each word as containing [ɑɹ]; of course, the rate at which each word was perceived as containing [ɔɹ] can be arrived at by simply subtracting the percentage given from 100%.
17. This division results in the word *informed* falling into this class. However, since as it is some of the merged (ɔɹ) class vowels do in fact fall along the border between the merged (ɔɹ) and unmerged (ɔɹ) vowels—which is simply at one edge of the merged (ɔɹ) space—this creates no real problem.
18. One of these other articulatory processes was looked at, as noted earlier in this article: length (in terms of vowel length, length of the following [ɹ], and the ratio of the length of each vowel to that of its following [ɹ]). However, no meaningful results could be found correlating perception and produced length.

REFERENCES

Boersma, Paul and David Weenink. 2005. *Praat: Doing Phonetics by Computer*, version 4. Available from http://www.praat.org/.

Bowie, David. 2003. "Early Development of the CARD-CORD Merger in Utah." *American Speech* 78: 31–51.

Carver, Craig M. 1987. *American Regional Dialects: A Word Geography*. Ann Arbor: Univ. of Michigan Press.

Cook, Stanley Joseph. 1969. "Language Change and the Emergence of an Urban Dialect in Utah." Ph.D. diss., Univ. of Utah.

Helquist, Val J. 1970. "A Study of One Phonological Variable in Urban and Rural Utah." M.A. thesis, Univ. of Utah.

Krahnke, Karl J. 1979. Sociolinguistic Variation and the Vowel System of Northern Utah: A Preliminary Look." In *Language and Linguistics Symposium 1979, April 5–6, 1979: Brigham Young University*, ed. Cheryl Brown, 6–11. Provo, Utah: Deseret Language and Linguistics Society.

Labov, William, Sharon Ash, and Charles Boberg. 2006. *Atlas of North American English.* New York: Mouton de Gruyter.

Labov, William, Malcah Yaeger, and Richard Steiner. 1972. *A Quantitative Study of Sound Change in Progress.* 2 vols. Philadelphia, Pa.: U.S. Regional Survey.

Lillie, Diane DeFord. 1998. "The Utah Dialect Survey." M.A. thesis, Brigham Young Univ.

Lisker, Leigh, and Arthur S. Abramson. 1970. "The Voicing Dimension: Some Experiments in Comparative Phonetics." In *Proceedings of the Sixth International Congress of Phonetic Sciences, Held at Prague, 7–13 September 1967*, ed. Bohuslav Hála, Milan Romportl, and Premysl Janota, 563–67. Prague: Czechoslovak Academy of Sciences.

Ong, Darryl, and Maureen Stone. 1998. "Three Dimensional Vocal Tract Shapes in /r/ and /l/: A Study of MRI, Ultrasound, Electropalatography, and Acoustics." *Phonoscope* 1: 1–13.

Paolillo, John C. 2002. *Analyzing Linguistic Variation: Statistical Models and Methods.* Stanford, Calif.: Center for the Study of Language and Information.

Pardoe, T. Earle. 1935. "Some Studies of Rocky Mountain Dialects." *Quarterly Journal of Speech* 21: 348–55.

Yaeger, Malcah. 1975. "Speaking Style: Some Phonetic Realizations and Their Significance." *Pennsylvania Working Papers on Linguistic Change and Variation* 1: 1–60.