TIMING OF PERCEPTION OF VOCALIC DISTINCTIVE FEATURES: IMPLICATIONS FOR VOWEL SYSTEM UNIVERSALS

Natasha Warner
Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

ABSTRACT

This paper examines factors which motivate the use of certain distinctive features in vowels across languages. Features which are cross-linguistically common may be those which listeners are able to perceive based on a brief portion of the acoustic signal, perhaps around a point of abrupt acoustic change. Past work on English vowels has supported this hypothesis. This paper presents results from a gating study of Dutch vowels on perception of several less common vocalic distinctions. Hierarchical clustering analyses reveal that the long/short distinction and the four-way height distinction require more of the signal to perceive than the high/low or front/back distinctions, supporting the importance of rapid perceptibility. However, front rounded vowels are distinguishable from other vowels even based on a very brief signal, contradicting the prediction.

1. INTRODUCTION

Although the vowel inventories of the languages of the world differ widely, some features are distinctive in many more languages than others. For example, most of the languages of the world distinguish at least three heights of vowels, as well as front from back vowels [1, 2]. The universally most common vowel system, /i, e, a, o, u/, uses these distinctions. However, contrastive nasalization, non-modal phonation types, advanced tongue root, or pharyngealization are found in far fewer languages. This study tests a possible explanation for why some contrasts are more common than others.

Stevens has argued many times [3, 4, 5, 6] for the importance in speech perception of brief regions around points of abrupt acoustic change. Stevens and his colleagues suggest [4, 7, 8] that the distinctive features most often used in the world's languages are the ones having cues located at these points of abrupt acoustic change, particularly those having a sudden onset or cessation of energy at some frequency. They identify the features [continuant], [sonorant], and [coronal] as primary based partly on this acoustic criterion.

Lang and Ohala [9] clarify Stevens' suggestion, hypothesizing that for both vowels and consonants, the cross-linguistically most common contrasts are those which listeners can perceive based on a very brief portion of the signal (such as 40 ms or less). They argue that these rapidly perceptible features will be found in most languages, and languages which use less common features distinctively will also make the more common distinctions. The less common contrasts are hypothesized to take longer for listeners to perceive.

For example, (nearly) all languages distinguish stops from nasals, thus using the feature [sonorant]. Cues such as the stop burst, or lack thereof, should be rapidly perceptible. Even languages with extremely small consonant inventories, such as Hawaiian, make this distinction. However, relatively few languages have distinctive secondary articulations, such as palatalization, pharyngealization, or uvularization. Most languages which do (e.g. Slavic languages or Arabic) have relatively large consonant inventories and also use the more common features. Secondary articulations are likely to require a longer portion of the signal to perceive.

Lang and Ohala [9] tested this hypothesis for the North American English vowels. They gated the vowels out at various points and asked listeners to identify the vowel. They performed hierarchical clustering analyses on the confusion data in order to determine what contrasts were salient when listeners were allowed to hear various portions of the vowel. They found some evidence that the 11 vowels they tested grouped perceptually into clusters around /i, e, a, o, u/ when listeners heard only the first 50 ms of the vowels. However, their study did not use enough subjects to be reliable. Furthermore, the vowel system they tested offers only a few uncommon vocalic features. It is important to investigate a variety of languages with additional distinctions.

In the current study, I test this hypothesis using the Dutch vowel system. Dutch is a particularly appropriate language for this test, because its vowel system includes several front rounded vowels, and can be analyzed as having a four-way height distinction. It also has a length distinction similar to that in English. Thus, Dutch offers a chance to replicate the previous findings and to test them further.

2. METHODS

One male native speaker of Dutch produced several tokens each of the 15 non-reduced vowels and diphthongs of Standard Dutch (Table 1) in an /hVt/ environment. Recording was done on a DAT tape in a soundproof booth. The speaker was from the Utrecht area, and stated that he spoke Standard Dutch.

	front unrounded	front rounded	back
high	i	y	u
upper mid	ɪ, e:	ʏ, ø:	o:
lower mid	ɛ		ɔ
low			ɑ, a:
Diphthongs	ɛi	œy	ʌu

Table 1. The Dutch vowel system. Analysis of [10], transcription of [11]. The unstressed reduced vowel, ə, is not included in the experiment or the table.

One should note from Table 1 that Dutch may be analyzed as having four levels of vowel height. There is some question as to which long and short vowels form pairs, but for phonetic, phonological, and historical reasons, the long high vowels are usually considered not to have short counterparts. The vowels /ɪ, ʏ/ are analyzed as upper mid, as the short counterparts of /e:, ø:/. Since /ɛ/ is lower than /e:/ and /a:, ɑ/ are low vowels, this analysis involves four vowel heights, even though Dutch does not have four vowels distinguished solely by height. Recent borrowings also use several additional vowels which require a four height analysis [10, 11]. A four level height distinction is cross-linguistically far less common than a three level one. Table 1 also shows that Dutch has several front rounded vowels, and a front rounded diphthong. Although many languages of Western Europe have front rounded vowels, distinctive rounding on front vowels is relatively uncommon in the world's languages.

The vowel system of the speaker for the experiment, as measured from the 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. Figure 1 shows that the long mid vowels are phonetically quite diphthongal, although they pattern phonologically with monophthongs [10]. The phonetics of this speaker's vowel system is similar to what is reported in previous literature on Dutch vowels [12, 13, 14, 15].

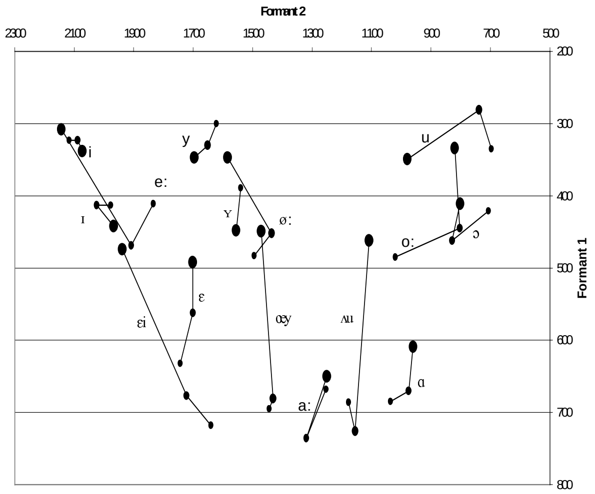

Figure 1. Formants of the speaker's vowels over time. Larger points are later in the vowel.

One production of each /hVt/ item was chosen for use in the experiment. Each item was final-gated at 9 points, located at 20 ms intervals: the shortest gate ended at onset of voicing, the next 20 ms after onset of voicing, etc, with 160 ms of the vowel included in the longest gate. All stimuli were gated to white noise, with a 5 ms transition period. The noise continued such that all stimuli had a total duration of 455 ms. Three training stimuli were also generated from each item, with gating end points at 30 ms before the end of the vowel, at the end of the vowel, and after the /t/ release at the end of the syllable.

58 native speakers of Dutch participated in the experiment. Subjects were tested individually in a soundproof booth, and heard first the 45 training stimuli, followed by the test stimuli. Each subject heard the test stimuli in a different random order (not blocked by gate). Subjects identified the vowels by using a mouse to click on one of 15 buttons, labeled with the orthographic representation of the vowels, on a computer screen.

To reach criterion on the training test, subjects had to identify at least 90% of the training stimuli correctly, and also had to identify at least one gate of each vowel correctly. (Thus, subjects who consistently misperceived the speaker's production of a particular vowel, even when allowed to hear the entire syllable, were excluded.) 9 of the 58 subjects failed to reach criterion, many because they consistently perceived /ø:/ as /œy/. This may be due to dialectal differences. Only the results of the remaining 49 subjects are discussed below.

3. RESULTS

3.1. Confusion matrices

Confusion matrices were compiled from the data, as shown in Tables 2 and 3.

	i	ɪ	e:	ɛ	ɛi	y	ʏ	ø:	œy	a:	ɑ	u	o:	ɔ	ʌu
i	34	15	0	0	0	0	0	0	0	0	0	0	0	0	0
ɪ	2	40	5	0	0	0	0	0	0	0	0	1	1	0	0
e:	0	19	6	22	2	0	0	0	0	0	0	0	0	0	0
ɛ	0	4	2	39	2	0	0	0	1	0	0	0	0	1	0
ɛi	0	0	0	34	12	0	0	0	3	0	0	0	0	0	0
y	1	0	0	0	0	42	5	1	0	0	0	0	0	0	0
ʏ	0	1	0	0	0	2	42	4	0	0	0	0	0	0	0
ø:	0	0	0	0	0	0	40	8	1	0	0	0	0	0	0
œy	0	0	0	13	2	0	1	0	31	0	2	0	0	0	0
a:	0	0	0	0	0	0	0	0	5	24	20	0	0	0	0
ɑ	0	0	0	1	0	0	0	0	2	2	38	0	1	4	1
u	0	1	0	0	0	0	0	0	0	0	0	42	1	5	0
o:	0	1	0	0	0	0	0	0	0	0	3	0	24	21	0
ɔ	0	0	0	0	0	0	0	0	0	0	0	4	3	42	0
ʌu	0	1	0	0	1	1	2	1	25	2	15	0	0	0	1

Table 2. Confusion matrix for the gate ending 20 ms after onset of voicing. Stimuli in left column, responses in top row.

The later confusion matrices (e.g. Table 3) show that the long mid vowels are misperceived until very late gates. Misperceptions of /e:, ø:/ and to some extent /o:/ as their short counterparts are common at late gates, but the other long vowels (/i, y, u, a:/) stop being misperceived earlier. The long mid vowels and the diphthongs are similar in the time course of their perception. The persistent misperception of the long mid vowels probably stems from their diphthongal character. Pols [13] found that when these vowels were excised from CVr syllables, listeners could not identify them correctly, and attributed this to the lack of diphthongization when these vowels are followed by /r/. Gating also removes the diphthongization at the end of the vowel, which appears to be an important cue for the long mid vowels, even though they are not true diphthongs [cf. 16].

	i	ɪ	e:	ɛ	ɛi	y	ʏ	ø:	œy	a:	ɑ	u	o:	ɔ	ʌu
i	49	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ɪ	0	47	2	0	0	0	0	0	0	0	0	0	0	0	0
e:	0	26	20	3	0	0	0	0	0	0	0	0	0	0	0
ɛ	0	0	0	49	0	0	0	0	0	0	0	0	0	0	0
ɛi	0	0	0	26	23	0	0	0	0	0	0	0	0	0	0
y	0	0	0	0	0	48	0	1	0	0	0	0	0	0	0
ʏ	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0
ø:	0	0	0	0	0	0	33	15	1	0	0	0	0	0	0
œy	0	0	0	3	1	0	0	1	42	2	0	0	0	0	0
a:	0	0	0	0	0	0	0	0	1	42	6	0	0	0	0
ɑ	0	0	0	0	0	0	0	0	0	0	49	0	0	0	0
u	0	0	0	0	0	0	0	0	0	0	0	49	0	0	0
o:	0	0	0	0	0	0	0	0	0	0	0	0	40	9	0
ɔ	0	0	0	0	0	0	0	0	0	0	0	0	2	47	0
ʌu	0	0	0	1	0	0	0	1	0	2	39	0	0	0	6

Table 3. Confusion matrix for the gate ending 120 ms after onset of voicing. Stimuli in left column, responses in top row.

3.2. Hierarchical clustering analysis

Hierarchical clustering analyses were performed on the data for each gate, using the average linkage between groups method, chi-squared dissimilarity measure. Hierarchical clustering analysis determines the perceptual dissimilarity between each pair of vowels, and combines similar vowels into clusters, which combine into larger clusters. The order in which stimuli combine into clusters shows which groups of vowels are most perceptually similar, and thus which distinctions are most clear, at a particular time point (gate). The results can be represented as a dendogram, showing the relative distance (dissimilarity) at which each pair of vowels or clusters form a larger cluster.

The dendograms in Figures 2 and 3 represent the results of the hierarchical clustering analyses for the gates ending at 40 ms and 60 ms after the onset of voicing. Both gates have several large clusters, one consisting of the low vowels, one of the non-low back vowels, one of the front rounded vowels, and one of the front unrounded vowels. This arrangement is typical of all but the latest gates, where there are few confusions.

Within the large non-low back, front rounded, and front unrounded clusters, there are smaller clusters separating the high from the mid vowels. For example, in both Figures 2 and 3, the cluster consisting of front rounded vowels contains smaller clusters, one consisting only of /y/ and the other consisting of /ʏ, ø:/. For both gates, /ʏ/ combines with /ø:/ at a small distance, showing that these two vowels are perceptually very similar. They then combine with /y/ at a much greater distance: they are more similar to /y/ than to other vowels, but much less similar to /y/ than they are to each other. Similarly, in both figures, /ɔ/ and /o:/ combine at a small distance, and the resulting cluster then combines with /u/ at a much larger distance.

The pattern is slightly different for the front unrounded vowels. At the 40 ms gate (Fig. 2), /i, ɪ/ (high vowels) and /ɛ, ɛi, e:, œy/ (mid vowels and diphthongs) form smaller clusters within the front unrounded cluster. (Although /œy/ is

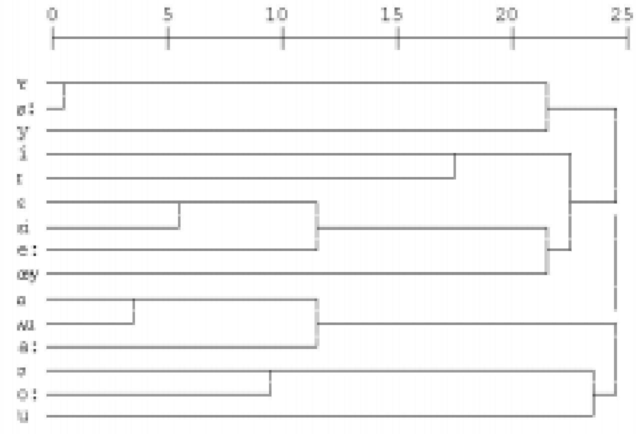

Figure 2. Dendogram for the 40 ms gate. Scale across the top shows normalized distance at which items combine.

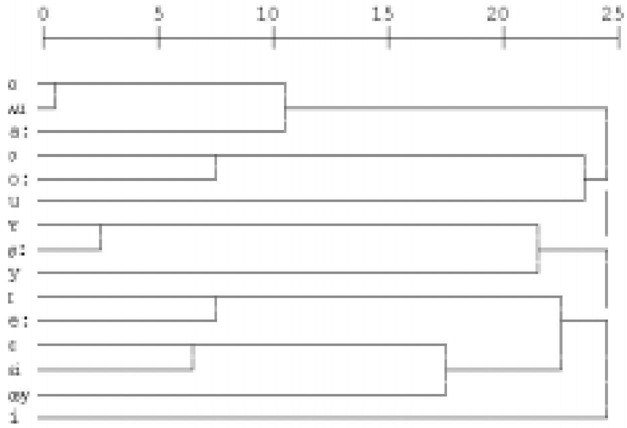

Figure 3. Dendogram for the 60 ms gate.

rounded, rounding has less effect on low vowels, and the beginning of this vowel is very low. Thus it often groups with unrounded vowels at early gates.) However, at the following gate (Figure 3), the front unrounded cluster contains three smaller groupings, one consisting of /i/ alone, one of /ɪ, e:/, and one containing /ɛ, ɛi, œy/. Thus, the two-way division into high vs. mid vowels at the 40 ms gate has changed to a three-way division into high, upper mid, and lower mid vowels at the 60 ms gate.

It is important that the high and mid clusters combine to form the larger clusters (non-low back, front rounded, and front unrounded). While the high vowels are perceptually distinct from the mid vowels, they are more similar to each other than they are to vowels from the other large clusters. For example, /u/ and /o:, ɔ/ are perceptually distinct, but they are more similar to each other than /u/ is to other high vowels or /o, ɔ/ to other mid vowels. The front rounded high and mid vowels are also more similar to each other than to their corresponding front unrounded (or back rounded) vowels.

4. CONCLUSIONS

This data partially supports and partially contradicts the prediction that universally common features are those which can be perceived over a short time window. Within the Dutch vowel system, the four level height distinction, the long/short distinction, and front rounding constitute less common distinctions. High/mid/low (three height levels) and front/back are the distinctions necessary for the universally most common vowel system. In this experiment, even at very short gates (20 ms and 40 ms after onset of voicing), listeners found the high/low and front/back distinctions very perceptible, but they also found the difference between front rounded vowels and front unrounded or back vowels clear. Furthermore, the front rounded/front unrounded/back contrasts were clearer than the difference between high and mid.

Thus, the prediction is not supported for front rounding. If front rounding can be distinguished more easily than high and mid vowels at the beginning of the vowel, why do fewer languages use front rounding than a high/mid distinction? There are several possible explanations. First, the vowels in this experiment were all recorded in the same environment (/hVt/) and presented with the initial consonant intact. Studies which excise vowels from context (either from a variety of isolated word contexts [13] or from connected speech [15]) find far more confusions between front rounded and other vowels than the current study does, and find approximately as many confusions between front rounded and other vowels as between high and mid vowels. Thus, it may be that the perception of front rounding is more affected by consonantal environment than the high/mid contrast is.

Furthermore, vowels are inherently rather long segments, at least when stressed. It may be that perceptibility within a brief window surrounding a point of abrupt acoustic change is important for some features, but not for others. Since vowels are relatively long segments, they might be more likely to manifest distinctions whose cues do not depend on abrupt changes. Diphthongization would be such a non-abrupt cue.

Finally, the number of vowels in the system may provide an explanation. In a five vowel system, it is clear why front rounding is usually not used distinctively. As Figure 1 shows, front rounded vowels are rather central in the vowel space. Thus if a small vowel system included one or more front rounded vowels, the vowels would not be well separated in the vowel space. However, most languages which have seven or eight vowels do include either front rounded or central vowels [1, 17]. Thus, in larger vowel systems, front rounding may not be cross-linguistically uncommon.

The results for the four level height distinction and the long/short distinction do support the hypothesis. Based on just 40 ms of the vowel signal, perceptual confusions group the vowels into high, mid, and low clusters. However, when listeners hear more of the vowel, they begin to distinguish upper mid from lower mid vowels. Thus, the universally common three-way height distinction can be made based on a brief portion of the vowel, but the much less common four-way distinction requires a longer sample of the vowel to perceive. The long and short vowels (in Dutch a difference of quality as well as duration, as in English) are also widely confused with each other based on short portions of the signal, but become clear when more of the signal is available.

Thus, two of the universally less common vocalic features of Dutch require longer to perceive than the more common distinctions. However, front rounding can be perceived better than the high/mid contrast even from very short portions of the signal. These results show that perceptibility over a short time window is a factor in the typology of vowel systems for some but not all vocalic distinctive features.

ACKNOWLEDGMENTS

I would like to thank Anne Cutler and John Ohala for discussion of this material, Roel Smits for help with the statistical analysis, and Niels Janssen for his effort in running the experiment. Errors are, of course, my own.

REFERENCES

[1] Liljencrants, J., and B. Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 48, 839-862.

[2] Ladefoged, P. and I. Maddieson. 1996. *The Sounds of the World's Languages.* Oxford: Blackwell.

[3] Stevens, K.N. 1971. The role of rapid spectrum changes in the production and perception of speech. In Hammerich, L.L., R. Jakobson, and E. Zwirner (eds.), *Form and Substance. Phonetic and Linguistic Papers Presented to Eli Fischer-Jørensen.* 95-101. Kobenhavn: Akademisk Forlag.

[4] Stevens, K.N. 1980. Discussion. *Proceedings of the Ninth International Congress of Phonetic Sciences*, Copenhagen, 1979, 3, 185-186.

[5] Stevens, K.N. 1985. Evidence for the role of acoustic boundaries in the perception of speech sounds. In Fromkin, V.A. (ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged.* 243-255. Orlando: Academic Press.

[6] Stevens, K.N., and S.E. Blumstein. 1981. The search for invariant acoustic correlates of phonetic features. In Eimas, P.D., and J. Miller (eds.), *Perspectives on the Study of Speech.* 1-38. Hillsdale, NJ: Erlbaum.

[7] Stevens, K.N., and S.J. Keyser. 1989. Primary features and their enhancement in consonants. *Language* 65, 81-106.

[8] Stevens, K.N., S.J. Keyser, and H. Kawasaki. 1986. Toward a phonetic and phonological theory of redundant features. In Perkell, J.S., and D.H. Klatt (eds.), *Invariance and Variability in Speech Processes.* 426-463. Hillsdale, NJ: Lawrence Erlbaum.

[9] Lang, C., and J.J. Ohala. 1996. Temporal cues for vowels and universals of vowel inventories. *Proceedings of the Fourth International Conference on Spoken Language Processing,* October 3-6, 1996, Philadelphia.

[10] Booij, G. 1995. *The Phonology of Dutch.* Oxford: Clarendon Press.

[11] Gussenhoven, C. 1992. Illustrations of the IPA: Dutch. *Journal of the International Phonetic Association*, 22, 45-47.

[12] Pols, L.C.W., H.R.C. Tromp, and R. Plomp. 1973. Frequency analysis of Dutch vowels from 50 male speakers. *JASA*, 53, 1093-1101.

[13] Pols, L.C.W. 1977. *Spectral Analysis and Identification of Dutch Vowels in Monosyllabic Words.* Soesterberg: Institute for Perception TNO.

[14] Pols, L.C.W. 1979. Some spectral and perceptual measurements on Dutch diphthongs. In *Anniversaries in Phonetics: Studia Gratulatoria dedicated to Hendrik Mol.* 270-285. Amsterdam: Institute of Phonetic Sciences.

[15] Koopmans-van Beinum, F.J. 1980. *Vowel Contrast Reduction: An Acoustic and Perceptual Study of Dutch Vowels in Various Speech Conditions.* University of Amsterdam dissertation, Academische Pers.

[16] Nearey, T.M., and P.F. Assmann. 1986. Modeling the role of inherent spectral change in vowel identification. *JASA* 80, 1297-1308.

[17] Lindblom, B. 1986. Phonetic universals in vowel systems. In Ohala, J.J., and J.J. Jaeger (eds.), *Experimental Phonology.* 13-44. Academic Press.