Uighur Vowels: An Acoustic and Perceptual Study
Author(s): Collin F. Baker
Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society: General Session Dedicated to the Contributions of Charles J. Fillmore (1994), pp. 68-81

Uighur Vowels: An Acoustic and Perceptual Study

Collin F. Baker
University of California at Berkeley

Introduction

The Uighur language of Northwest China, with at least eight to ten million speakers, is Turkic, closely related to neighboring Kazakh and Uzbek and partially mutually intelligible with modern Turkish. One might suppose, therefore, that, like modern Turkish, it would have eight vowels, symmetrically ±high, ±back, and ±round. Like Turkish, Uighur has vowel harmony, with some vowels underspecified to various degrees and, furthermore, in Uighur the harmonic features can spread from either stem to affix or affix to stem, so it is not too surprising that there should be disagreement about the number of underlying vowels. Some authors, such as Hsu (1992) treat Uighur as having eight vowels, but others, such as Hahn (1991:33) regard it as a nine-vowel system, including three degrees of height in the front. Fig. 1 compares these two analyses:

(Hsu 1992)	front		back	
round	-	+	-	+
high	i	ü	ɨ	u
low	e	ö	a	o

(Hahn 1991)	front		back	
round	-	+	-	+
high	i	ü	ɨ	u
mid	e	ö		o
low	ä		a	

Fig. 1. Two analyses of Uighur vowels

In working with an informant who speaks the Kashgar dialect, we found that we had difficulty in transcribing certain vowels in a way that would accord with the informant's judgments, particularly with regard to [e] vs. [ä] (approximately IPA [ɛ] vs. [æ]). Hahn says that [e] "occurs almost exclusively in certain roots that are derived from foreign words and proper names, where such roots have not yet been 'nativized', i.e., have not yet been made to conform to native phonological principles..." (1991:37), but [e] seems to be reliably found in at least a few native words, such as eqiš 'flowing' and eǧiz 'mouth'. There are two symbols in the Uighur alphabets for these vowels (see Fig. 2.), yet there is also considerable variation in the pronunciation of words containing them. In some cases, [e] occurs in informal speech where careful speech has [ä].

To help resolve the question of whether we were dealing with an eight or a nine-vowel system, an acoustic study was performed . After the acoustic data had been analyzed, a perceptual experiment was also performed, to see how the informant himself and other hearers (non-native) would classify tokens of all the vowels. The results indicate that there are nine vowels in Uighur, at least in the idiolect of our informant, although the issue is not completely resolved. More generally, the results suggest that the common practice of plotting the vowels of a language in the F1-F2 plane may not adequately represent them or the phonetic cues which native speakers use to distinguish them, especially in languages where rounding is a distinctive feature.

kona yeziq (isolated form)	yeŋi yeziq	Hahn (1991) phonemes	IPA	graphs in this paper
ئى	i	i	i	i
ئې	e	e	e	e
ئە	ə	ä	æ	A, ae
ئا	a	a	ɑ	a
ئو	o	o	o	o
ئۇ	u	u	u	u
ئۈ	ü	ü	y	U
ئۆ	ɵ	ö	ø	O
	i	ɪ	ɨ	I

Fig. 2. Alphabets for writing Uighur

Theories of vowel quality

Theories of vowel quality can be divided into two broad classes, "target" models and "non-target" models (Strange 1989). The traditional theories are target models, based on the idea that for each vowel of a language there is some acoustic target, some set of values of acoustic parameters which constitute the perfect realization of that vowel. Individual tokens of vowels in utterances are thus better or worse examples depending on how well they approximate those values. This view can be dated back at least to Wilhelm von Helmholz; and was reinforced by the advent of the sound spectrograph in the 1930s and 40s, which made possible more accurate measurement of the fundamental frequency (f_0) and the formant frequencies (F1, F2, F3, etc.). Since then, many linguists have believed that the basic determinants of vowel quality are F1 and F2, and that these correspond roughly to the inverse of vowel height and frontness, respectively (Joos 1948).

One problem with defining vowels in these terms is that formant frequencies depend not only on the positions of the articulators, but also on the size and shape of the vocal track itself. Peterson and Barney (1952) measured the frequencies of the first three formants of American English vowels of 76 men, women, and children. They found that the vowels for an individual speaker could usually be distinguished reasonably well on a plot in the F1-F2 plane, but that the vowel spaces of different speakers overlap considerably. In particular, men, women, and children have substantially different vowel spaces, so that a given point in F1-F2 space might represent one vowel for a man and another for a child.

There has been considerable debate on the question of how we adjust our perception of vowels to compensate for these differences among speakers (a process known as "speaker normalization"). One suggestion has been that vowels are distinguished mainly on the basis of the ratios of formant frequencies rather than the absolute frequencies themselves. These ratios are much more constant both across individual speakers, and across groups such as men, women, and children. But different vowels, with quite different formants, may have identical formant ratios (even within speakers), so we cannot simply discard formant frequencies in favor of formant ratios (for a summary of the history of formant ratio theories, see Miller 1989).

Furthermore, the theory that there is a single target for each vowel, whether it is defined in terms of formant frequencies, their arithmetic differences or their ratios, has been challenged and become more elaborated as a result. For example, Stevens and House (1963) showed that the consonants preceding and following a vowel affect the formant frequencies. Lindblom (1963) found similar effects for Swedish, and also that the effect increases with rate of speech; he suggested that this results from the speaker's failure to reach the articulatory target (called "target undershoot"), especially in rapid speech.

Yet speakers of a language do rather well at identifying even isolated vowels, even when tokens produced by several speakers are presented in a random sequence (Assman, Neary, and Hogan 1982), which seems surprising given the above findings. If the formant frequencies of vowels (and their ratios) vary depending on the speaker, the surrounding consonants, and the rate of speech, how can native hearers hear them apart? Miller (1989) is a careful attempt to find acoustic parameters that will reliably distinguish the tokens of the vowels of American English. Citing research on auditory perception, Miller argues for describing the vowels in a perceptual space defined in terms of the $\log_{10}$ of the ratios F3/F2, F2/F1 and one additional ratio relating F1 to the mean f_0, which is supposed to aid in speaker normalization.

Recently, a number of researchers have suggested that vowels should not be described in terms of articulatory/acoustic targets at all. For example, Strange *et al.* (1976) showed that vowels in a consonantal context are correctly categorized much more often than vowels in isolation, and Strange *et al.* (1983) found that vowels from which the "steady state" central portion had been excised, leaving only the transitions into and out of the vowel, were recognized as well as vowel centers from which preceding and following transitions had been removed. This suggests that the dynamic information contained in C-V and V-C transitions is at least as important as the "steady-state" information as defined in traditional target models. On the other hand, Rakerd and Verbrugge (1985), using isolated vowels and /dVd/ syllables, and asking subjects for ratings of similarity, found two perceptual dimensions (D1 and D2) which corresponded fairly well with the F1 and F2 of the vowels. Rackerd and Verbrugge found that duration was also a significant factor in vowel discrimination, even among English monophthongs.

Assman, Neary, and Hogan's (1982) article is perhaps the most directly relevant to the present study. They performed both an acoustic analysis of vowel tokens and a perceptual study in which the same tokens were classified by a group of native speakers. There are important differences, however: First, in the study by Assman *et al.*, the speakers, experimenters, and listeners were all native speakers of the same dialect of Canadian English. The present study depends primarily upon one native speaker of Uighur, although the perceptual experiment includes data from two listeners who are native speakers of American English but also familiar with Uighur. Second, Assman *et al.* used a recording of ten speakers, reading from a list of keywords, and carefully eliminated any erroneous pronunciations; their tokens were thus far more uniform and free of errors of classification from the beginning than the Uighur tokens studied here. As always, there is a trade-off between naturalness and control; their elicitation method meant that their tokens constitute reading pronunciations, which may differ from vowels occurring in natural conversation.

Assman *et al.* found that the nature of the response task influenced the error rate for vowel identification; listeners who responded by marking a word with the same sound as the one heard on tape were wrong about 15% of the time, but when the same listeners were asked to repeat the word they had heard, they spoke the incorrect vowel only about 5% of the time. There were no significant differences for categorization of isolated vowels vs. those in the context /p_p/, contra Strange *et al.* 1976. They suggest that in some of the earlier studies, categorization differences due to consonantal context were confounded with differences due to different response tasks.

Furthermore, they found that under good conditions with trained listeners, even isolated vowels from mixed speakers are miscategorized only 5.4% of the time, partially refuting the theory that listeners need consonant transitions and some time for speaker normalization to take place, although there is a slight but significant improvement when speakers are blocked, so that listeners can adjust to each speaker in turn. (Some other results reported in Assman *et al.* will be discussed below, in connection with the results of the present study.)

Acoustic Study

Methodology

The informant was an adult male native speaker of the Kashgar dialect of Uighur who is also fluent in English. In order to assure that samples of all the vowels were obtained, the examples in Hahn (1991:34-58) were used as the basis of the recording session (although Hahn's phonological analysis is not being adopted here). Every effort was made to obtain tokens which would be as near as possible to those of natural speech, given the constraints of making a recording of high acoustic quality. The order of presentation was randomized, so that the original groupings would not influence the informant. If the subject were simply to be asked to read the examples, that might prejudice the results in favor of the vowels represented orthographically; if prompted in Uighur, that might also influence the outcome. Therefore, the informant was asked in English to translate each of the glosses given by Hahn. Sometimes this resulted in the expected Uighur word; sometimes a second prompt was needed in Uighur suggesting another possible way of expressing the idea, and sometimes the informant did not recognize the word reported in Hahn as a possible translation of the gloss. Often the informant repeated the Uighur word or phrase several times, giving several tokens of the same vowel(s), or offered several ways of translating the gloss. The entire session was recorded in an anechoic room using a high quality analog tape recorder; the Uighur utterances were then transcribed, and the accuracy of the transcription was checked afterward by the informant.

Eighty-five sections of the recording were digitized at 10,000 Hz sampling rate on a Macintosh computer using an 16-bit DSP board, then each waveform was carefully examined, and the endpoint and the duration of the central portion of each vowel was recorded; so far as possible, the portion marked was steady state in terms of both the shape of the waveform and the sound of the vowel, even if this meant marking a relatively short segment. A total of 352 usable tokens of monophthongs were marked in this way.

The digitized samples were then processed on a Unix workstation, using the Waves-ESPS formant extraction program to find the first four formants by autocorrelation LPC. The order of the LPC was 12, and the calculation was done over a 50 ms. window moved in 10 ms. steps. A measure of the probability of voicing was also calculated by the program every 10 ms. Only those portions of the vowels which had a probability of voicing higher than 90% as calculated by the extraction program **and** were within the "hand-measured" endpoints were used. The remainder of the acoustic analysis is based on the duration of each vowel center and the values of the first three formants at 10 ms. intervals within it.

Results

The median duration of the vowels was 46 ms; this may seem quite short, but is reasonable in the light of the strict definitions of what constitutes the steady state portion of the vowel. The distribution was highly skewed; there were two long tokens (around 200 ms.), resulting from hesitation or emphasis on the pronoun [u] 'he'.

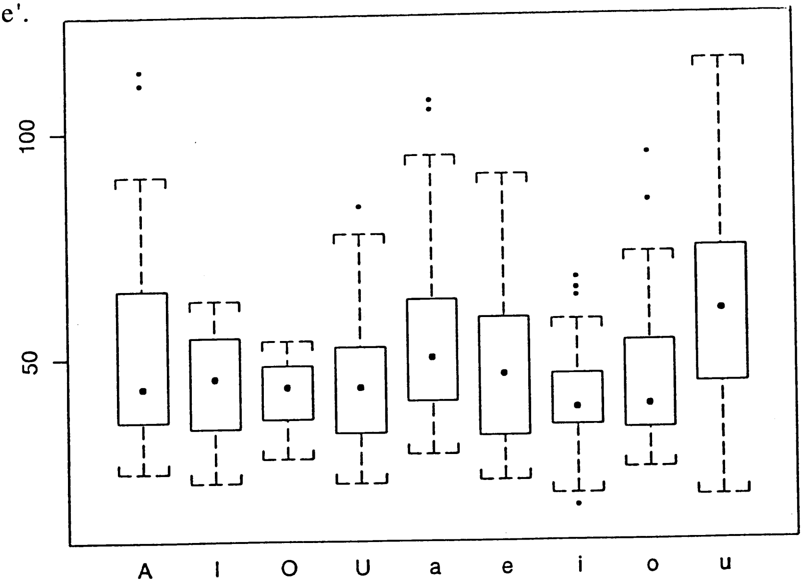

Fig. 3. Durations of vowel tokens by vowel.

Fig. 3 shows the durations grouped by vowel in a Box chart. (The ends of the boxes mark the first and third quartiles, and the dots in the centers mark the medians.) As the chart suggests, the difference in the durations are partially predictable from the type of vowel (an ANOVA gave $p < 0.01$ for the model duration = vowel). but the converse is not true; a vowel with a duration of 55 ms., for example, could easily be [a], [o], or [u]. This suggests that duration may be one of the factors in vowel discrimination, but will not be decisive by itself.

Indeed, the data is very rich and therefore difficult to display adequately in two dimensions. One useful way is to plot all the values for F1 and F2 which have been calculated. This shows us the trajectory of the formant values in F1-F2 space at 10 ms. intervals in the vowel. Figure 4. is an example of such a trajectory, for two repetitions of the phrase yäp maŋmak 'keep on eating', with the axes inverted so that front vowels are to the left and high vowels toward the top of the chart, as in traditional vowel charts based on articulation. The long connected lines show the initial glides starting very high and front and ending fairly low, between mid and front, while the two repetitions of maŋmak show all four vowels in the same general area, quite low and somewhat back. Plots of the trajectory based on the logarithms of formant ratios, as suggested by Miller (1989), show a similar pattern.

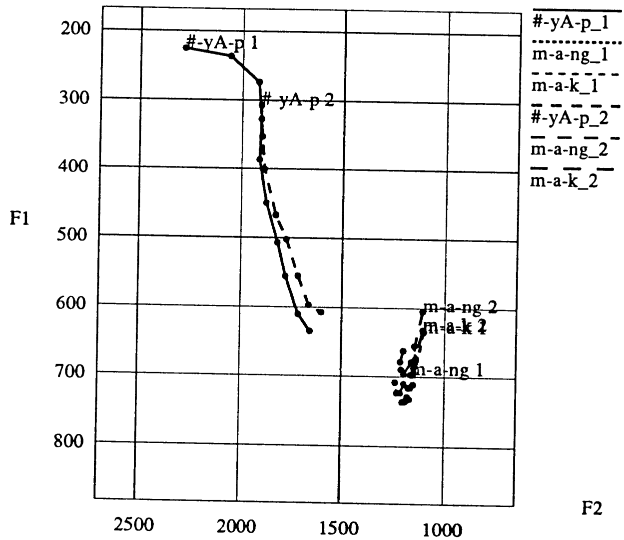

Fig. 4. Values of F1 and F2 for the phrase yäp maŋmak

These kinds of plot provide a level of detail which is lost when we average formant values either over an entire vowel token or over a group tokens of a vowel type. But we do need to make generalizations; one way of generalizing is to choose good tokens of each vowel type and treat every point in such tokens as a valid example of a set of parameter values for that vowel. We can then simply draw a convex hull around these points to define a region which can be identified with the vowel. If we do this for all the vowels, we get a chart like Figure 5 for all the vowels of Uighur in F1-F2 space. This can then be overlaid on other trajectories, giving an idea of how long the trajectory is in the space of a given vowel and how close it comes to the center. The "tails" of the trajectory show the influence of the consonantal context, which will be discussed in more detail below.

A more traditional method of generalizing about the formants of vowels is to average the readings for many tokens, to get a mean value for each vowel type; Fig. 6 shows a graph of such values, along with F1-F2 values for American English vowels derived from Peterson and Barney (1952). Note that, although there are about the same number of vowels in the two languages, the Uighur vowels are much closer together in F1-F2 space. This is due in part to the presence of front rounded vowels in Uighur, whose lowered F2s (and to a lesser extent, F1s) cause them to coincide with some central vowels in that plane. Another factor may be errors in the classification of vowels in the present study, which causes them to appear as outliers, far from the center of vowel regions.

Fig 5. Convex hulls for all vowels of Uighur

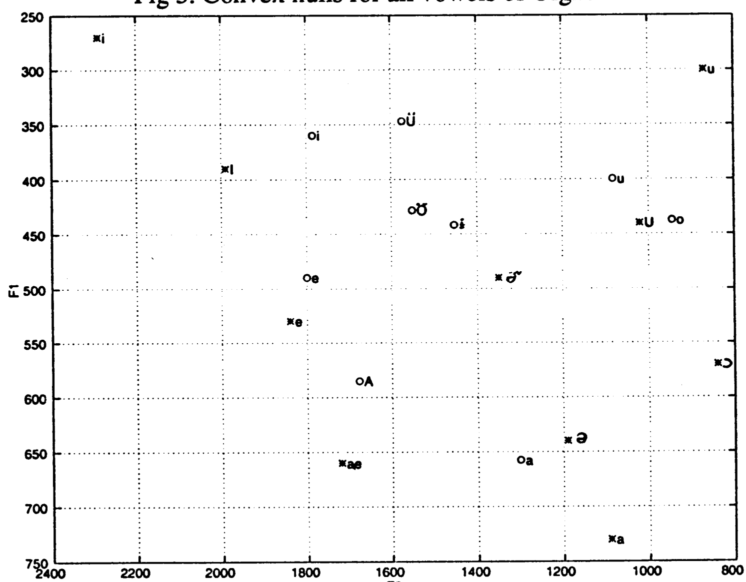

Fig. 6. Mean F1-F2 values for Uighur (o) and American English (*) vowels.

Insofar as we can separate vowels on the basis of F1 and F2, we would expect that at least the vowels in the corners of the triangle, [i], [a], and [u] should be well separated. Fig. 7 (a) shows the mean F1 and F2 for each token of these three vowels; as expected, the regions are fairly separate, although there is still a little overlap. In this regard, neither the ratios of the first three formants nor the logarithms of these ratios (shown in Fig. 7 (b)) seem to do any better than (or even as well as) the formant frequencies at defining separate regions for each vowel.

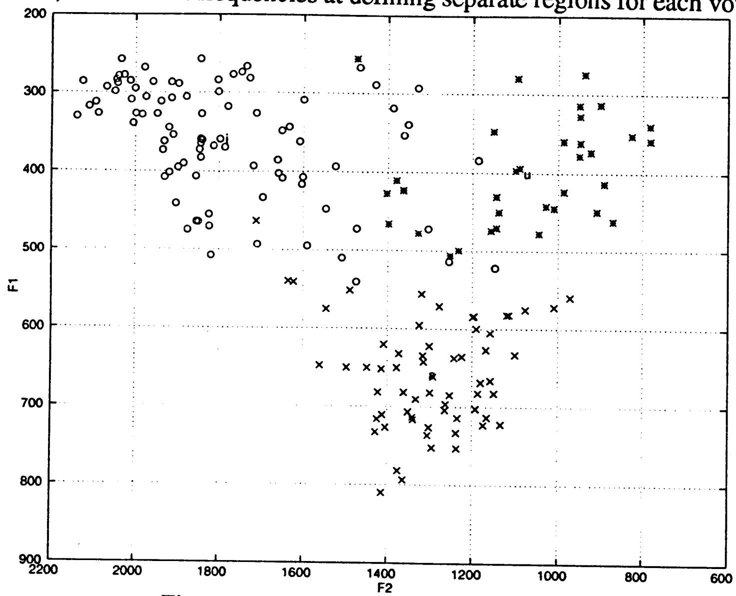

Fig.7(a). Uighur [i], [a], and [u]. F1 x F2

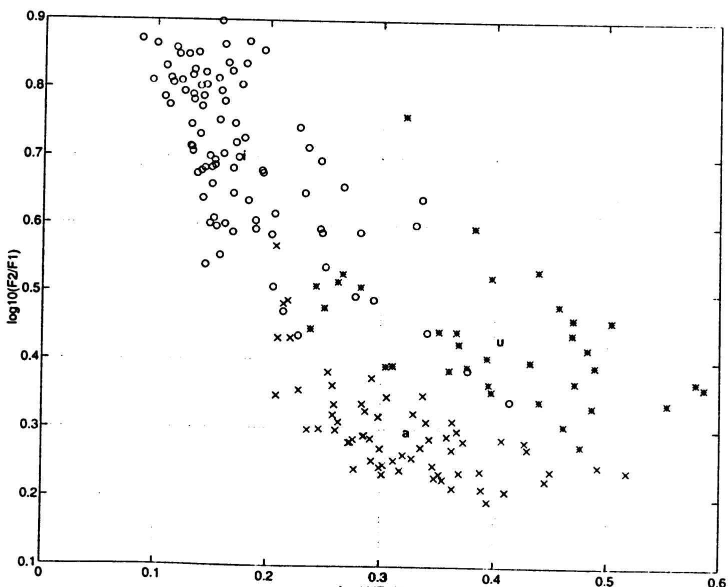

Fig.7 (b). Uighur [i], [a], and [u]. log(F2/F1) x log(F3/F2)

Perceptual study

Methodology

Given that none of the methods discussed above for graphing the vowels produce a complete separation of the vowel regions in the F1-F2 plane, it seems fair to ask ourselves how well native speakers can distinguish tokens of the vowels. If they can do so reliably on the basis of isolated tokens, this suggests that the acoustic clues are present in the speech signal, even if not reducible to simply F1 and F2. If native speakers have difficulty distinguishing vowels without a sentential context, this would suggest that higher level processing, involving semantics and lexical recall is necessary, and we would not expect to find everything in the acoustics of the signal.

To answer these questions, a second experiment was carried out, in which our informant heard digitized Uighur sounds and was asked to classify each of them into one of nine vowel types. For comparison purposes, two non-native speakers of Uighur were also asked to do the same classification task. In this experiment, two sets of stimuli were used, short stimuli consisting of the centers of the vowels which had been marked already, and a new set of longer stimuli, each comprising a full syllable, extracted from the same recordings. The latter were used to test the influence of the consonantal context; some were (monosyllabic) words which could be pronounced in isolation, while others were merely parts of words. The short stimuli were tapered over the first and last 10 ms. intervals to reduce distortion and "popping". The stimuli were presented randomly within each set, and the subjects responded by clicking with a mouse on one of nine buttons, corresponding to the eight graphemes of the Uighur alphabet plus a central vowel written with ɨ (barred i). Note that this meant asking the informant to make all the usual distinctions and one more, deciding which vowels were reduced to something like schwa, although he had already had some practice in making this distinction.

Results

Figure 8 shows the results of the perception experiment for the native speaker informant (abbreviated NS); the left side gives the number of times each vowel type appeared and the number of times it was classified as each of the types. Perfect agreement between the classification on the basis of listening to the central part of the vowel and the spelling given earlier would be represented by having all of the scores on the diagonal. The only case where this occurred was in the classification of the short stimuli of [ä] by the native speaker (16 of the 16 presentations). Overall, the NS classified the short stimuli correctly 60% of the time and the long stimuli 61% of the time. The relatively low success rate on the short stimuli is understandable, given the very short segments that were presented in the experiment, but the equally poor results on the longer stimuli are surprising. Looking at Figure 8 in a little more detail, we notice that the poor results on the long stimuli are due mainly to problems with [ɨ] and [e]. Some difficulty with [ɨ] is more or less predictable, since it is not represented in the orthography. With regard to [e], a check of the response data showed that many of the errors were in the misclassification of [e] as [ä] in presentations of the loan word [rentigen], 'X-ray' (from the name "Röntgen"). If we omit all the tokens of [e] from the calculation, we get 58% success for the short stimuli and 89% for the long ones. (While there were many misclassifications of [e] as [æ], there were very few in the opposite

Frequencies

Short stimuli

Response→ / Stimulus ↓	ä	ɨ	ö	ü	a	e	i	o	u	Tot
ä	**16**									16
ɨ		**14**				2			11	27
ö		1	**19**	2						22
ü	1	10	10	**33**	4	1		2		61
a					**35**				1	36
e	21		2		1	**37**	3	1		65
i	6	13	4	12	1	6	**5**			47
o		7			1			**25**		33
u								8	**11**	19

Long stimuli

Response→ / Stimulus ↓	ä	ɨ	ö	ü	a	e	i	o	u	Tot
ä	**25**				1					26
ɨ		**9**	1	20	11					41
ö			**37**							37
ü	1	21		**62**						84
a					**26**					26
e	100	4			4	**55**				163
o								**35**		35

Percentages

Short stimuli

Response→ / Stimulus ↓	ä	ɨ	ö	ü	a	e	i	o	u
ä	**100**								
ɨ		**52**				7			41
ö		5	**86**	9					
ü	2	16	16	**54**	7	2		3	
a					**97**				3
e	32		3		2	**57**	5	2	
i	13	28	9	26	2	13	**11**		
o		21			3			**76**	
u								42	**58**

Long stimuli

Response→ / Stimulus ↓	ä	ɨ	ö	ü	a	e	i	o	u
ä	**96**				4				
ɨ		**22**	2	49	27				
ö			**100**						
ü	1	25		**74**					
a					**100**				
e	61	2			2	**34**			
o								**100**	

Fig 8. Vowel Perception by Native Speaker

Frequencies

Short stimuli

Response→	ä	ɨ	ö	ü	a	e	i	o	u	Tot
Stimulus ↓										
ä	**21**	2			1					24
ɨ		**24**	4						6	34
ö		1	**19**	1					1	22
ü		34	45	**26**					6	111
a		1			**35**			3		39
e	20	8				**50**	8			86
i	6	27	2	5		17	**16**		5	78
o			1	1				**23**	9	34
u			4	1				2	**17**	24

Long stimuli

Response→	ä	ɨ	ö	ü	a	e	i	o	u	Tot
ä	**19**									19
ɨ		**48**		3						51
ö		1	**28**	1						30
ü		21	12	**41**					1	75
a					**18**					18
e	2	2			2	**126**	1			133
o		2						**32**		34

Percentages

Short stimuli

Stimulus	ä	ɨ	ö	ü	a	e	i	o	u
ä	**88**	8			4				
ɨ		**71**	12						18
ö		5	**86**	5					5
ü		31	41	**23**					5
a		3			**90**			8	
e	23	9				**58**	9		
i	8	35	3	6		22	**21**		6
o			3	3				**68**	26
u			17	4				8	**71**

Long stimuli

Stimulus	ä	ɨ	ö	ü	a	e	i	o	u
ä	**100**								
ɨ		**94**		6					
ö		3	**93**	3					
ü		28	16	**55**					1
a					**100**				
e	2	2			2	**95**	1		
o		6						**94**	

Fig. 9. Vowel Perception by Non-native Speakers

direction. This reinforces the idea that [e] may be an variant of [æ] used in informal speech.)

Figure 9 shows the results for the non-native speakers (NNS), which were 51% correct for the short stimuli, and 89% correct for the long. The NNSs were both able to discriminate [ä]/[e] as well as any of the other pairs, which is not surprising, since both were native speakers of English. The NS fared better than the NNSs on distinguishing the rounded front vowels, [ö] and [ü], but actually did worse than the NNSs on [ɨ] and [i], which may be due to the lack of a symbol for [ɨ] in the Uighur alphabet.

Conclusions and discussion

With regard to one of the initial questions, whether [æ] and [e] are two separate vowels in Uighur, the results are not completely conclusive. The areas of the tokens overlap somewhat in the F1-F2 plane, but no more than many other adjacent pairs of vowels. There is some indication the NNSs are resolving these sounds more precisely than the NS, perhaps because the NNSs are both English speakers, and the difference is definitely phonemic in English (cf. Blankenship 1991). But our informant is also able to correctly categorize the difference more than half the time, and there are two symbols in Uighur orthography, so the hypothesis that there are two low front vowels is likely to be true.

Assman *et al.* found that error rates were much higher for tokens which had been "gated" to remove the initial and final transitions, leaving a 100 ms. central portion of the vowel. Error rates for these gated vowels were much higher in the mixed speaker condition (13.75%) than in the blocked speaker condition (9.50%). They conclude that the information required for speaker normalization is still present in the "steady state" central portions of the vowels, but that the higher error rates overall are due to the removal of clues from formant transitions and duration. The error rates for NS vowel recognition of short stimuli in the present study are higher than those reported in Assman *et al.*, probably because the stimuli were approximately half as long as theirs.

Finally, Assman *et al.* created a series of parameters derived from the acoustic data to be used as predictors of categorization; linear discriminant analysis was used to measure the relative success of various combinations of these parameters. They were able to achieve categorizations that were 80 to 90% correct by means of some of these parameters, although the correlations for individual tokens were less impressive. Unfortunately, the use of linear discriminant analysis requires a normal distribution for each of the underlying acoustic variables; this criterion was not true of the Uighur data in the present study, so different statistical methods will need to be used to model the categorization from the acoustic parameters.

Like Assman *et al.*, we find very large differences across vowels with regard to the reliability with which they can be distinguished. This is contrary to what is predicted by theories such as Lindblom (1986), which suggest that over time, vowels will tend to spread out in acoustic space so that they become maximally differentiable, and thus a more efficient system for encoding information. Natural language seems rather inefficient in this regard, and, furthermore, neither English nor Uighur orthography has been kind enough to provide consistent spellings for the vowels, particularly [ə] and [ɨ], (although in general Uighur orthography does much better than English). We have also found acoustic evidence of some of the phonological processes in Uighur described in Hsu (1992) and Lindblad (1990), such as fronting, reduction to schwa, etc. These may account for some of the discrepancies between the orthography and the acoustic parameters, reflecting the phonological conservatism common to many writing systems, although it is not possible to go into detail here.

With regard to methodology, the technique of plotting tracks of individual tokens of vowels through time reveals much fine detail which is lost when researchers average tokens together in order to describe "typical" or "standard" values for vowels. On the other hand, as mentioned above, we have also found strong evidence that descriptions of vowels in terms of points or areas of the F1-F2 plane are inadequate, especially where rounding is phonemic. The question then becomes how we can represent this kind of temporal detail in a multidimensional space, both graphically as an aid to understanding, and mathematically, as we seek to understand the relation between the acoustics and the perception of vowels.

These results suggest several directions for future research. One would be to measure the effects of surrounding consonants upon the formants of the vowels, based on the methods used in Lindblom (1963). In his work, very consistent F1 and F2 target values were found for vowels provided that the observed values were adjusted by an amount related to the surrounding consonants. Such an adjustment might produce a better clustering of Uighur vowels, also. Whether or not Lindblom's method produces a better separation of the vowels, it seems likely that a more precise model of the Uighur perceptual results in terms of the acoustic data can be found. Although linear discriminant analysis may not be appropriate (due to the non-normal distribution as noted above), there may be tree-based procedures that will yield comparable models. Finally, it should also be possible to measure the perceptual similarity of the vowel tokens directly, using a different technique. Pairs of tokens could be played for NSs and NNSs, asking just for a response of "same" or "different". This kind of response should be relatively free of any orthographic influence, and would also help point out any tokens that are inconsistently classified.

(I am indebted to John Ohala for numerous suggestions with regard to both data collection and interpretation in this study; any errors that remain are my own.)

Bibliography

Assman, Peter F., Terrance M. Neary, and John T. Hogan. 1982. Vowel Identification: Orthographic, perceptual and acoustic aspects. Journal of the Acoustical Society of America. 71. 4. 975—989.

Blankenship, Barbara. 1991. Vowel Perception in a Second Language. UCLA Working Papers in Phonetics. 80 (Dec.). 38—64.

Hahn, Reinhard F. 1991. Spoken Uighur. Seattle: University of Washington Press.

Hsu, Wen. 1992. Uighur Phonology of Compounding. Unpublished ms. Linguistics Dept., University of California at Berkeley.

Joos, Martin. 1948. Acoustic Phonetics. Language. 24 (Suppl). 1-136.

Lindblad, Vern M. 1990. Neutralization in Uyghur. Unpublished M. A. Thesis, Linguistics Program, University of Washington.

Lindblom, Björn. 1963. Spectrographic Study of Vowel Reduction. Journal of the Acoustical Society of America. 35. 11. 1773—81.

Lindblom, Björn. 1986. Phonetic Universals in Vowel Systems. In John J. Ohala and Jeri J. Jaeger (eds.) Experimental Phonology. Orlando, Fla.: Academic Press.

Miller, James D. 1989. Auditory-perceptual interpretation of the vowel. Journal of the Acoustical Society of America. 85. 5. 2114—33.

Peterson, G., and H. Barney. 1952. Control methods used in a study of vowels. Journal of the Acoustical Society of America. 24. 175—84.

Rakerd, Brad and Robert R. Verbrugge. 1985. Linguistic and acoustic correlates of the perceptual structure found in individual differences scaling study of vowels. Journal of the Acoustical Society of America. 77. 1. 296—301.

Stevens, K. N. and A. S. House. 1963. Perturbations of vowel articulations by consonantal context: An acoustical study. Journal of Speech and Hearing Research. 6. 111-128.

Strange, Winifred, Robert Verbrugge, D. Shankweiler, and T. Edman. 1976. Consonant context specifies vowel identity. Journal of the Acoustical Society of America. 60.213-224.

Strange, Winifred, J. J. Jenkins, and T. L. Johnson. 1983. Dynamic Specification of Coarticulated Vowels. Journal of the Acoustical Society of America. 74. 695-705.

Strange, Winifred. 1989. Evolving Theories of Vowel Perception. Journal of the Acoustical Society of America. 85. 5. 2081—87.