Confusions between Dutch Consonants under Various Conditions of Noise and Reverberation

L.C.W. Pols and I. Stoop
Amsterdam and Leyden, the Netherlands

1. Introduction

Ever since the Miller and Nicely (1955) consonant identification experiment there has been a vivid interest in representing the perceptual differences and similarities between consonants. Idenfication under noisy or other disturbing conditions, paired or triadic comparison, scaling, and memory recall are some of the procedures used to achieve confusion or similarity data.

Miller and Nicely only had rather simple means available to process their confusion matrices and furthermore described their results in terms of predefined distinctive feature systems. We would prefer to use multidimensional scaling techniques and let the data more or less speak for themselves. Over the years this has been done several times with the original Miller and Nicely data, e.g. Shepard (1974), Wish and Carroll (1974), Soli and Arabie (1979), and Kroonenberg and De Leeuw (1980).

However, within the framework of a project about speech intelligibility and listening comfort in noisy and reverberant conditions, we had the opportunity to collect a large amount of new consonant identification data. We feel that there are various interesting aspects to these new data and to the way they have been processed:

- they concern a language different from English, namely Dutch;
- it is a large data set (17 consonants, 28 acoustically different conditions, 6 speakers, 5 listeners);
- initial, medial, and final consonants in CVCVC words were identified;
- not just different types of noise were used as acoustic disturbances but also reverberation, plus combinations of the two;
- advanced multidimensional scaling techniques were applied to the data;
- the present material is relevant with respect to noise legislation and listening comfort, although this point will not be discussed any further in this paper (see Pols, 1981a), nor will the links with the speech transmission index STI related to speech intelligibility (see Pols, 1981a and Houtgast et al., 1980).

2. Experimental Procedure

Three male and three female Dutch speakers read 20 different lists with 51

CVCVC nonsense words embedded in short carrier phrases, for instance /hɛt wort *limal* ɛində/. The 17 Dutch consonants /p,t,k,b,d,f,s,χ,v,z,h,m,n,l, r,w,j/ occurred three times in each list. Since in Dutch several voiced consonants do not occur in final position, there were 11 different final consonants /p,t,k,f,s,χ,m,n,ŋ,l,r/. The consonants were combined with 12 vowels and three diphthongs. Five subjects listened to these recorded word lists under various conditions of noise and reverberations, and had to identify the three consonants in each CVCVC word. The 28 different conditions were a mixture of four reverberation times (T = 0,05, 1 and 1,5 s), five signal-to-noise ratios (SNR > 50, + 15, + 8, + 1, and -6 dB), and five noise spectra, see Table I. The different noise spectra mainly differed in their high-frequency energy and were supposed to be representative of various types of ambient noise.

Table I. Specification of all 28 listening conditions in terms of type of noise spectrum used, signal-to-noise ratio, and reverberation time

T in S	SNR in dB					Type of noise spectrum
	50	+15	+8	+1	-6	
0	1	2	3	4	5	Speech noise
0.5	6	7	8	9	10	Speech noise
1	11	12	13	14	15	Speech noise
1.5	16	17	18	19	20	Speech noise
0.5			21	22		Traffic noise
0.5			23	24		Indoor traffic noise
0.5			25	26		Train/airplane noise
0.5			27	28		Industrial noise

After two days of training the listeners got, on each of the following days, all 28 conditions in random order for one speaker, plus some standard conditions for practice and reference. The identification experiment was computer controlled and all responses were stored on-line for subsequent data processing.

3. Results

For an evaluation of the various percentages correct scores under the various conditions we refer to Pols (1981a, 1981b); in this paper we will concentrate on the structure of the confusion matrices. The present data require a so-called three-mode model, namely stimuli x responses x conditions. These models are also referred to as individual difference models; these so-called individual differences are replaced in our data by differences caused by the acoustic conditions. For the time being we have neglected the real individual differences in our data, namely those caused by different speakers and different listeners, and used cumulative data. Appropriate multidimensional scaling programs like PARAFAC (Harshman, 1970), INDSCAL (Carroll

and Chang, 1970), and ALSCAL (Takane et al., 1977) represent the stimuli and/or the responses as points in a multidimensional space with fixed axes, whereas the conditions are represented as factors weighing the overall configuration. Differences between these programs have to do with the structure of the input data the optimization algorithm, freedom of axes etc.

Kroonenberg and De Leeuw (1980) recently presented the method of principal component analysis by alternating least squares algorithms to solve the three-mode model in its most general form. Results from this program, called TUCKALS, will be presented here and we will limit ourselves to the initial-consonants data. A mean-squared loss function is used to minimize the difference between the low-dimensional model and the original data. A 3-dimensional stimulus configuration is presented in Fig. 1. The response configuration happens to be almost identical to this one. If one studies the configuration of Fig. 1 one will realize that clusters of consonants are positioned in the centre and at the four corner points of a tetrahedron. One can distinguish the following clusters: /l,r,w,j,h/, /z,s/, /v,f,χ/, /m,n/, and /p,t,k/ plus perhaps /b,d/. Although quite different from, for instance, the

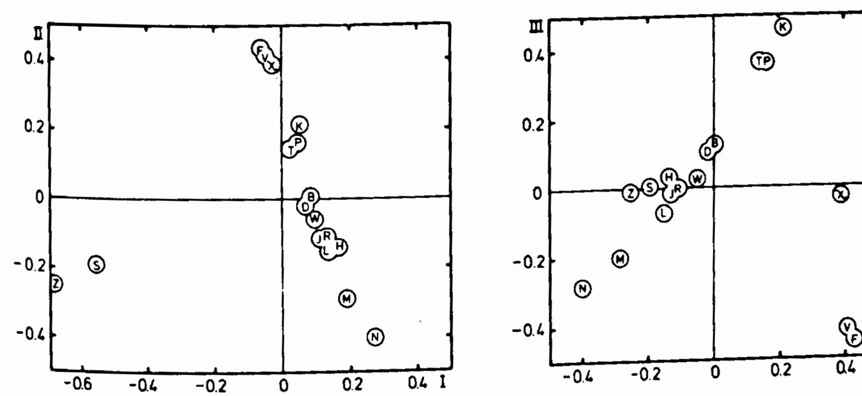

Figure 1. Stimulus configuration of initial consonants in the I-II and II-III plane of TUCKALS2 solution in three dimensions, based on double-centered confusion matrices for 28 different conditions.

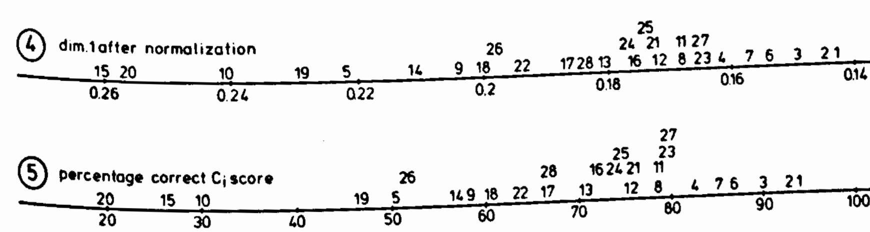

Figure 2. Representation of the first, and most important, dimension for the 28 conditions. The lower panel gives the percentage correct scores for the initial consonants for the same 28 conditions.

configuration found by Soli and Arabie (1979) by using INDSCAL on the Miller and Nicely data, this configuration seems to be quite attractive for Dutch consonants and certainly reflects the major structure in the confusion matrices. In those matrices one sees, for instance, almost no confusions between /f/ and /s/, or between /v/ and /z/, and more confusions between /p/, /t/, and /k/ than between /p/ and /b/, or between /t/ and /d/. This of course, is partly related to the types of disturbances used.

Despite the freedom given to the program to use three dimensions to represent the conditions, this representation of the 28 conditions turns out to be one-dimensional, see the upper panel of Fig. 2. The lower panel in this figure represents the percentage correct score for the initial consonants. Apart from a few deviations, e.g. conditions 26, 10, 15 and 20, there is a striking similarity. This could be an indication that both (speech) noise and reverberation have similar effects on consonant intelligibility and confusability behavior.

References

Carroll, J.D. and Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition. *Psychometrika* 35, 283-320.

Harshman, R.A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multimode factor analysis. *UCLA Working Papers in Phonetics No. 16*.

Houtgast T., Steeneken, H.J.M., and Plomp, R. (1980). Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics. *Acustica* 46, 60-72.

Kroonenberg, P.M. and De Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* 45, 69-97.

Miller, G.A. and Nicely, P.E. (1955). An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am.* 27, 338-352.

Pols, L.C.W. (1981*a*). Effect of reverberation and ambient noise on consonant intelligibility. *Report IZF 1981 C-5* (in Dutch).

Pols, L.C.W. (1981*b*). Consonant intelligibility in reverberant and/or ambient noise conditions. *Proc. 4th FASE Symposium on Acoustics and Speech, Venice, Italy*, 87-90

Shepard, R.N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika* 39, 373-421.

Soli, S.D. and Arabie, P. (1979). Auditory versus phonetic accounts of observed confusions between consonant phonemes. *J. Acoust. Soc. Am.* 66, 46-59.

Takane, Y., Young, F.W., and De Leeuw, J. (1977). Non-metric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika* 42, 7-67.

Wish, M. and Carroll, J.D. (1974). Applications of INDSCAL to studies of human perception and judgements. In: *Handbook of Perception, Vol. II*, Carterette, E.C. and Friedman, M.P. (eds.). New York: Academic Press, 449-491.