ON THE LEXICAL ASPECTS OF VOWEL DISPERSION
THEORY: DUTCH CASE

Louis ten Bosch
Inst Phonetic Sciences, University of Amsterdam, Amsterdam, The Netherlands

Abstract
The 'vowel dispersion theory' states that the structure of the vowel inventory in a language can be explained by optimization of acoustic inter-vowel contrast, given articulatory boundary conditions for each vowel. In this paper, the primacy of the acoustic properties is questioned by considering the possible effect of the lexicon on vowel dispersion. Here, the need for acoustic contrast between two vowels is assumed to be determined by the functional load of the vowel opposition. The results for Dutch indicate that the functional 'load' explains a part of the acoustic structure of the Dutch vowel inventory. Since the model is based on the lexicon only, we emphasize the used methodology, rather than the language-specific results.

1 Introduction
The set of phonemes in a language shows a large variety across languages. Universal trends in the structure of phoneme inventories (known as 'phonological universals') have been observed for a long time and attempts have been made to formulate them explicitly (e.g. Ruben, 1976; Crothers, 1978; Maddieson, 1984, 1991; Liljencrants & Lindblom, 1972; Lindblom, 1986 and later). Quantal Theory: Stevens, 1989; c.f. Ten Bosch & Pols, 1989). In general, the phonetic models of the structure of vowel systems start from two principles: (a) the reduction of articulator effort, and (b) the optimization of inter-vowel acoustic contrast. There is much debate about the adequacy of these principles and their relative weighting. It is well known (see e.g. Ten Bosch, 1991) that a specification of the weighting is essential for the outcome of the optimization, but also less attention has been paid to the relation between the principle of acoustic vowel contrast and the functionality of this contrast (see Lindblom, 1972, 1986; Ten Bosch, 1991, chapter 4; Vallee, 1990). Moreover, with respect to the implementation of the contrast and effort principles, more elaborate models are available now and the vowel dispersion model as well as a general segment inventory model could now be based on articulatory synthesis models and advanced auditory models (An example of the use of more elaborate models is given by the SPEECH MAPS project, 1994).

In this paper, we want to address the point that the principle of 'acoustic contrast' is not based on the functional load of vowel oppositions. For example, if a language has three vowels /a/, /i/ and /u/ that are spectrally specified by three target positions and many minimally pairing words with /i/ and /u/ and only a few with /a/ the need for acoustic contrast between /a/ and /i/ and both other vowels is less. The need for contrast between /i/ and /u/ is a result of the functional load on the lexicon and the frequency of words. Important aspects of the model are focussed upon in the following sections. Results will be presented for the Dutch case. The results are discussed in the concluding section.

2 Influence of lexical structure
Let us assume there are N vowels. For each vowel pair (v1, v2) we select those words from the lexicon that form phonemically minimal pairs with respect to v1 and v2, resulting in a list L1 consisting of words containing v1 that each has one corresponding minimal opposing word containing v2 in the list L2. Additionally, the lists L1 and L2 are constructed so as to contain words with the same grammatical category to allow word confusion that is syntactically possible. Our basic assumption here is that the need for contrast between v1 and v2 is determined by the probability of confusion between L1 and L2, in other words, by the frequency of words being confused by any other vowel from v1, . . . , vN, given the confusion probabilities P(vi —> vi) and uniform distribution of vowels. The exponents Aj that are determined by the lexicon modify the unbiased case into the lexically-balanced case.

P(v1 —> v2) = exp(—Ce12). By substitution in (1) this implies that the following expression is to be minimized: D = Σv1, v2 Aj P(v1 —> v2) exp(—Ce12), in which C denotes a constant that is related to the overall scaling of the acoustic space.

2 The definition of acoustic distance
The distance d1 between vowels v1 and v2 is here determined by the Euclidean distance between the first two formant frequencies, in ERB. The ERB-transformation is performed in order to agree with the frequency selectivity of the human auditory system (Patterson, 1976; Glasberg & Moore, 1990). The formant representation is chosen for two reasons: to allow a match between model predictions and phonologically specified vowel systems, and the findings (e.g. by Wexley-Port & Altschuler, 1985) that Euclidean distances based on bark-transformed formants may highly correlate with judged dissimilarities between vowels.

3 Experimental set-up and results
On the basis of the previous sections, the experiment was set-up as follows. Lists of all lexical items of the same grammatical category in Dutch have been extracted from the CELEX database (CELEX, 1990). The twelve Dutch monophthongs (denoted a, i, u, e, o, E, I, A, y, U, OE, OE), the last two vowels figuring in 'put' and 'peut' in Dutch were selected for comparison. Diphthongs were not taken into account. For each vowel pair (v1, v2), two lists where constructed with corresponding phonemically minimal word pairs with the same grammatical category. For example, the two vowels /O/ and /E/ yield certain lists with 'bof' (Eng. 'bone') and 'beef' (‘bed') figuring in it. The minimal pair /pof/ —> /pef/ (‘rotten' — 'save') is not included since they differ in grammatical category.

On the basis of expression (1), all coefficients Aj were determined. Next, optimal vowel positions were looked for that minimized expression (1). This was done by Kruskal's algorithm, by searching positions in a two-dimensional space, such that P(v1 —> v2) = exp(—Ce12). For the application of Kruskal's algorithm, C = 1 was taken. The optimal systems were found by minimization of the 'stress' which could be defined in a linear or monotonic fashion. Vowel systems were determined for eight

where the word lists L1 and L2 correspond to the distinct vowel pair (v1, v2) and NF denotes a normalisation factor depending on the size of the lexicon. The above expression is symmetric in v1 and v2, since the 'donor' word v1 and the 'receiver' word v2 play an equal role. The psycho-linguistic interpretation of this equal role is that the confusion between a certain given word containing v1 and a minimal pair containing v2 depends on the frequency of u2. It is known that, broadly speaking, the 'accessibility' of words increases with its token frequency; in the above expression it is assumed that this relation is linear.

The consequence is that the former expressions for D are exchanged by the new expression

\[ D = \sum_{v_1, v_2} A_{ij} P(v_1 \rightarrow v_2) \]  

where \( A_{ij} \) are constants that are entirely determined by the structure of the lexicon:

\[ A_{ij} = f(u_1)f(u_2)/NF \]

Writing \( A_{ij}P(v_i \rightarrow v_j) = c_{ij} \), \( D = \sum c_{ij} \) can be approximated by 1 - \( (1 - c_{ij})(1 - c_{kl})..._N \) in other words \( D = \prod_{i,j} (1 - c_{ij}) \) is to be maximized. This latter expression is approximated by

\[ \prod_{i,j} \left( 1 - P(v_i \rightarrow v_j) \right) \]

which reveals a lexically-determined weighing of the expression

\[ \prod_{i,j} (1 - P(v_i \rightarrow v_j)) \]

which returns the probability of v1 not being confused by any other vowel from v1, . . . , vN, given the confusion probabilities P(vi —> vi) and uniform distribution of vowels. The exponents Aj that are determined by the lexicon modify the unbiased case into the lexically-balanced case.

1 Inter-vowel confusion
The second aspect of the model is the relation between inter-vowel confusion and inter-vowel acoustic distance. This aspect is a common feature of each vowel dispersion model. Many models have been proposed (Lindblom, 1972; psychological categorization models, c.f. Smits & ten Bosch, 1994, statistical models). Here we will use
combinations of three binary factors (stress: linear versus monotonic; receiver freq.: token versus lexical; lexical lists: nouns + pronomina only versus all categories). The latter factor refers to the construction of the lists \( L \), whether these consist of nouns and pronomina only, or of all categories. This exception is based on the following table presenting relative lexical and token frequencies for 10 syntactical categories (indicated in the first column). Among the PREP, there are hardly any minimal pairs. The VERB category is excluded since it only contains infinitives.

<table>
<thead>
<tr>
<th>CATEG.</th>
<th>rel. lex. fr.</th>
<th>rel. token fr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>13.5</td>
<td>9.5</td>
</tr>
<tr>
<td>ADV</td>
<td>1.4</td>
<td>8.2</td>
</tr>
<tr>
<td>ART</td>
<td>0.0</td>
<td>10.7</td>
</tr>
<tr>
<td>C</td>
<td>0.1</td>
<td>6.6</td>
</tr>
<tr>
<td>EXP</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>N</td>
<td>72.3</td>
<td>19.1</td>
</tr>
<tr>
<td>NUM</td>
<td>0.2</td>
<td>1.0</td>
</tr>
<tr>
<td>PREP</td>
<td>0.1</td>
<td>13.1</td>
</tr>
<tr>
<td>PRON</td>
<td>0.1</td>
<td>13.3</td>
</tr>
<tr>
<td>V</td>
<td>11.6</td>
<td>18.0</td>
</tr>
</tbody>
</table>

In the following table, the results obtained from Kruskal's algorithm are summarized. For each combination, these results were rank correlated (Spearman) with the actual formant data (derived from Koopmans-Beinum, 1980 and from Van Son & Pols, 1990).

<table>
<thead>
<tr>
<th>combi</th>
<th>Spearman</th>
</tr>
</thead>
<tbody>
<tr>
<td>mtn</td>
<td>0.75</td>
</tr>
<tr>
<td>mtf</td>
<td>0.70</td>
</tr>
<tr>
<td>mln</td>
<td>0.68</td>
</tr>
<tr>
<td>mlf</td>
<td>0.66</td>
</tr>
<tr>
<td>ltn</td>
<td>0.63</td>
</tr>
<tr>
<td>ltf</td>
<td>0.64</td>
</tr>
<tr>
<td>lln</td>
<td>0.53</td>
</tr>
<tr>
<td>lif</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Combinations are indicated by a three-letter combination, referring to the combination monotonous – linear, token – lexical, and (noun+pronomina) (‘noun’) – all categories (‘full’). The difference between combination number 6 and 7 is significant, as well as is the difference between 1 and 4, 2 and 5, 3 and 6, and larger differences. The results are optimized across many (> 200) random start configurations.

Among the monotonic options, the mtn option yields the optimal Spearman correlation with actual data (token frequency, nouns + pronomina). The corresponding vowel system is shown in figure 1. The contour lines connect the formant positions corresponding to ‘equal articulatory effort’ as proposed in ten Bosch (1991). The 12 monophonograms are plotted in the figure in such a way that the resulting configuration resembles the actual situation (Kruskal's data are specified up to an overall factor, up to rotations and reflections in the formant space). Among the linear options, the ’iff’ combination yields the highest Spearman correlation. In this setting, Kruskal's algorithm attempts to optimally match the inter-vowel distances on the basis of the inter-vowel confusion probabilities, based on token frequencies and all syntactical categories. The corresponding optimal vowel system in the ‘iff’-case is shown in figure 2.

4 Discussion.

The table presented above shows that the match between predicted and actual vowel system is larger in the monotonous case than it is in the linear case. In fact, the condition in the linear case is harder to meet. Given the monotonic and linear option, the results for the token frequency (slightly) outperform the results obtained with the lexical frequency. This is in line with our expectation. The differences between the options (noun+pronomina) (‘noun’) – all categories (‘full’) are small and in fact not significant.

Both figure 1 and 2 show that the lexical structure of Dutch explains a part of the structure of the Dutch vowel system. There are, however, a few remarkable errors. In the monotonic option (figure 1), the position of the short /i/ and /A/ are remarkable. Globally, the triangle-like structure is preserved, but especially the short vowels are not located in coherence with their known acoustic specification. The distance between /A/ and /O/ is larger than expected. This is related to the fact that the number of minimally opposing words for these vowels is large (ten Bosch, 1991). Also in figure 2 (referring to the linear option), the /i/, /a/ and /u/ do not span the vowel triangle any more. The short /A/ lies further from the center than /a/ does. Also here, the distance between /A/ and /O/ is larger than expected.

In general, the localisation of the vowels /U/ and /O/ is more distinct here, the distance between /A/ and /O/ is larger than expected. This is related to the fact that the number of minimally opposing words for these vowels is large (ten Bosch, 1991). Also in figure 2 (referring to the linear option), the /i/, /a/ and /u/ do not span the vowel triangle any more. The short /A/ lies further from the center than /a/ does. Also here, the distance between /A/ and /O/ is larger than expected.

In general, the localisation of the vowels /U/ and /O/ is more distinct here, the distance between /A/ and /O/ is larger than expected. This is related to the fact that the number of minimally opposing words for these vowels is large (ten Bosch, 1991). Also in figure 2 (referring to the linear option), the /i/, /a/ and /u/ do not span the vowel triangle any more. The short /A/ lies further from the center than /a/ does. Also here, the distance between /A/ and /O/ is larger than expected.

The integration of the length opposition into an acoustic contrast measure based on spectral and durational contrasts is troublesome (see e.g. ten Bosch, 1991). How duration is to be included remains unclear.

ACKNOWLEDGMENT

This research is sponsored by the University of Amsterdam and by the Dutch foundation NWO.

REFERENCES


