a corresponding input [+round]". Align(X,L,Y,L) stands for "Align the left edge of every X to the left edge of some Y".

5. I assume that the same constraint(s) that will account for VhV cases (recall from footnote 2 that the two Vs are identical) will also account for the identity of VV in a bimoraic syllable.

6. The [i...a] and [u...a] forms are addressed in Section 5.1.

7. Golston (1996) argues that all properties of representations should be encoded as exceptions to constraints. I have used more familiar input representations here, for example an input feature [+high] as opposed to a lexically necessary violation of a constraint prohibiting the feature [+high]. The cover constraint VIsE is one of the types of constraints necessary in a full-fledged DOT analysis, where the presence of a feature in a surface form is indicated by a lexically necessary violation of a constraint on feature quality.

8. "Gradient" evaluation of constraint C means that n+1 violations of C are worse than n violations of C.

9. Golston (1996) uses the symbols R and A for necessary violations of a Root constraint and an Affix constraint respectively. (R) and (A) are used for inverse violations of Root and Affix constraints. The Tiv data do not require this distinction, hence the different formalism.

10. As pointed out to me by Heidi Harley, forms without [+round] (like yina 'be less than') also show the interaction of Faithlo * VisE.

11. These forms all have a [+round] feature too; I have no explanation for the absence of [e...i] and [a...i] forms. Recall that "gh" is a digraph for a single segment, as is "ts" and "ng".

Rapid perceptibility as a factor underlying universals of vowel inventories'

Natasha Warner
University of Arizona and Max Planck Institute for Psycholinguistics

1. Introduction

Although the vowel inventories of the languages of the world differ widely, some features distinguish vowel phonemes in many languages, while others are used by relatively few languages. For example, a very large proportion of the languages of the world distinguish at least three heights of vowels (using the features [high] and [low]), and also distinguish front from back vowels (Liljencrantz and Lindblom 1972, Ladefoged and Maddieson 1996), as in the universally most common vowel system, /i, e, a, o, u/. However, much smaller numbers of languages make distinctive use of features such as nasalization, non-modal phonation types (e.g., creaky or breathy voice), advanced tongue root, pharyngealization, or length. This study tests a possible explanation for why some contrasts are more common than others. In so doing, it addresses the relationship between formal (in this case featural) and functional (in this case perceptual) aspects of vowel inventories.

Stevens has argued many times (Stevens 1971, 1980, 1985, Stevens and Blumstein 1981) for the importance of brief regions around points of abrupt acoustic change for perception of speech (cf. also Furui (1986) and Liu (1996)). Stevens and Keyser (1989) and Stevens (1980) suggest that the features which are most often used distinctively in the world’s languages are the ones having cues located at these points of abrupt acoustic change. They identify the features [continuant], [sonorant], and [coronal] as the primary distinctive features based partly on this acoustic criterion.

Lang and Ohala (1996) take this idea further, hypothesizing that for vowels as well as consonants, the cross-linguistically most common contrasts are those which listeners can distinguish based on a very brief portion of the signal (such as 40 ms or less). For example, all or nearly all languages distinguish stops from nasals, thus using the feature [sonorant], which Stevens argues depends on an abrupt change in amplitude. Cues such as the stop burst, or lack thereof, should be rapidly perceptible. However, relatively few languages make distinctive use of secondary articulations, such as palatalization, pharyngealization, or uvularization. These are expected to require a longer portion of the signal to distinguish.
Lang and Ohala (1996) tested this hypothesis for the vowels of North American English by collecting confusion data for gated vowels (vowels of which only the initial portion is presented). They performed hierarchical clustering analyses to determine what contrasts were salient when listeners were allowed to hear various portions of the vowel. (Both the experimental method and the statistical analysis will be explained in detail below, as the same methods are used in the current study.) They found that the vowels were grouped perceptually into clusters around the five universally most common vowels /i, e, a, o, u/ when listeners heard only the first 50 ms of the vowels. Tekieli and Cullinan (1979) also found that vowel height and frontness/backness could be perceived relatively well early in the vowel (within the first 10 ms), but that the tense/lax distinction required a longer portion of the signal to perceive. However, in both of these studies, the main cross-linguistically less common distinctive feature is tenseness. It is important to investigate additional distinctions.

In the current study, I use the Dutch vowel system to test the hypothesis that universally common distinctions can be perceived rapidly. Dutch is particularly appropriate because its vowel system includes several less common distinctions beyond those which have been tested before (Lang and Ohala 1996). Standard Dutch includes several front rounded vowels, and can be analyzed as having a four-way height distinction. It also has a length or tenseness distinction similar to that in English (that is, with considerable difference in vowel quality as well as a durational difference). This paper presents a confusion study for gated versions of all the Dutch vowels. Hierarchical clustering analysis and multidimensional scaling analysis are used to test the hypothesis.

Thus, this paper uses experimental methods (a perception study) and statistical analyses to test a proposed functional explanation for a formal construct. Distinctive features, a formal phenomenon, are used differently by various languages. The hypothesis that the most commonly used ones are those which offer listeners perceptual cues early in the signal is the functional explanation to be tested.

2. Methods

2.1 Materials

One male native speaker of Standard Dutch produced several tokens each of the 15 non-reduced vowels and diphthongs of Standard Dutch in an /hVt/ environment, which has relatively little influence on the quality of the beginning of the vowel. Recording was done on a DAT tape in a sound-treated recording booth. The Dutch vowel inventory appears in Table 1. The analysis given here is that of Boooj (1995), but with the vowels which only occur in recent loanwords omitted. The transcription is that of Gussenhoven (1992). The high vowels are transcribed without length marks because they are of approximately the same duration as the short vowels in most environments, although they pattern phonologically with the long vowels. There is also a reduced vowel, [ə], which occurs only without stress.

There is some question as to which short vowels are counterparts of which long vowels, but on grounds of phonetic locations in the vowel space, duration, phonological alternations, and historical origins, the long high vowels are generally assumed to be without short counterparts. /i, e/ are considered upper-mid, and /e/ lower-mid. Since /a, ã/ are low vowels, this means that Dutch may be analyzed as having four distinctive vowel heights, although it does not have four vowels distinguished solely by height. The transcription of vowels which form tense/lax or long/short pairs reflects the fact that both duration and quality differ in these vowels, as in English.

<table>
<thead>
<tr>
<th>Table 1. The Dutch vowel system.</th>
</tr>
</thead>
<tbody>
<tr>
<td>front unrounded</td>
</tr>
<tr>
<td>high</td>
</tr>
<tr>
<td>upper mid</td>
</tr>
<tr>
<td>lower mid</td>
</tr>
<tr>
<td>low</td>
</tr>
<tr>
<td>diphthongs</td>
</tr>
</tbody>
</table>

The vowel system of the speaker for the experiment, as measured from the tokens chosen for use as stimuli, is shown in Figure 1. The first, second, and third formants of each item were measured at three points in time, one-quarter, half, and three-quarters through the duration of the vowel. (Only the first and second formants are shown in the figure.) Several important points are evident from this figure. First, the long mid vowels are phonetically quite diphthongal (as is shown by the length of the lines connecting the measurement time points), although they pattern phonologically with monophthongs.

Second, this figure demonstrates one of the phonetic reasons for considering /i, ë/ to be upper-mid, and not high, vowels: they have approximately the same height in the vowel space as the beginning points of the long mid vowels. Third, the front rounded vowels are relatively far back in the vowel space, appearing more central than front. Finally, most of the change in the diphthongs /ei, ëy, ëu/ is late in the diphthong: they have very little change between the first two measurement points, but a rapid change between the measurement points at half and three-quarters of the duration of the diphthong. These observations have been noted in the literature on Dutch vowels before (Pols, Tromp and Plomp 1973, Pols 1977,
training stimuli (described above), followed by the 135 test stimuli. Each subject heard the test stimuli in a different random order (not blocked by gate). Each stimulus was presented only once to each subject. Subjects identified the vowels by using a mouse to click on one of 15 buttons on a computer screen. The buttons were labeled with the orthographic representation of the Dutch vowels, and also had a real word illustrating the vowel next to them on the screen. The data from 9 subjects was excluded because they either identified less than 90% of all training stimuli correctly, or they identified all three training stimulus gates of a particular vowel incorrectly. Since the training stimuli allowed listeners to hear (nearly) the entire vowel, or even the entire CVC syllable, consistent failure to identify training stimuli correctly suggests a difference in the vowel system attributable to dialect. A confusion matrix (frequency of each type of response to each type of stimulus) was tabulated from the responses of the remaining 49 listeners, and was used as the input for the analyses described below.

3. Results

3.1 Hierarchical clustering analysis

A hierarchical clustering analysis groups stimuli into clusters based on their similarity. The results are presented in terms of the abstract distance at which two stimuli or clusters of stimuli combine to form a (larger) cluster. The analysis begins with each vowel as a separate unit (no clusters), and then groups the most similar vowels together into a cluster, and then the next most similar together, and so on, until all of the vowels form one large cluster. What is of interest is which vowels group together to form sub-clusters. The order in which various items combine into clusters is shown in a dendogram.

For a given gate point, if the vowels which are distinguished by some phonological feature combine into a cluster at a small distance, that indicates that those vowels are rather similar to each other, and thus that the phonological feature distinguishing them is not very perceptible at that gate. The hypothesis that less common distinctive features require a relatively long portion of the signal to be perceived predicts that vowels which are distinguished by an uncommon feature will form a cluster at early gates, because at those gates they will be perceived as similar. Particularly, there should be clusters around the areas of the most commonly distinguished vowels, perhaps around the Dutch vowels /i/, /e/, /a/, /o/, /u/. The long and short vowel pairs should cluster together, and the front rounded vowels should be included in nearby clusters (for example, with /y/ in the same cluster as either /i/ or /u/).
3.1.1 The four large clusters
A hierarchical clustering analysis was carried out for each gate point. The results of this analysis for the gate ending 40 ms after the onset of voicing are shown in Figure 2. At this gate, the vowels fall into four large clusters, consisting of the low vowels (plus the back diphthong) /æ, a, au/, the back non-low vowels /o, ð, ð/, the front rounded vowels /i, y, ø/, and the front unrounded vowels (plus the front rounded diphthong) /i, t, e, e, øy/.

The grouping into the four large clusters low, back non-low, front rounded, and front unrounded is typical of the hierarchical clustering results for most gates, although it is less clear for the very late gates, where there are few confusions. The very first gate, which allowed listeners to hear only the /h/ of the /hVt/ syllable, also shows a somewhat different pattern, as shown in Figure 3. However, even at this gate, the low, non-low back, and front rounded clusters appear. (Although the fricative /h/ can contain much information about an upcoming vowel, gating the stimuli to noise made it difficult to use the information in the frication noise.)

![Dendogram showing combination of clusters for the 0 ms gate. The horizontal dimension shows the normalized distance at which each pair of stimuli or clusters combine.](image)

The fact that the front rounded diphthong /øy/ falls into the front unrounded cluster rather than the front rounded one may seem anomalous. This is true of all gates except the 0 ms, 20 ms, and 160 ms gates. (At the 20 ms gate, /øy/ clusters with the low vowels. At the 0 ms and 160 ms gates, it does form a cluster with the front rounded vowels, but at great distance, and with some front unrounded vowels included in the same cluster.) The grouping of /øy/ with the front unrounded vowels reflects its frequent misperception as /l/. Although /øy/ is rounded, rounding has less acoustic effect on low vowels, and its beginning point is rather low. Furthermore, there is no low or lower-mid front rounded monophthong in the Dutch system for /øy/ to be misperceived as.

3.1.2 Results for vowel height
At the 40 ms gate (Figure 2), within each of the large non-low clusters, there are smaller clusters separating the mid vowels from the high vowels. Thus, the non-low back cluster contains smaller clusters, one consisting of /o, ø/ and the other only of /u/. The front rounded cluster has sub-clusters of /y/ and /r, ø/, and the front unrounded cluster contains one smaller cluster of /i, t/ and another of /e, e, ø/ (along with /øy/). This arrangement of the mid and high vowels within the larger clusters is apparent at each gate from 20 ms to 80 ms (after which the long high vowels are too rarely misperceived to form such clusters). These smaller clusters of mid and high vowels within the larger clusters mean that each high vowel is more similar to its corresponding mid vowel than it is to the other high vowels in the system. Each mid vowel is also more similar to its corresponding high vowel than it is to other mid vowels. That is, the front/back and rounded/unrounded distinctions are clearer from the first 20–80 ms of the vowel than the high/mid distinction is. This is not the grouping predicted by the hypothesis: the front rounded vowels
form a cluster by themselves, and have little perceptual similarity with either the front unrounded or the back vowels.

The 60 ms gate also has smaller high and mid clusters within the larger non-low clusters, but a slightly different pattern appears within the front unrounded cluster, as shown in Figure 4. There are three subclusters within the front unrounded cluster, one consisting only of the vowel /i/; one containing /e, ɪ/, and another containing /e, ɛ, ɛɪ/. Thus, instead of clusters of high and mid vowels, there are clusters of high, upper-mid, and lower-mid vowels. The 80 ms gate shows the same pattern, and at later gates, /e, ɪ/ and /ɛ, ɛɪ/ continue to form upper-mid and lower-mid clusters, although /i/ is rarely misperceived. The change from just high and mid clusters at the 40 ms gate to high, upper-mid, and lower-mid clusters at later gates shows that listeners are only able to distinguish high from mid (and low) vowels based on the information in the first 40 ms of the vowel, but they are able to distinguish upper-mid from lower-mid vowels based on the information in the first 60 ms. Thus, as predicted by the hypothesis, listeners require more of the signal to distinguish four vowel heights than three.

3.1.3 Results for vowel length
In the dendogram representation of a hierarchical clustering analysis, the distances at which stimuli combine are normalized, so one cannot compare them across different gates. One can only compare the composition of the clusters, not the distances at which they form, across gates. However, the non-normalized distance coefficients, which represent the average dissimilarity of all pairs of stimuli in the clusters being combined, can be compared across gates to determine whether a particular set of vowels becomes more or less similar over time. This provides a useful way to examine the role of the long/short vowel distinction. The long/short vowel pairs form clusters at nearly all gates (although /ɪ/ varies as to whether it forms a pair with /ɪ/ or /ɛɪ/), so the composition of the clusters alone does not tell us how listeners’ use of this distinction changes over time. The non-normalized distance coefficients for the formation of long/short vowel clusters at the 40 ms, 100 ms, and 140 ms gates appear in Table 2. These coefficients show that for all four long/short vowel pairs, the distance at which they join the cluster increases from earlier to later gates. Thus, although the vowels in long/short pairs are very similar to each other, the perceptual distance between them does increase as listeners hear more of the vowel. The distinction between the long and short vowels becomes clearer as more of the vowel is heard.

Table 2. Non-normalized distance coefficients for clustering of long and short vowels at several gates

<table>
<thead>
<tr>
<th>Vowel pair or cluster</th>
<th>40 ms gate</th>
<th>100 ms gate</th>
<th>140 ms gate</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ɪ/ - /ɛɪ/</td>
<td>2.63</td>
<td>3.28</td>
<td>6.71</td>
</tr>
<tr>
<td>/ɪ/ - /ɛɪ/</td>
<td>5.27</td>
<td>7.05</td>
<td>8.91</td>
</tr>
<tr>
<td>/æ, /ɛ, /ɪ/</td>
<td>6.07</td>
<td>7.97</td>
<td>9.70 (160 ms gate)³</td>
</tr>
<tr>
<td>/ɛ - /ɪ/</td>
<td>not a direct cluster</td>
<td>3.38</td>
<td>7.76</td>
</tr>
</tbody>
</table>

3.2 Multidimensional scaling analysis
A hierarchical clustering analysis shows which vowels are perceptually most similar, but it does not show explicitly what acoustic characteristics listeners use in distinguishing the vowels. A multidimensional scaling analysis⁴ was performed on the same confusion matrix data to address this question. This analysis incorporates the confusion data from several gates simultaneously, to determine what perceptual dimensions account for listeners' confusions overall. However, the analysis becomes inaccurate if confusion matrices with extremely few confusions are included, so only the first six gates were used. (See van der Kamp and Pols 1971 and Wright 1986 for other examples of the application of this method to the perception of vowels.)

A multidimensional scaling analysis uses similarity data, in this case based on confusions, to place the items (vowels) in a multidimensional space which represents
the perceptual dimensions listeners use in perceiving the vowels. One can then examine the dimensions the analysis produces to determine what acoustic characteristics they correspond to. The analysis produces solutions using various numbers of dimensions. Solutions using more dimensions will always fit the data better, because they allow more ways for variation in each vowel to be accounted for. However, it is unlikely that human listeners actively classify vowels on, say, six or more dimensions. If the solution with three dimensions provides nearly as good a fit to the data as the solutions with four or more dimensions, one can conclude that the solution with three dimensions is at least a sufficient representation of the perceptual space used by human listeners, and one can then turn to determining what acoustic characteristics those three dimensions represent. In this case, the three dimensional solution provides a good fit to the data (Table 3). There are relatively gradual decreases in the average $R^2$ (goodness of fit of the model) between the six dimensional and the three dimensional solutions, but a large decrease between the three dimensional and the two dimensional solutions, indicating that a two-dimensional solution does not model listeners’ perception well.

Table 3. Average $R^2$ for the multidimensional scaling analysis using the first six gates, for solutions with various numbers of dimensions

<table>
<thead>
<tr>
<th>Number of dimensions</th>
<th>Average $R^2$ across gates</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>.89</td>
</tr>
<tr>
<td>5</td>
<td>.86</td>
</tr>
<tr>
<td>4</td>
<td>.79</td>
</tr>
<tr>
<td>3</td>
<td>.74</td>
</tr>
<tr>
<td>2</td>
<td>.56</td>
</tr>
</tbody>
</table>

Figures 5 and 6 show the stimulus locations as calculated by the multidimensional scaling analysis for the three dimensional solution. Figure 5 shows the locations of the vowels in the first and second dimensions, and Figure 6 shows the first and third dimensions. In Figure 5, the vowels seem to be arranged approximately in the shape of the vowel space, with /i/ and /i/ having high values for both dimensions, /u/ having a low value for dimension one but a high value for dimension two, and /æ, a, au/ having somewhat low values for dimension one and very low values for dimension two. Thus, from a visual inspection of the figure, dimension one seems to reflect frontness/backness of vowels, and dimension two to reflect vowel height. However, the front rounded vowels are lower in dimension two than one would expect. The correlation of the values on dimension one with the second formant measured at one quarter through the vowel is $r = .86$ (p < .001). The correlation of the second dimension with the first formant measured at one quarter

through the vowel is $r = -.79$ (p < .001).5 As Figure 6 shows, the third dimension primarily seems to separate the front rounded vowels (except the front rounded diphthong) from all the other vowels. This dimension might reflect rounding, although the low values for /u, o, au/ are problematic. The correlation of values on this dimension with the third formant is $r = -.76$ (p < .002).

Multidimensional scaling analyses were also carried out using the data from the first 4 or 5 gates instead of the first 6, as the number of confusions at the sixth gate might already be too small to be useful. Results were similar to the results with six gates. Since the four-dimensional solution in the analysis with six gates provides a somewhat better fit to the data than the three-dimensional solution, and it is reasonable to suppose that listeners might use four dimensions in perceiving vowels, that solution was also examined, but the dimensions were not readily interpretable (i.e., they did not map onto any known characteristics of vowels). Therefore, the three-dimensional solution to the analysis with data from 6 gates, above, will be adopted.

Van der Kamp and Pols (1971) and Pols et al. (1969) have also applied multidimensional scaling analysis to perceptual data on Dutch vowels. They used either a single period excised from vowels or resynthesized completely steady state vowels, rather than natural gated vowels, as stimuli. Both found that a three dimensional
4.1 Front rounding: The feature for which the hypothesis is not supported

The results regarding the perception of front rounded vowels contradict the hypothesis that cross-linguistically uncommon features require a longer portion of the signal to distinguish than common features do. This is particularly clear from the hierarchical clustering analyses at the early gates, in the fact that the front rounded vowels cluster together instead of forming clusters with their corresponding front unrounded vowels, for example. This leads to the question of why, if front rounding can be distinguished so reliably based even on the first 20 ms of a vowel, it is not used distinctively in more of the world’s languages. In particular, since front rounding is more clear than the difference between high and mid vowels from the early portions of the vowel, why is it that far more languages distinguish high from mid vowels than front rounded from front unrounded and back vowels? There are several possible explanations for this.

First, the perception of front rounding may be more affected by consonantal environment than the high/mid distinction is. In this study, all vowels were recorded in the environment /hVT/ because Lang and Ohala (1996) found that responses at the crucial early gates were strongly influenced by consonantal context if /d/ or /g/ was used as the initial consonant instead. Some earlier studies excised vowels from a variety of contexts, though. Pols (1977) recorded the Dutch vowels in all possible CVC environments, excised the vowels from the syllables, and presented them to listeners in isolation for identification. His results (1977:110) show a large number of confusions between front rounded vowels and both front unrounded and back vowels. In order to compare these results to those of the current study, I performed a hierarchical clustering analysis on the data from Pols’ (1977) study, and found that in his data, the front rounded vowels form clusters with front unrounded vowels instead of with each other. /e/ forms a cluster with the mid front rounded vowels, and /æ/ forms a cluster with /i, e, u/ (Figure 7).

Koopmans-van Beinum (1980) presented 12 Dutch vowels recorded in isolation, excised from isolated words, and excised from connected conversational speech to listeners for identification. The last condition, of course, produced many confusions, with more confusions of the front rounding distinction than the high/mid distinction, except for the back high and mid vowels. Thus, it may be that when consonantal influence on the vowels varies, the high/mid distinction is more perceptible than front roundedness. Consonantal effects on the vowel may be more detrimental to the perception of front rounding than to the perception of height. A second possible explanation for the discrepancy between the ready perceptibility of front rounding in this experiment and its cross-linguistic low frequency lies in the acoustic cues for vowels. Vowels, at least when stressed, are inherently rather long segments. Crystal and House (1982, 1988) find that the average duration of English vowels is longer than the average duration of all types of consonants except
Languages with seven-vowel systems usually either have an upper-mid/lower-mid distinction or have two central or front rounded vowels, and neither of these patterns clearly predominates. Thus, when separation of the vowels within the vowel space does not dictate otherwise, a front rounding distinction may not be cross-linguistically uncommon. However, this does not explain why front rounding is less common than the high/mid distinction, even though it is more readily perceptible even over a very short time window.

4.2 Features for which the hypothesis is supported

Unlike front rounding, both the four-level height distinction and the long/short distinction support the hypothesis that less common features require a longer duration of the signal to perceive. The change in the hierarchical clustering analysis for the front unrounded vowels from the 40 ms gate to the 60 ms gate is particularly convincing: at 40 ms, there is a high vowel cluster /i, i/ and a mid vowel cluster /e, e/ (also including /e/) as well as a low vowel cluster consisting of /a, a, a/. However, at 60 ms, there are high, upper-mid, lower-mid, and low vowel clusters. Thus, as the differences among /i, i, e, e/ become clearer, listeners begin to distinguish four vowel heights instead of three. The universally very common three-level height distinction can be perceived based on cues in the very beginning of the vowel, but perception of the less common four-level height distinction requires a longer portion of the signal.

Although there is some question as to which vowels form long/short pairs in Dutch, the results discussed in Section 3.3.4 above demonstrate that the length distinction is more perceptible at late gates than at early ones. This might seem to be an obvious consequence of truncating segments which are differentiated by duration, but in Dutch, as in English, there are substantial differences in vowel quality as well as in duration between long and short members of a pair. Some researchers consider this a tenseness distinction rather than a length distinction (Gussenhoven 1992, van Oostendorp 1995). Thus, there are potential cues to this distinction even early in the signal, but they are not strong enough to make this distinction readily perceptible at early points in the vowel. The results for long and short vowels in Dutch replicate Lang and Ohala’s (1996) results for English vowels: they also found that long/short (or tense/lax) vowel pairs tended to form clusters, and that these clusters were stronger at earlier gates. Although a length or tenseness distinction is not unusual cross-linguistically, it is certainly less common than the
5. Conclusions

The results support the hypothesis that some distinctions appear in many languages because they are readily perceptible from cues in a brief time window around an abrupt acoustic change. However, some distinctions which are perceptible based on a short time window may not be cross-linguistically common because of other influences on the phonological system, such as separation of vowels in the vowel space or a differential effect of surrounding context. Thus, the choice of distinctive features in languages' inventories is related to rapid perceptibility, but this is not the only factor which contributes to universals of vowel systems.

An interesting direction for future research might be the exploration of a yet wider variety of vocalic distinctive features, using several languages. The current paper extends past work on English, particularly on the tense/lax distinction, to a four-way height distinction and a front rounding distinction, and finds support for the hypothesis for some but not all features. Future work on timing of perception of distinctively nasalized vowels, vowels with creaky or breathy voice, vowels with advanced tongue root, as well as perhaps back unrounded vowels, would elucidate which types of cross-linguistically less common features require a long portion of the signal to perceive, and which, like front rounding, are immediately perceptible.

Work on a wider variety of languages and distinctive features might also allow one to examine speed of perception of very uncommon vs. somewhat uncommon vs. very common distinctions. The current work divides vocalic distinctive features only into those which are used in a typical 5-vowel system and those which are less common than that. However, distinctive use of non-modal voicing for vowels is probably even less common than distinctively nasalized vowels, for example, and work on a wider variety of languages would allow one to determine whether these finer distinctions are also related to perceptual factors.

The current work relates formal properties of the vocalic system, the phonological features which distinguish vowels, to a functional motivation, rapid perceptibility of acoustic cues. The results confirm that at least in the area of vowel inventories, formal and functional aspects of language are closely related.

Notes

* I would like to thank John Ohala, Anne Cutler, Allard Jongman, Carlos Gussenhoven, Terry Nearey, and Roel Smits for discussion of this research and of statistical methods, as well as two anonymous reviewers for their helpful comments. I would also like to thank Niels Janssen for his effort in running the experiment, and John Nogengast for technical assistance. Errors are, of course, my own.

1. The average linkage between groups method was used, with a chi-squared dissimilarity measure (Norušis & SPSS, 1994).

2. The method of hierarchical clustering (Norušis & SPSS, 1994) involves first looking for the two items which are most similar to each other and combining them into a cluster. It then looks for the next most similar items, and either adds one item to the first cluster, or combines two items into a separate cluster. This process continues until all items have been combined into a single large cluster. Thus, the process necessarily involves all items being combined into one cluster comprised of rather dissimilar items, and also involves each item appearing by itself at the initial stage of the analysis. What is of interest is which items combine into clusters of intermediate size. Although all hierarchical clustering analyses will show each vowel starting out as separate, and all vowels being combined into one large cluster, the composition of the clusters in between these stages, and the order in which vowels join those intermediate clusters, can vary.

3. At the 40 ms and 100 ms gates, /a/ combines with /a/ only after it has already formed a cluster with /a/. However, at the 140 ms gate, where there are very few confusions, /a/ forms a cluster more closely with /a/, /e/ than with /a, a/, and this cluster is therefore not representative of the distance at which the long and short vowel combine. Therefore, the value for the cluster /a, a/ at the 160 ms gate is given here.

4. The INDSCAL method, chi-squared dissimilarity measure (Norušis & SPSS, 1994) was used.

5. F1 may not be an exact reflection of what listeners perceive as vowel height, but it is generally accepted that F1 varies inversely with vowel height. Similarly, F2 alone may not be the exact perceptual cue for frontness, but it is clearly related to it. Rounding is more difficult to map directly to an acoustic characteristic, but rounding does tend to lower F3. See Johnson (1997) and Stevens (1998) for general discussion of the relationship between formant frequencies and vowel quality.

6. This effect of consonantal environment might seem to imply that perception of front roundedness would differ across languages depending on consonant inventory. However, the major effect of consonantal environment is probably determined by place of articulation, and since most languages have at least labial, coronal, and dental consonants, there may not be much cross-linguistic difference along these lines in degree of influence on perceptibility of front rounding.