Institute of Phonetic Sciences,
University of Amsterdam,
Proceedings 19 (1995), 39-50.

LEXICALLY-BASED VOWEL DISPERSION: A CASE STUDY FOR DUTCH[1]

Louis F.M. ten Bosch

Abstract

The 'vowel dispersion theory' states that the acoustic structure of the vowel inventory in a language can be explained by optimizing the acoustic inter-vowel contrast, under the constraint of articulatory conditions. In this paper, the primacy of the acoustic principles of contrast and effort is questioned by considering the possible effect of the lexicon on vowel dispersion. As an extreme point of view, the need for acoustic contrast between two vowels will be assumed to be determined only by the 'functional load' of the vowel opposition. This functional load is determined by the lexicon. The results for Dutch indicate that the functional load explains at least a part of the acoustic structure of the Dutch vowel inventory. Since the model is tested for one language 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 phonological trends in the structure of phoneme inventories (which have become known as 'phonological universals') have been observed for a long time and attempts have been made to formulate them explicitly, both in linguistic and in phonetic terms (e.g. Ruhlen, 1976; Crothers, 1978; Koopmans-van Beinum, 1980; Koopmans-van Beinum, 1983; Maddieson, 1984; Liljencrants & Lindblom, 1972; Lindblom, 1986; Quantal Theory: Stevens, 1989; Ten Bosch & Pols, 1989; Ten Bosch, 1991; Svantesson, 1995; Schwartz, Boë & Vallée, 1995; Iivonen, 1995). Some of these models are based on a phonological viewpoint, while other models use more acoustic-phonetic principles. These acoustic-phonetic models, aiming at the explanation of the structure of *vowel* systems, have shown to be quite successful, if they are matched to the phonological data available.

Broadly speaking, most phonetic models of the structure of vowel systems start from two principles: (a) the reduction of articulatory effort, and (b) the optimization of inter-vowel acoustic contrast. These principles are a direct consequence of the fact that vowels have to be produced and are meant to be perceived, and that the corresponding effort at the speaker's and the listener's side is likely to be minimized. Such a minimization principle has already been recognized by linguists in the past century. In fact, such a principle means that vowel systems, or, more generally, segment inventories, are governed by a principle of 'least effort', such that an inventory of

[1] This is a revised version of a paper, published in the Proceedings of the 1995 International Conference of Phonetic Sciences, Stockholm.

which segments are hard to pronounce or in which segments are hard to distinguish perceptually will be less favourable than a system in which segments are easy to pronounce and easy to identify. The easier the inventory, the more probable its existence in real life, and as a consequence, one can say that existing systems tend to 'minimize' the involved required effort of production and/or perception.

There is much debate, however, about the adequacy of these principles. First of all, purely linguistic principles such as (vowel) symmetry are not taken into account at all. These linguistic principles will not be discussed here. Secondly, one might debate the relative weighing of the articulatory and perceptual principle, and the exact quantification of effort. The perceptually-based quantification of the difference of segments in general is troublesome since many segments have a dynamic character, i.e. change over time, and the perceptual difference is not known to simply integrate over time. Only for (steady) vowels, more or less substantial results have been reported (by Kewley-Port and Atal, 1989, for example). At the production side, it is well known that languages exist with a very rich complex consonant inventory, and it would certainly be not correct to claim that a language just strives for minimal articulatory effort, although indeed there is a slight tendency to simple consonant systems (Maddieson, 1984). Moreover, it is well known (see e.g. ten Bosch, 1991) that, given quantifications of perceptual contrast and articulatory effort, a numerical specification of the weighing between them is essential for the outcome of the model. We here leave aside the problem of the difference between context-dependent and context-independent notions of effort and *ditto* for perceptual contrast.

With respect to the quantification of contrast and articulatory effort of segments in isolation, more elaborate models become available now. (An example of the use of more elaborate models is given by Abry et al., 1994). This means that new, more complicated phonetic models can be designed that aim at the explanation of phonologically and phonetically specified segment inventories. In particular, vowel models that attempt to explain the phenomenon of vowel dispersion that is observed in the majority of languages, could now be based on articulatory synthesis models and advanced auditory models.

In this paper, we want to address a totally different point, in which the structure of vowel systems is based on the 'functional load' of vowel oppositions. For example, if a hypothetical language has only three vowels /a/, /i/ and /u/, and many minimally pairing words with /i/ and /u/ and only a few with /a/, the *need* for acoustic contrast between /a/ and both other vowels will be less than the need for acoustic contrast between /i/ and /u/. This difference will in one way or another be reflected in the acoustic distances between these three vowels: the need for acoustic difference between /i/ and /u/ will be larger then between the other two vowel pairs. In the present model, it will be assumed that the need for acoustic contrast between two vowels is directly based on the capability of these two vowels to distinguish words in the lexicon. This need will therefore be related to (a) the lexical distribution of minimal word pairs, (b) the (token) frequency of words, and (c) a model that relates inter-vowel distance with inter-vowel confusion.

In the next sections, the notion of functionality, as well as a model to relate lexical structure with inter-vowel distances will be discussed. Next, results will be presented for the Dutch case. A discussion follows in the concluding section.

2. Lexical structure

Let us assume that a language has exactly N stable vowels of which the acoustic realisation is context-independent. Context-independency is an essential technical constraint: the vowel dispersion model aims at an explanation of the structure of the 'vowel system' of the language, without reference to actual pronunciations in consonantal environments. For each vowel pair (v_1, v_2) we can select all those words from the lexicon that form phonemically minimal pairs with respect to v_1 and v_2, resulting in two lists L_1 and L_2. The list L_1 consists of words containing v_1 each having one corresponding minimally pairing word containing v_2 in the list L_2. So the vowel pair (v_1, v_2) and the lexicon completely specify the lists L_1 and L_2, and these lists are independent of the ordering of the vowels. The notion of 'phonemically minimal' can be based (is to be based) on the norm phonetic transcription of the words. Additionally, the lists L_1 and L_2 can be constrained so as to contain words within the *same* grammatical category to allow word confusion that is syntactically possible.

As an example, the /E/-/I/ opposition leads to two Dutch lists. If we select the noun pairs, two lists are obtained, one containing (among other words) /sEnt/, /dEs/, /klEp/, /sxEp@r/, /b@klEmIN/, the other list containing the corresponding pair members /sInt/, /dIs/, /klIp/, /sxIp@r/, /b@klImIN/. Here '@' denotes the schwa, and 'N' the velar nasal. The two short vowels /O/ and /E/ yield two lists with /bOt/ (Eng. 'bone') and /bEt/ ('bed') figuring in it. However, the minimal pair /rOt/ -- /rEt/ ('rotten' - 'save') will never be included in any list since these words differ in grammatical category. In order to give an idea of the size of these lists: the number of minimal one-syllable noun pairs is 4295, for two syllables 1175, and for three syllables 251. These data are based on CELEX (1990).

The basic assumption here is that the need for contrast between v_1 and v_2 is determined by the probability of confusion between the words figuring in the lists L_1 and L_2, more precisely, by the (lexical or token) frequency of each word in L_1 and in L_2. If the frequency of word w is denoted by f(w), the probability of the overall word confusion due to vowel confusion is given by

$$\sum_W f(w)\, P(w) / ls$$

The sum is taken over all words w in the lexicon, ls denotes the lexicon size, and P(w) denotes the probability of confusing a word w with a minimally pairing word that differs by (just) one vowel. This confusion probability can be rewritten as

$$\sum \left\{ P(v_1 > v_2)\, f(w_1)\, f(w_2) \right\} / NF$$

the sum taken over each vowel pair (v_1, v_2) and all minimal word pairs (w_1, w_2) from L_1 and L_2, where the word lists L_1 and L_2 correspond to the distinct vowel pair (v_1, v_2) as described above. NF denotes a normalisation factor depending on the size of the lexicon. $P(v_1 > v_2)$ denotes the probability of acoustically confusing the 'stimulus' vowel v_1 as v_2. The above expression is symmetric in w_1 and w_2, since the 'donor' word w_1 and the 'receiver' word w_2 have been assumed to play an equal role. The psycholinguistic interpretation of this equal role is that the confusion between a certain *given* word containing v_1 and a minimally pairing word containing v_2 only depends on the (token) frequency of w_2. For example, if the acoustic signal represents /sEnt/, the probability of perceiving /sInt/ increases with the token frequency of the word /sInt/. Accordingly, if the inter-vowel confusion probability between /I/ en /E/ is symmetric, and if /sEnt/ has a token frequency of 5000 and /sInt/ has a token frequency of 50000, then it follows from the formula that the probability of perceiving /sInt/ is 10 times higher than the probability of perceiving /sEnt/ *once the counterpart word has been presented.* If the probability of perceiving /E/ when /I/ is presented P(I > E)) is ten times higher than P(E > I), the probability of perceiving /sInt/ is equal to the probability of perceiving /sEnt/. The inter-vowel confusion is compensated by the token frequency of the 'patient' word, and that is just what the formula does. It is known that, broadly speaking, for the listener the 'accessibility' of words increases with their token frequency; in the above expression it is assumed that this relation is linear. This is, of course, a very drastic simplification. If the accessibility of words would be completely *in*dependent of its (token) frequency, the above formula should be adapted to

$$\sum \left\{ P(v_1 > v_2)\ f(w_1) \right\} / NF$$

An appropriate expression could be

$$\sum \left\{ P(v_1 > v_2)\ f(w_1)\ f(w_2)^{\tau} \right\} / NF$$

τ denoting an exponent (between 0 and 1, vowel-independent and word-independent) to be estimated or determined by psycholinguistic data.

Resuming, we have the following situation. On the basis of the lexicon, pairs of lists can be determined for each vowel pair in the language. By the formula given above, we have an expression for the confusion between minimally pairing words in terms of the known frequencies of these words and the unknown probabilities $P(v_i > v_j)$. The flow of the arguments is now as follows: since there is a relation between the acoustic specification (for example: formant position) of vowels and the confusion probabilities $P(v_i > v_j)$ on the one hand, and between $P(v_i > v_j)$ and the overall probability of confusing words on the other hand, there exists an (indirect) relation between the acoustic specification of vowels and the overall probability of confusing minimal word pairs. The (optimization of) functional contrast between vowels can therefore directly be defined in terms of (minimization of) the probability of confusing minimal word pairs. Therefore, an acoustic specification of the Dutch vowel system can be looked for that minimizes the overall confusion between minimal word pairs. Consequently, it can be attempted to find the 'optimal' vowel system for e.g. Dutch, if we have a sufficiently long list of frequent Dutch words, an adequate norm description of each word in terms of phonemes, all the (token and type) frequencies, and a model relating acoustic distance between vowels and the probability to confuse them. In the next sections, we will discuss the aspects inter-vowel confusion, acoustic distance, and the experimental set-up and results.

The formulae given above allow a neat interpretation in terms of probability theory. This will be discussed in Appendix 1.

3. Inter-vowel confusion

As observed earlier, an important aspect of the model is the relation between inter-vowel confusion and inter-vowel acoustic distance. This aspect is in fact a common feature of each vowel dispersion model. Many models have been proposed, for example based on the classical vowel dispersion model (Liljencrants & Lindblom, 1972).

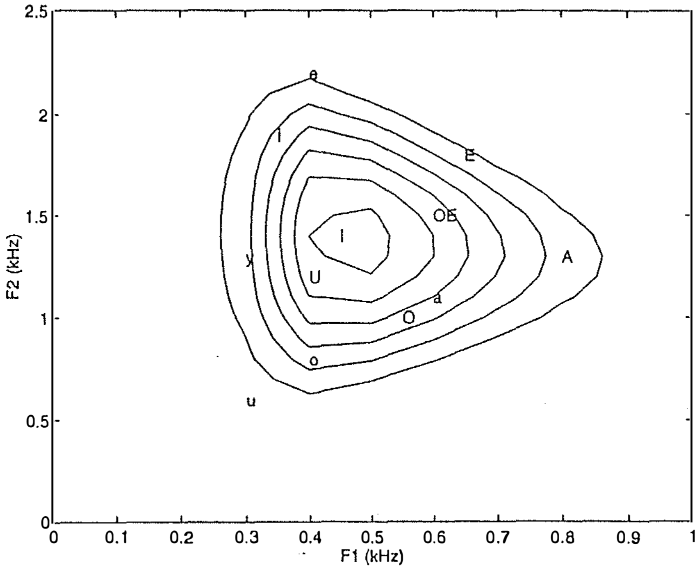

Figure 1. The optimal vowel configuration by Kruskal's algorithm using monotonic stress and token frequencies, based on nouns and pronomina.

Here we will assume an exponential relation between the inter-vowel confusion probability $P(v_1 > v_2)$ and the inter-vowel acoustic distance d_{12}:

$$P(v_1 > v_2) = \exp(-C.d_{12})$$

with C a positive (scaling) constant that is related to the overall scaling of the acoustic space. The assumption implies that $P(v_1 > v_2) = P(v_2 > v_1)$, so $P(v_2 \mid v_1) = P(v_1 \mid v_2)$, i.e. the confusion matrix for vowels is symmetric. Asymmetrical confusion matrices may be used, but the quantitative aspects in this model will increase in difficulty. The model states that if the acoustic distance is zero, the confusion probability is maximal. Lindblom (1986) suggest a relation of the form $P(v_1 > v_2) = 1/d_{12}^2$, which evidently yields singularities if the acoustic distance is small.

4. The definition of acoustic distance

The distance d_{ij} between vowels v_i and v_j is here determined by the Euclidean distance between the first two formant frequencies, after a transformation from Hz to an ERB-scale. The ERB-transformation is applied in order to optimally agree with the frequency selectivity of the human auditory system (Patterson, 1976; Glasberg & Moore, 1990). The *formant* representation is chosen for two reasons: (a) to allow a match between model predictions and phonologically specified vowel systems, and (b) the findings that Euclidean distances based on bark-transformed formants may highly correlate with *judged dissimilarities* between vowels (e.g. Kewley-Port & Atal, 1989). The differences between the bark-representation and the ERB-representation are in this respect of minor importance.

5. 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

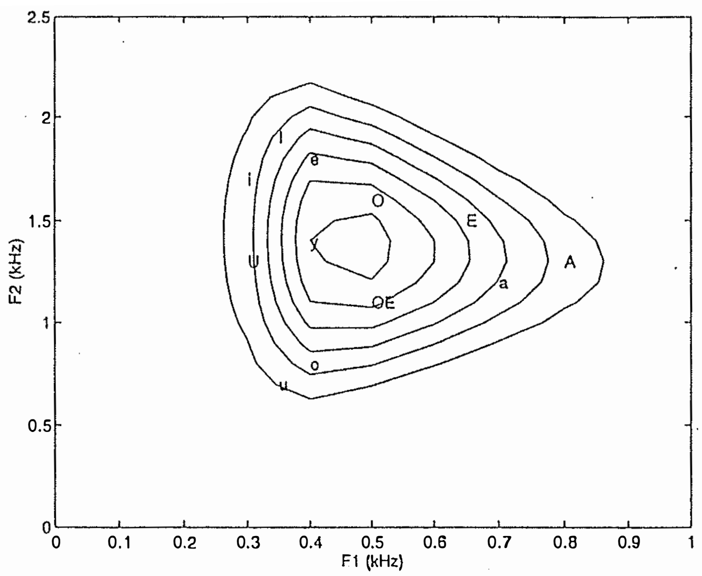

Figure 2. The optimal vowel configuration by Kruskal's algorithm using linear stress, and token frequencies, based on all grammatical categories.

CELEX database (CELEX, 1990). The twelve Dutch monophthongs (denoted a, i, u, e, o, E, O, I, A, y, U, OE, the last two vowels figuring in two Dutch words with orthography 'put' and 'peut') were selected for comparison. The schwa and the three diphthongs were not taken into account. As explained above, for each vowel pair (v_1,v_2), two lists where constructed with corresponding phonemically minimal word pairs with the same grammatical category.

On the basis of the lists, the following coefficients

$$A_{ij} = \sum \left\{ f(w_1)\ f(w_2) \right\} / NF$$

were determined, both by taking the token frequencies as well as the type (lexical) frequencies. Next, vowel positions were searched such that

Table I. Relative lexical (type) and token frequencies for 10 grammatical categories in Dutch. Data from the CELEX database (1990).

CATEGORY	rel. lexical freq.	rel. token freq.
ADJ	13.8	9.5
ADV	1.4	8.2
ART	0.0	10.7
C	0.1	6.6
EXP	0.1	0.0
N	72.3	19.1
NUM	0.2	1.0
PREP	0.1	13.1
PRON	0.1	13.3
V	11.6	18.0

$$D = \sum A_{ij} P(v_i > v_j)$$

was minimized (see Appendix for details). This minimization was done by Kruskal's algorithm, by searching positions in a two-dimensional space, such that the Kruskal stress between the distances in the output of the Kruskal algorithm and the distances on the basis of

$$d_{ij} = -\log(P(v_i > v_j))/C$$

was minimized. For the application of Kruskal's algorithm, C = 1 was taken. (The value of C is not relevant for the result of Kruskal's algorithm, as long as it is fixed during minimization.) In order to study the robustness of the found vowel configurations, vowel systems have been determined for all eight combinations of three important binary factors: *stress*, *word frequency definition*, and the *structure of lexical lists*. The Kruskal factor *stress* refers to the possibility of finding an optimum vowel configuration by using a linear ('l') or monotonic ('m') fashion in terms of the mismatch between the matrix of actual inter-vowel distances on the one hand, and the desired distances on the other hand. A monotonic fit is just an ordinal fit. The second factor (*word frequency definition*) refers to the possibilities of defining the frequency of a word on the basis of token frequency ('t') or on the basis of lexical ('l') frequency. The third factor (*structure of lexical lists*) refers to the construction of the lists L_i, whether these lists are constrained so as to contain nouns ('n') and pronomina (PRON) only, or to contain all categories ('a') instead. This third factor is based on the following table presenting relative lexical and token frequencies for 10 syntactical categories (indicated in the first column of table I.). Articles (ART), expletives (EXP), adjectives (ADJ) and other, numerically minor categories are not considered. Among the prepositions (PREP), there are hardly any minimal pairs. The verb (V) category, although showing a high type and token probability, is excluded from figuring in the lists L_i since it only contains infinitives.

In table II, the results obtained from Kruskal's algorithm are summarized in terms of Spearman rank correlations on the basis of the inter-vowel distances between the model output and actual formant data for Dutch (derived from Koopmans-van Beinum, 1980 and from Van Son & Pols, 1990). The second column specifies the used factor combinations by a three-letter code using 'm' or 'l', 't' or 'l', and 'n' or 'a', referring to the combination of the three binary factors *stress*, *word frequency definition*, and

Table II. Results of 8 different Kruskal optimizations. The Kruskal stress factor, word frequency definition, and structure of lexical lists is specified in the second column. The Spearman rank correlation between the output of Kruskal's algorithm and the actual data is specified in column three. For a description of the factors see the text.

	factor comb.	Spearman
1	m, t, n	0.75
2	m, t, a	0.70
3	m, l, n	0.68
4	m, l, a	0.66
5	l, t, n	0.63
6	l, t, a	0.64
7	l, l, n	0.53
8	l, l, a	0.54

structure of lexical lists: monotonous - linear, token - lexical, and (noun + pronomina) - all categories. 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 initial vowel configurations, to avoid locally optimal solutions.

Among the monotonic options (option 1 to 4 in table II), the 'm, t, n' option yields the highest 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 monophthongs are plotted in the figure in such a way that the resulting configuration optimally resembles the actual (F1-F2) situation. This has been done by rotating, shifting and/or mirroring the output of Kruskal's algorithm so as to optimize the match between the model solution and the known actual formant data. This post-processing of the output in the formant space is allowed (and required) since it is only specified up to an overall omnidirectional scaling factor, up to rotations, and up to line reflections in the formant space.

Among the linear options (option 5 to 8 in table II), the 'l, t, a' 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 this is shown in figure 2.

6. Discussion

Table II 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. Evidently, the condition in the linear case is harder to meet, since monotonicity involves a relaxation of the linear constraint. 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) ('n') and all categories ('a') are small, and most likely not significant.

Vowel triangle traceable
Both figure 1 and 2 show that the lexical structure of Dutch explains at least a part of the structure of the Dutch vowel system. This is interesting, since the structure was based on optimization on the basis of minimal word pairs only, without any reference to acoustic-phonetic interpretations of the phoneme symbols. There are, however, a few discrepancies. In the monotonic option (figure 1), the acoustic 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 acoustic distance between /A/ and /O/ is larger than expected. This is related to the fact that the number of minimally opposing words for these acoustically close vowels is surprisingly large for Dutch (ten Bosch, 1991). Also in figure 2 (referring to the linear option), the /i/, /a/ and /u/ do not span the vowel triangle. For example, the short /A/ lies further from the center than /a/ does. Also here, the distance between /A/ and /O/ is larger than expected. In both options, the location of the vowels /U/ from Dutch 'put') and /OE/ (from 'peut') is not precise. Nevertheless, the triangle-like structure of the vowel system, at least for the monophthongs, is traceable.

Long versus short vowels
Apart from the question how to integrate *diphthongs* (these are entirely excluded here), there is another issue to be addressed, viz. the distinction between long and short vowels. In fact, we studied the 12 monophthongs without any reference to length differences. 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 therefore unclear. A difference in duration contributes to the overall perceived dissimilarity between vowels, and one might think of an expression such as

$$\mathrm{diss}(v_1, v_2) = \mathrm{expr}(d_{spec}(v_1, v_2), d_{dur}(v_1, v_2))$$

in which $\mathrm{diss}(v_1, v_2)$, $d_{spec}(v_1, v_2)$ and $d_{dur}(v_1, v_2)$ denote the overall dissimilarity between the vowels v_1 and v_2, the dissimilarity based on the spectral distance between the vowels, and the dissimilarity between the vowels as a consequence of a difference in (acoustic) duration. 'expr' denotes an expression that is still to be determined. It is however, a problem of subtle weighing between all these factors to get interpretable output of any optimization algorithm.

Metric in the vowel space
In fact, inter-vowel confusion and the definition of the acoustic-phonetic metric involves more care. In this respect, the choices in the model can easily be elaborated. A possible improvement of the definition of 'acoustic contrast' may involve the use of the first *cepstral* coefficients based on a spectral representation of an acoustic 'norm realisation' of each vowel. In automatic speech recognition systems, the cepstra prove to be a robust acoustic representation of speech segments given context-dependence and speaker variability. The distance between vowels may in that case be based on the Mahalanobis distance (weighted Euclidean distance), if necessary with diagonal covariance matrices. The relation between vowel confusion and this elaborated distance measure, however, is a more psycholinguistic aspect of the model. As is well-known, there is a difference in judged dissimilarity of vowel like segments when presented in isolation compared to the case in which they are presented in context. A vowel dispersion model should account for that difference or at least correct for possible bias effects.

Asymmetry
Another aspect that might be relevant for the generalizibility of the model concerns the possibility of having an asymmetric vowel confusion matrix based on a symmetric vowel distance matrix. It has been observed (Weenink, personal communication) that vowel confusion matrices for the short vowels only show a tendency for a vowel stimulus to be perceived with a lower first and second formant, specially if the vowel stimuli are short (a bit shorter than their average duration in spontaneous speech). This means, for example, that the probability of an /A/ being confused with an /O/ is much larger then the probability of an /O/ being confused with an /A/. The perception experiments have been performed by extracting stable vowel portions, taken from the mid portion of the vowel.

This suggests that the vowel confusion matrix, although based on the symmetric distance matrix, results by a (psycholinguistically or psycho-acoustically motivated) bias towards the stimuli with smaller formants.

Output validation
A problem that arises when vowel dispersion models are enriched with more sophisticated 'modules' is the validation of the output. In most cases the acoustic specification of the acoustic-phonetic data, to which the output of the Kruskal algorithm should be matched, is insufficient to justify the use of complicated model designs. For example, if one is tempted to explain the 'structure' of the vowel inventories of the languages in the word, by setting up an acoustic-phonetic model and by matching its output with phonological databases, it is of no importance to have the model super-specify the acoustic properties of the phonological segments, since this is not relevant in the matching procedure.

Conclusion
In this paper, a model has been presented that aims at the explanation of the Dutch vowel inventory by using a lexically based contrast. The model is based on a number of explicit assumptions, concerning the validity of the relation between vowel confusion and vowel distance and the symmetry of the confusion matrix, the use of the probabilities in the way described above, and the entire neglection of the direct need for acoustic contrast itself. It furthermore does not take into account notions such as the dynamic interpretation of contrast and articulatory effort, i.e. contrast and effort in context. Probably, the structure of vowel inventories is a result of a mixture of linguistic, acoustic-phonetic and pragmatic factors that cannot be disentangled properly.

Acknowledgement

This research is sponsored by the University of Amsterdam and by the Dutch Organisation for Scientific Research NWO. Valuable comments by Louis Pols and editorial support is gratefully acknowledged.

References

Abry, C., Badin, P., and Scully, C. (**1994**). 'Sound-to-gesture inversion in speech: The *Speech Maps* approach'. ESPRIT Research Report N^{o}6975. In Varghese K., Pfleger S., and Lefèvre J.P. (Eds) *Advanced speech applications*, Springer Verlag: Berlin, 182-196.

Bosch, L.F.M. ten (1991). *On the structure of vowel systems. Aspects of an extended vowel model using effort and contrast.* Doctoral dissertation, University of Amsterdam.

Bosch, L.F.M. ten, and Pols, L.C.W. (1989). 'On the necessity of quantal assumptions'. *Journal of Phonetics* **17**, 63-70.

Bosch, L.F.M. ten (**1995**). 'On the lexical aspects of vowel dispersion theory: Dutch case'. *Proceedings of ICPhS 95*, Stockholm, 420-423.

CELEX (1990). A program for retrieval of lexical information (for Dutch, English, German). Centre for lexical information, University of Nijmegen, The Netherlands.

Crothers, J. (1978). 'Typology and universals of vowel systems'. In: *Universals of human language.* Vol. 2: Phonology (J.H. Greenberg, ed.). Stanford, Cal., Stanford Univ. Press. 93-152.

Glasberg, B.R., and Moore, B.C.J. (1990). 'Derivation of auditory filter shapes from notched-noise data'. *Hearing Research* **47**, 103-138.

Iivonen, A. (1995). 'Number of possible basic vowel qualities and their psychoacoustical distance measure'. *Proceedings ICPhS*, Stockholm, 404-407.

Kewley-Port, D. and Atal, B. (1989). 'Perceptual differences between vowels located in a limited phonetic space.' *J. Acoust. Soc. Am.* **85**, 1726-1740.

Koopmans-van Beinum, F. (1980). *Vowel contrast reduction: An acoustic and perceptual study of Dutch vowels in various speech conditions.* Doctoral dissertation, University of Amsterdam.

Koopmans-van Beinum, F. (1983). 'Systematics in vowel systems.' In: *Sounds and Structures. Studies for Anthonie Cohen.* (M. van den Broecke, V. van Heuven, W. Zonneveld, eds.) Foris, Dordrecht, 159-171.

Maddieson, I. (1984). *Patterns of sound.* (Cambridge studies in speech sciences and communication). Cambridge Univ. Press.

Liljencrants, J. and Lindblom, B. (1972). 'Numerical simulation of vowel quality systems: the role of perceptual contrast'. *Language* **48**, 839-862.

Lindblom, B. (1986*). Phonetic universals in vowel systems. In: Experimental Phonology* (J. Ohala and J. Jager, eds.). Academic Press, Orlando, Florida. 13-44.

Patterson, R.D. (1976). 'Auditory filter shapes derived with noise stimuli'. *J. Acoust. Soc. Am.* **59**, 640-654.

Ruhlen, M. (1976*). A guide to the languages of the world. Language Universals Project*, Stanford Univ. Press.

Schwartz, J.-L., Boë, L.J., Vallée, N. (1995). 'Testing the dispersion-focalization theory: phase space for vowel systems.' *Proceedings ICPhS*, Stockholm. 412-415.

Svantesson, J.-O. (1995). 'Phonetic evidence for the great Mongolian vowel shift.' *Proceedings ICPhS*, Stockholm. 416-419.

Van Son, R., and Pols, L.C.W. (1990). 'Formant frequencies of Dutch vowels in a text, read at normal and fast rate'. *J. Acoust. Soc. Am.* **87**. 1683-1693.

Vallée, N. (1990). *Typology des systemes vocales.* Report de l'Institut de la Communication Parlée, Grenoble (Fr.).

Appendix

In section 2, we have observed that a lexically-based expression D, indicating the lexically-based *dispersion* of a vowel system, is basically of the form

$$D = \sum A_{ij}\, P(v_i > v_j)$$

the sum to be taken over all vowel pairs, where A_{ij} are constants that are entirely determined by the structure of the lexicon:

$$A_{ij} = \sum \left\{ f(w_1).f(w_2) \right\} / NF$$

where the sum is taken over all words w_1 in L_1 and all words w_2 in L_2. Observe that D is to be minimized (1/D might therefore be a better definition of dispersion, from a purely numerical point of view).

It is possible to interpret the relation of the formulae above in terms of probability theory. Writing $A_{ij}\, P(v_i > v_j) = e_{ij}$, it is assumed that these values e_{ij} are small (this means that the probability of confusing minimally pairing words is still quite small, much less than 1). In that case, $D = \Sigma e_{ij}$ can (in first order) be approximated by

$$1 - (1-e_{12})(1-e_{13})\ldots(1-e_{(N-1),N})$$

which is still to be minimized, in other words, $\Pi\,(1-e_{ij})$ is to be maximized, the product to be taken over all vowel pairs. This latter expression is approximated by

$$\prod (1 - P(v_i > v_j)) \;{*}{*}\; A_{ij}$$

(** denoting the power function) which reveals a lexically-determined weighing of the 'flat' unbiased expression

$$\prod (1 - P(v_i > v_j))$$

which returns the probability of v_i *not* being confused by any other vowel from $v_1, ..., v_N$, given the confusion probabilities $P(v_i > v_j)$ and a uniform *a priori* distribution of the vowels. The exponents A_{ij} that are determined by the lexicon modify the unbiased case into the lexically-balanced case.