N-DIMENSIONAL METRICAL FORMALISM OF THE PERCEPTUAL SPACE OF POLISH PHONEMES

Wojciech Myślecki, Wojciech Majewski

Institute of Telecomunication and Acoustics
Technical University of Wrocław, Poland

ABSTRACT

This paper presents a method of metrical scaling of nonmetric perceptual space of Polish phonemes perceived by listeners in a presence of additive disturbances and frequency distortions. The experimental material consisted of 10 confusion matrices of Polish phonemes obtained by means of subjective tests for 10 various listening conditions. It was assumed that the confusion matrices estimate the subjective proximity between the phonems. The Shepard's algorithm of N-dimensional analysis of proximity [1,2] was used to establish a space arrangement of investigated phonemes.

INTRODUCTION

The over-all effects of additive disturbances and frequency distortions upon the average intelligibility of human speech are by now well-understood. Most of the existing studies present the results in terms of the articulation score, i.e. the percentage of the spoken words, logatoms, syllables or phonems that the listeners hear correctly. As a consequence, therefore, all of the listener's error are treated as equivalent and no information of the perceptual confusion is available. Perhaps the major reason that confusion data are not a popular subject of investigations is the time and cost of collecting them. An example of the investigations, where phoneme confusion data were used, is the study of Myślecki and Majewski [3] related to the mean conditional entropy of transmission channel (CETC). The base to calculate CETC was a phoneme confusion matrix which, in this case, constituted an experimental estimator of channel matrix for given transmission condition.

The fact that the authors were in possession of the phoneme confusion matrices for various transmission conditions are urged them to a closer look at the problem of mutual configuration of Polish phonemes in the listener's perceptual space.

It was decided to applicate the Shepard's multidimensional scaling technique [1,2] as to obtain a metrical formalism of the nonmetric perceptual space of phonemes. In this technique it is necessary to determine:

- experimental similiarities of investigated objects,
- a strategy of conducting the iterative process,
- optimal values of the constant multipliers for vectors for the approach to monotonicity and to minimum dimensionality,
- number of the iterations before a rotation to principal axes and for terminating the iterative process.

After this it is possible to estabilish a final resolution consisted of:

1° the minimum number of dimensions of the Euclidean space required such that the distances in this space are monotonically related to the initially given proximity measures

2° an actual set of orthogonal coordinates for the points in this minimum space.

The aim of this study is to determine the above mentioned values and solutions for 36 Polish phonemes.

EXPERIMENTAL PROCEDURE

Subjective measurements were carried out for 10 different conditions of speech transmission obtained by means of specially designed model of transmission channel. Masking noise and another additive disturbances of different levels (white noise, overhearing, hum) and frequency distortions (band limiting) were introduced to change a quality of speech transmission.

The test material consisted of phonetically and structurally balanced logatom lists (one or two syllable nonsense word lists) that were read by professional male speaker. For each condition of transmission, i.e. for each measuring point, four lists of 100 logatoms (1520 phonemes) were read. The listening tests were carried out in a listening studio by means of SN-60 *Tonsil* headphones.

Table 1. The transmission conditions and phoneme intelligibility scores for 10 measuring points.

№	I_{ph} %	S/N_n* dB_A	S/N_d* dB_A	Type of disturb.	Band limitation
1	73	−3	X	X	200÷4000Hz
2	75	−3	X	X	600÷2000Hz
3	86	+3	X	X	600÷2000Hz
4	87	+6	+6	hum	400÷2500Hz
5	89	+6	+6	overhear.	400÷2500Hz
6	90	+12	+6	hum	400÷2500Hz
7	91	+6	X	X	200÷4000Hz
8	92	+12	+6	overhear.	400÷2500Hz
9	95	+12	+18	overhear.	400÷2500Hz
10	99	+30	X	X	400÷2500Hz

I_{ph} phoneme intelligibility
S/N_n signal to white noise ratio
S/N_d signal to additive disturbance ratio
* S/N_n and S/N_d were measured independently before band limitation.
X without a disturbance

The listening team consisted of 12 subjects (age: 20÷24) of normal hearing. The measuring procedure was based on ISO recommendation (DP 4870, 1976). As the results of this experimental procedure applied to each of 10 measuring points a phoneme intelligibility and confusion matrix were obtained (data for all listeners and 4 logatom lists have been pooled). Table 1 summarizes the articulation data obtained for all of 10 measuring points. The S/N ratios and band limiting conditions are there also given.

METHOD OF COMPUTATION

Generally, the problem of multidimensional scaling is to find N points whose interpoint distances match in some sense (here, in the rank order sense) the experimental proximity measure (here, the confusions between 36 Polish phonemes). In this study all the computations were carried out in accordance with a program presented in [1,2], where a computing time conserving strategy was adapted. In the choosen strategy the iterative procedures start with larger values of a constant multiplier α for vectors for the approach to monotonicity

$$\alpha_{ijk}=\frac{\alpha[s_{ij}-s(d_{ij})](x_{jk}-x_{ik})}{d_{ij}} \qquad (1)$$

where

$$d_{ij} = [\sum_{k=1}^{N-1}(x_{ik} - x_{jk})^2]^{1/2} \qquad (2)$$

x_{ik} = coordinate for vertex *i* on axis *k*
s_{ij} = rank of the proximity measure
$s(d_{ij})$= rank of the distance (in N-dimensional space) corresponding to s_{ij}

and a constant multiplier β for vectors for the approach to minimum dimensionality

$$\beta_{ijk}=\frac{\beta[s_{ij}-\bar{s}](x_{jk}-x_{ik})}{d_{ij}} \qquad (3)$$

where
$\bar{s}$=the mean of N(N-1)/2 proximity measures.

Larger values of α and β promote faster though less accurate convergence between the configuration in the perceptual space and its N-dimensional Euclidean formalism. Next, when the criterion for terminating (departure from monotonicity) the iterative process

$$\delta=\frac{2\sum_{j=2}^{N}\sum_{i=1}^{j-1}[s_{ij}-s(d_{ij})]^2}{N(N-1)} \qquad (4)$$

attains its minimum, a rotation to principal axes is performed. The results of the rotation could not be taken as the final solution, but they can however be used to estimate the number of dimensions that can be eliminated. The final solution can then be reached (in a "new" reduced space) by iterating with a small value of α and β set to 0.

CALCULATIONS AND RESULTS

At a start of calculations we have to determine the values of α and β constant multipliers. Shepard [1] has undertaken a systematic exploration of the effects of these two parameters upon a convergence. He used however a set of artificial proximity measures, i.e. a known distance function and a known configuration, so it was not obvious, that his results could be directly adapted to phoneme perceptual space. For $\alpha=0.4$ and $\beta=0.2$, recommended by Shepard, we calculated the $\delta(n)$ criterion (4) for matrix № 1 (worst transmission conditions) and for matrix № 10 (best conditions). The obtained curves $\delta(n)$, n-number of iterations, are quite similiar for matrices 1' and 10' and pass through a minimum for n=3 and then increase again (see Fig. 1).The comparison of the curves from Fig. 1. with the Shepard's appropriate curve [1] shows their similiarity, hence we can conclude that the investigated configuration has no strong influence on the choice of α and β multipliers.

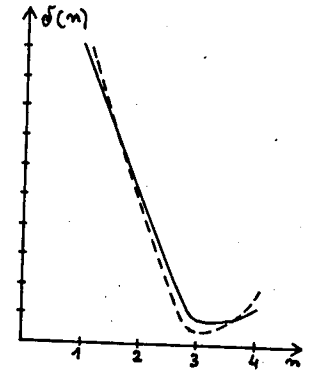

Fig. 1. Departure from monotonicity $\delta(n)$ for the worst (dashed line) and best (solid line) transmission conditions.

To the further calculations for theall of 10 confusion matrices it was decided to rotate to principal axes after the third iterations. The results obtained for 10 matrices after the rotation have shown, that the fractions of variance accounted for by the two first (in order of their importance) rotated axes were from 0.991 to 0.998.

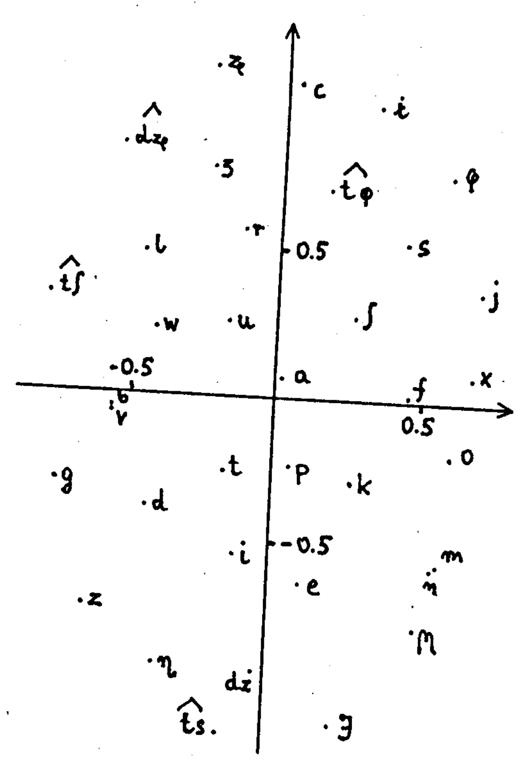

Fig. 2. Final configuration for matrix № 10 (best transmission conditions)

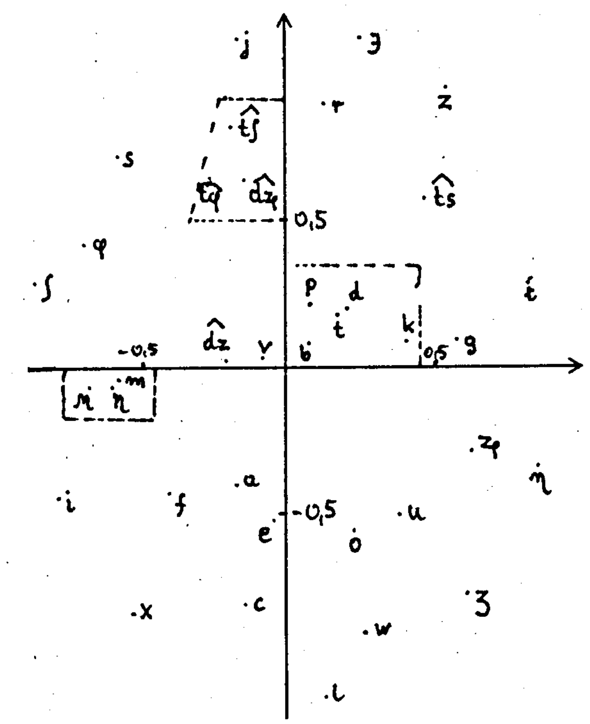

Fig. 3. Final configuration for matrix № 5 (median transmission conditions)

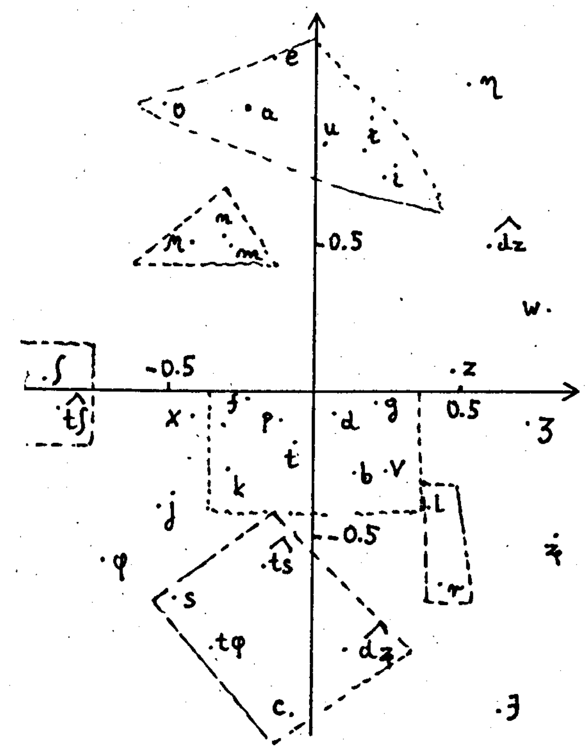

Fig. 4. Final configuration for matrix № 1 (worst transmission conditions)

From this the sufficiently close metrical formalism of perceptual space of 36 Polish phonemes might be inferred to have two dimensions. The final metrical configurations for the all of 10 investigated confusion matrices have been reached in the two-dimensional Euclidean space with the two multipliers setted so that $\alpha=0.2$ and $\beta=0$. The departure from monotonicity $\delta(n)$ attained its minimum during the iterations from 49 to 58, and its value was between $1.8 \cdot 10^{-3}$ and $2.9 \cdot 10^{-3}$. The examples of the final two-dimensional configurations of 36 Polish phonemes for the confusion matrices № 10,5 and 1 are shown in Fig. 2÷4.

SUMMARY

In this study an attention was focused on the application of Shepard's multidimensional scaling method to achieve metric formalism of Polish phoneme perceptual space. It was proved that the strategy and the parameter values recommended by Shepard [1,2] enable N-dimensional scaling of the phoneme perceptual space. The two-dimensional Euclidean space was sufficient for metric representation of arrangement of 36 Polish phonemes with an error, i.e. a departure from monotonicity, less than 0.3%.

REFERENCES

[1,2] Shepard R.N., The analysis of proximity: multidimensional scaling with unknown distance function, Psychometrika, 1962, 27, part I, pp.125-140, part II, pp.219-246

[3] Myślecki W., Majewski W., Relations between subjective and objective measures of speech transmission quality evaluation, Proc. 6-th FASE Congr., 1986, Sopron, Hungary