MEASUREMENT OF THE PERCEIVED SIMILARITY
IN FACE RECOGNITION

F. Crenna, G.B. Rossi, and L. Bovio

Measurement Laboratory, Dept. of Mechanics – DIME, University of Genova, Italy. Crenna @dimec.unige.it

Abstract: The perception of similarity among faces is investigated, to put face measurement on a sound foundation. The goal is to obtain insight in the process of human face recognition, and to apply these ideas to automatic face recognition procedures, in order to increase their efficiency and robustness, when operating in difficult measurement conditions. Two procedures are investigated, the eigenface method and the repere-points approach, that have complementary properties.

Keywords: Perceptual measurement, face recognition, multidimensional measurement, measurement with persons, eigenfaces.

1. INTRODUCTION

Face recognition for security includes identification problems, to reserve the access to critical areas to authorized people, and recognition problems, to identify unknown subjects [1]. In this paper we focus on the latter case, where so the subject is unknown and the task is to verify if he/she is included in a given data base of subjects. This case is most common in security applications, and includes the identification of suspect persons in crowded spaces.

Among all the available biometric methods for automatic recognition, face recognition is rather diffused in security applications, since it offers the chance to identify the subject without his/her collaboration [2]. This is a big advantage since it enables applications based either on existing video surveillance systems, or on specific face recording systems properly placed, for example at a ticket counter.

Face recognition may be considered as a classification on the basis of a face measurement result. The measurement process is applied to both a reference set of face images, constituting the set of known subjects, and a test face image recorded in the field. By comparing the measurement results for the reference and the test faces, the recognition algorithm returns an identification decision.

In this paper we will focus specifically on the measurement process, more than the decisional algorithm. A face, as a measurand, may be considered as a surface in the three dimensional physical space, including an infinite number of points. Available measurement methods may be divided in two classes: detail and general methods. In the next paragraph we will briefly introduce the two measurement techniques for face characterisation with particular attention to the influence of measurement

conditions. In fact since the recognition decision is based on the comparison of measurement values obtained under reference conditions (subject data base) and in the field (subject under test) and measurement uncertainty results in ambiguity or wrong recognition decisions [3-5]. Form this point of view human recognition is by far more robust then automatic recognition systems, so we are interested in comparing some features of human recognition referring to particular measurement properties used in automatic systems. If the findings of such an approach will be significant and useful we propose to introduce them in the automatic algorithm to improve its robustness, in particular with respect to the influence quantities.

In the following we will present the test methods and procedures designed to investigate face similarity from the human point of view, then we will introduce some preliminary experimental results.

The proposed methodological approach is very general and can be applied to different face measurement methods and its results implemented in different decisional algorithms.

2. FACE RECOGNITION

A general scheme of an automatic face recognition system is presented in figure 1. The algorithm devoted to the recognition decision is the last stage and it requires two distinct measurement inputs: one regards the face of subject to be identified, generally recorded in the field, may be in difficult conditions such as face orientation, lighting conditions or image resolution; the other is a set of measurements obtained on a face image data base, possibly recorded under controlled conditions (i.e frontal images, diffused illumination and similar conditions).

Measurement uncertainty is the fundamental cause for imprecise decisions, and it implies a decisional risk [12].

The measurement process is designed according to a specific recognition method as presented in paragraph 3, and it may require a preliminary phase to obtain a structured set of parameters characteristics of the face under investigation , by processing the raw measurement data.

The decisional algorithm implements a classification according to a specific method. Such classification can be schematised considering the distance in a 'face space', defined by the measurement method used, between several points representing faces in the data base and a single point representing the face under test. The data base face nearest

to the face under test is the one recognised. [12]. Of course current methods require more complex procedures, but this simplified procedure depicts clearly the problem we are going to investigate. In fact our interest in this case is to investigate if according to the human perception the face space presents a Euclidean behaviour, or if distances are somehow distorted.
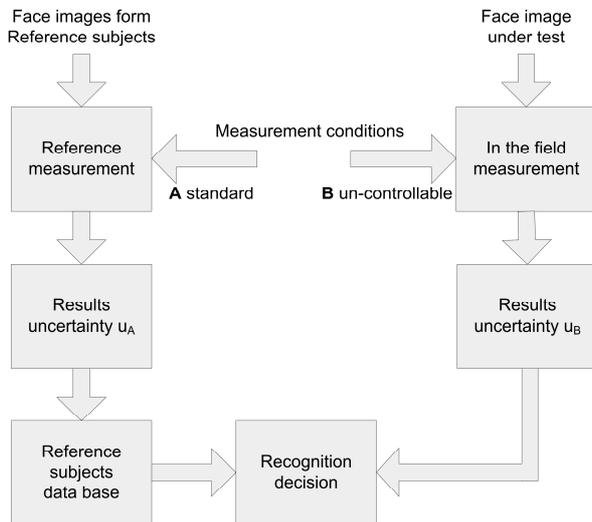

Figure 1. Measurement in recognition.

3. FACE MEASUREMENT METHODS

Available face measurement methods may be divided in two classes: detail and general methods. They are based either

- on the measurement of the position of a set of specific, *repere*, points, such as the mouth or the eye corners [1],

-on the measurement of a set of characteristics obtaining a picture of the overall appearance of the face [1-6].

Of course working conditions are critical for both approaches, as they affect face measurement results [4,5]. As an example consider different face expressions, the presence or absence of glasses, or the orientation of the face. From a measurement point of view, these are influence quantities affecting the measurement results by increasing the measurement uncertainty that will consequently decrease the reliability of the recognition decision [7].

As already presented in previous studies [7-9, 12] the effect of influence quantities affects the measurement results and the recognition efficiency. In a previous paper we proposed a probabilistic approach in order to deal with all the amount of information available and to give the user the possibility to evaluate the reliability of the recognition decision. In this research we are going to investigate human perception to be able to introduce in the automatic algorithm some significant aspect of perceptual face recognition, so that the algorithm should become more robust in particular with respect to the influence quantities.

We can consider human recognition as more similar to a general method, since it takes into account specific details of the face only when the recognition task is particularly difficult. Beside that humans are able to recognise in a reliable way also when operating conditions are not favourable, for example when the orientation of the face is considerable or a slight camouflage is used. For this reason it is particularly interesting to investigate the perceived similarity, so that some peculiar aspects of perception can be introduced in the automatic algorithms.

The methodological approach is very general and can be applied to different face measurement methods. We present here two different experiments to evaluate perceptual face similarity for a general and a detail recognition method. For the former experiment we have considered the *eigenfaces* recognition method [10-11]. Eigenfaces enables a general measurement of the characteristics of a face image, by reducing the dimensionality of the problem through the principal components analysis (PCA). First of all PCA identifies the principal directions in the face space along which the differences among faces are maximised. These directions are identified by eigenvectors of the covariance matrix that in this specific case are just called eigenfaces.

By selecting only the most important eigenfaces it is possible to reduce the dimensionality of the problem, limiting the description of the differences to a large fraction of the overall variance among faces. Our test case consists of a data base 640 x 480 face images of 28 subjects and we have selected the first 12 eigenfaces to describe about the 80% of the variance.

Figure 2 presents some of the first eigenvectors and it is clear how they resemble the image of a face. More detail about this and other similar general methods can be found in [3,6,10,11,13,14].

Figure 2: the first most important eigenfaces
as selected by PCA.

Once defined the eigenfaces, each face image is decomposed on such a base obtaining a measurement of the face along each direction. So each face is characterised by a set, or a vector, of projections of the face itself along each eigenface. Of course this is possible for both reference face images and the face under test, so the recognition algorithm will have at its disposal a set of positions in the face space, that in our test case consists of twelve dimensions.

A basic recognition algorithm considers the Euclidean point to point distance between test and reference face images; the reference face nearest to the test face will be assigned as the recognised one. The introduction of probability distributions accounting for the measurement uncertainty enables the implementation of probabilistic methods [8].

We plan to introduce some perception aspects in the algorithm by modifying the metrics used to evaluate the

distance among faces, in such a way that it won't be Euclidean anymore. Some previous studies focused on psychological and cognition aspects, confirm this approach [15].

Similarly, in the case of a detail method, the positions of some specific, *repere*, points on the face surface are considered. The number of points is variable and it depends on the specific method considered but in general they may be several tenths. Positions are measured on face images recorded with two dimensional or three dimensional methods, and the vector of positions , or coordinates of each repere point, is the basic input for the decisional algorithm. In general it may be pre-processed in order to take care of recording perspective or face orientation, but in any case such points are the inputs to the recognition algorithm.

In this case we are interested in understanding if, from the perceptual point of view, some of these repere points present more importance than others, in order to construct a sort of priority scale of repere points that could help in focusing the attention of the automatic algorithm to specific points of the images or to specific images in which face orientations are most favourable.

3. PERCEPTION TESTS

In order to investigate the perceived similarity we have designed and set up a jury test.

Such an approach requires careful design in order to obtain robust and metrologically valid results [22]. Possible approaches in order of quantity of obtainable information are:
- construction of a simple perceived similarity order
- construction of an order of the differences in similarity
- construction of a ratio of perceived similarity given a reference standard.

Besides the underling measurement scale, test method can be classified according to the stimuli available to the user. In this case the stimuli will be face images, and we can produce a test considering the evaluation of the similarity among fixed face images (*fixed stimuli*), or we can give the user the possibility to alter the face images to obtain a specific result (*variable stimuli* approach). Another important design parameter is the number of stimuli available at a time. We can consider a minimum of a couple of stimuli, for example when dealing with a perceptual evaluation in direct comparison with a reference, or all the available stimuli, as when dealing with a direct ranking of perceived similarity among a set of face images.

Different choice produce different tests set up and different situations for the user evaluation so also results may be different.

According to our previous experience [16-18] we have designed three independent tests, based on variable and fixed stimuli and on interval or ratio measurement. After proper processing a comparison of the results can validate a difference or a ratio scale according to the compatibility of the results [17], this gives the results a considerable robustness since they are obtained independently of the test method.

4. TEST PROCEDURES

Since the eigenface method we are considering is based on a mathematical procedure for decomposition and reconstruction of face images, we have been able to design a variable stimuli test interface, as shown in figure 3. After pressing the start button the interface presents to the user two images: a real image of a face and a reconstructed image based on the selected eigenfaces. The user can alter the reconstruction by moving the available sliders, until he/she has obtained the best perceived similarity with the real image. When giving the test instructions we take care to advice the user that some slider may have a non-perceivable effect so they may be neglected.

When the user is satisfied with the obtained similarity he presses the next button and another face image form the set of six, is presented. So from each subject we obtain the levels for six eigenfaces for 6 different images. This enable the identification of effects specific for an eigenface or for a particular face.

During test design we had care to set the default value of the sliders not in the position of the best mathematical reconstruction but we have applied some offset so that the user has to move the controls even to obtain the standard mathematical result. Another trick regards the real face image: since the eigenface reconstruction generally is not a clear, well defined, face image, we have defocused the original face image in order to increase the confidence of the user to obtain a similarity, on the contrary the real and the reconstructed images would be far different in any case.
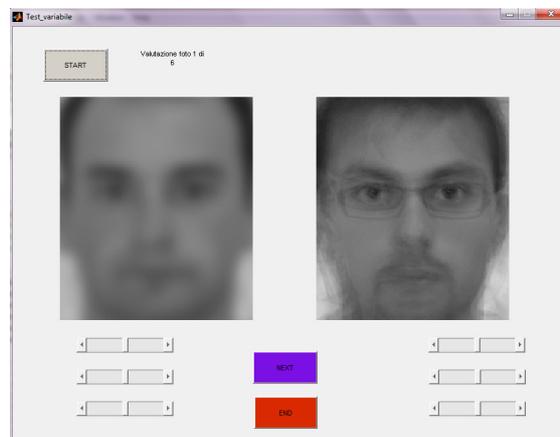

Figure 3: test interface for variable stimuli

In the second test procedure the user has to order some face images according to the perceived similarity with a reference image, by moving each image on a similarity axis from the least to the most similar, taking care of the differences in perceived similarity. Such a test is designed to obtain results on an interval scale. An example of the test interface for the detail method is shown in figure 4. The similarity axis corresponds to the red arrow, and the user can move face images only in the horizontal direction while the vertical order remains unchanged. On the pressure of the *next* or *end* buttons the software code reads the position of the face images along the *x*, or horizontal axis. Further

processing assigns a zero score to the position corresponding to the minimum similarity and a hundred to the maximum one.

When dealing with eigenfaces we have considered a set of 5 different reconstruction; one of them corresponds to the face reconstruction perceived as most similar in the previous test modality.

When detail, or repere method, is considered it is not possible to produce an alteration of the face according to a user input, due to the heavy graphical computation needed. So we have artificially modified the face images, by altering the position of some repere points, working with a three dimensional model of the face.

The graphical interface used to produce the 3D face models, gives us the possibility to save a face image in specific viewing conditions. So it is possible to simulate the view we could obtain with a given camera, with known optical and sensor characteristics, at a given viewing angle. In such a way simulated face images will be treated as images of real faces by the automatic recognition code, so that it will be possible to compare the effect of repere points modification on both perceptual and automatic recognition processes.
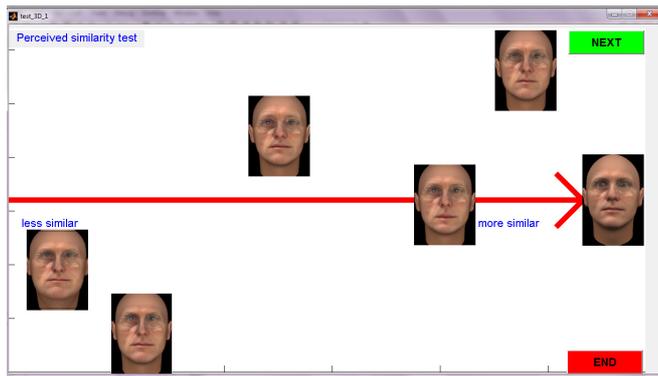

Figure 4: test interface for fixed stimuli on an interval scale

Finally a third test method requires the numerical evaluation of the perceived similarity of two face images, one of which is always the same reference, while the other varies on the same set of five face images used in the previous test. The test interface is presented in figure 5. The user has the possibility to perceive a fixed reference image, or anchor, presenting a standard similarity, corresponding to a given value, for example 100% (upper right), with respect to the real image of the subject's face (left). Then he/she has to evaluate the perceived similarity of a different face image (bottom right), by giving a numerical evaluation according to the standard. This test, called magnitude estimation, aims to obtain results on a ratio scale for the perceived similarity of faces.

Results validation will be possible comparing results obtained with the interval and magnitude tests, using the same face image set, or results obtained with the variable stimuli and the interval tests which have at least one face images in common.
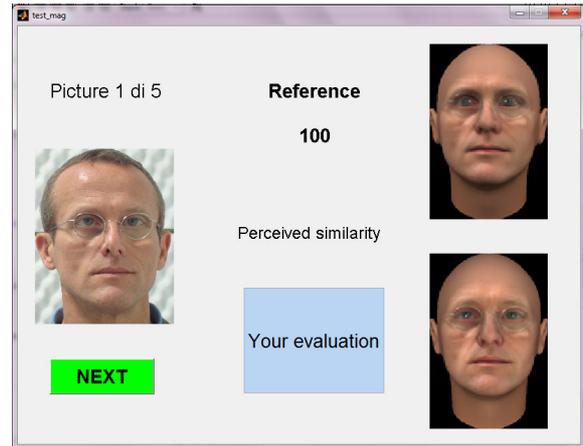

Figure 5: test interface for magnitude estimation of perceived similarity.

5. PRELIMINARY RESULTS

In the following we present some preliminary results obtained for the eigenface method with the variable stimuli test set up.

(a)

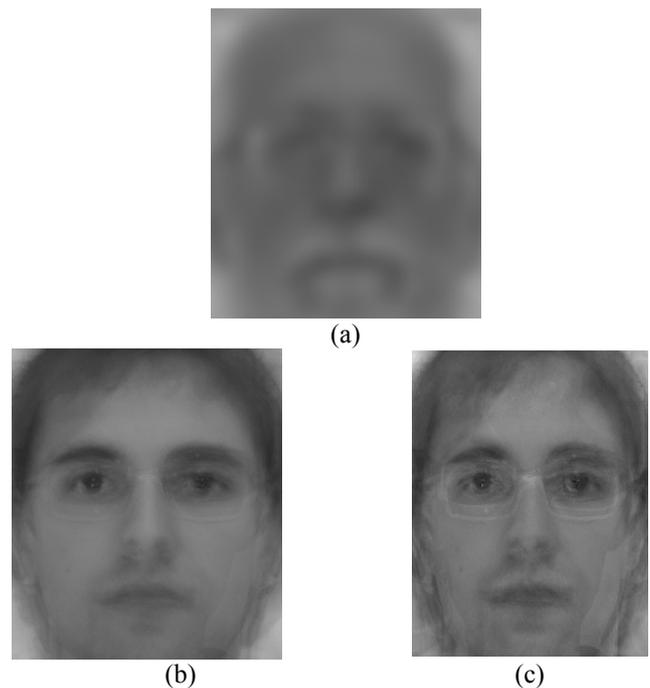(b)   (c)

Figure 6: reference image (a), mathematical reconstruction as presented to the user (b) and the one modified to be perceived as more similar (c)

We have tested the eigenface decomposition for 6 different face images: 5 male, 1 female, 1 with bear, 2 with glasses. The user had the possibility to alter up to ±3 times the projection of the face image along the direction of 6 eigenfaces, as to say that the position of the face in the eigenface space can be varied up to three times. The 6 eigenface we selected considering the first 3 most important

eigenfaces, corresponding to the three largest eigenvalues and three least important, among the 12 eigenfaces selected.

An example of maximum similarity matching obtained by one subject is given in figure 6. First on the top the original face image (6a), bottom left the mathematical best reconstruction (6b), bottom right the user generated reconstruction.

The first experimental data obtained from a rather small jury has been analysed by 3 way analysis of variance.

Results show:
-   a significant contribution of the eigenface number, with a probability level of about 2%, less than the usual 5% limit, confirming that there is an effect on perception connected to the eigenface;
-   a non-significant effect of subjects constituting the jury, confirming that the previous sensitivity to the eigenface number is not typical of a specific subject but it is general, as the jury elements are non-significant;
-   a non-significant effect of the specific face image, confirming that the perceptual effect of the eigenface number is independent on the face image considered but it is a general effect.
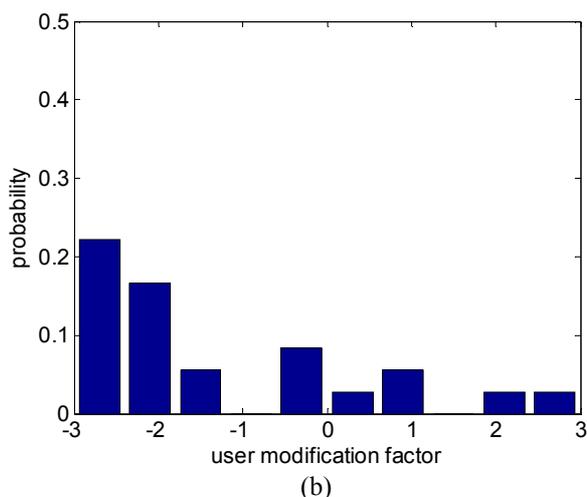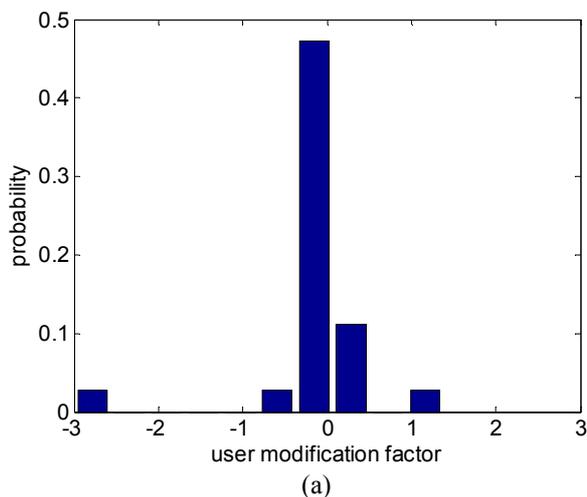

(a)

(b)

Figure 7: probability distribution of results for
an important eigenface (a)
and for one with much less contribution to variance (b)

6. CONCLUSIONS

Although at an initial investigation stage, results confirm that the projections along the eigenvectors, or eigenfaces, corresponding to the lower eigenvalues, do have an noteworthy impact on similarity perception. This means that although such directions account for a minor amount of the global variability, as compared to the others, still they are very important from the perceptual point of view. One possible conclusion can be that in considering the distances along such directions with the same weight than the other we are neglecting their perceptual importance. If this is the case, a different weighting curve should be introduced. On the other hand, it is not easy to identify a systematic contribution, from these preliminary data. As an example, Figure 7 shows how different the probability distributions for two eigenfaces can be: the former refers to an eigenface that accounts for a large amount of global variance; the latter one that provides a limited contribution. A larger effect for the latter is evident but, with the data available at the moment, it is difficult to figure out a possible set of weights the enables the construction of a new metrics.

REFERENCES

[1]   Stan Z. Li, Anil K. Jain eds 'Handbook of Face Recognition', Eds., Springer-Verlag, 2004.
[2]   Zhao, W.Y., Chellappa, R., Rosenfeld, A., Phillips, P.J., 'Face recognition: A literature survey', ACM Comput. Surv.35 (4), 399–458, 2003.
[3]   G. Shakhnarovich, B. Moghaddam, 'Face Recognition in Subspaces', in 'Handbook of Face Recognition', Eds. Stan Z. Li and Anil K. Jain, Springer-Verlag, 2004.
[4]   W Zhao, R Chellappa, 'Face processing, advanced modeling and methods', Academic Press, Elsevier 2006
[5]   J. Zhang, Y. Yan, M. Lades, Face Recognition: Eigenface, Elastic Matching and Neural Nets, Proc. IEEE, 85, 9, 1997.
[6]   J. Ruiz del Solar, P. Navarrete, 'Eigenspace-based face recognition: a comparative study of different approaches', IEEE Trans. on Systems, Man and Cybernetics, Part C, 35, 3, 2005, 315-325.
[7]   F. Crenna, G B. Rossi, L Bovio, 'Effects of measurement conditions on face recognition: An experimental and numerical investigation', IMEKO TC21 workshop MSPI 2010, NPL, Teddington, UK.
[8]   F. Crenna, G. B. Rossi, L. Bovio, 'Probabilistic characterization of face measurement', Proc. IX Advanced mathematical and Computational Tools in Metrology, AMCTM 2011, Gothenburg, Sweden 19-21 June 2011
[9]   Rossi G. B., Crenna F. 'A probabilistic approach to measurement based decisions', Measurement 39 (2006) 101-119
[10] M. Turk , A. Pentland, Face recognition using eigenfaces, in Proceedings of International Conference on Pattern Recognition, pp.586-591, 1991.
[11] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience,.3 71-86, March 1991.
[12] G Betta, D Capriglione et al, 'Face-based recognition proposals for the metrological characterization of feature-based approaches' , Meas. Sci. Technol. 22 (2011) 124005 (10pp)
[13] Wendy S.Yambor, Analyzing PCA-based Face Recognition Algorithms: Eigenvector Selection and Distance Measures,

Computer Science Department Colorado State University, U.S.A , 2000.

[14] W.S. Yambor, "Analysis of PCA-based and Fisher Discriminant-Based Image Recognition Algorithms", M.S. Thesis, Technical Report CS-00-103, Computer Science Department, Colorado State University, July 2000.

[15] Wenger, M. J., & Townsend (2001). Faces as gestalt stimuli: Process characteristics. in Face perception and cognition, edited by M. J. Wenger and J. T. Townsend. Erlbaum Press.

[16] Rossi G. B., Crenna F., Panero M. Panel or jury testing methods in a metrological perspective, Metrologia, 42 (2005) 97-109

[17] F. Crenna, G. B. Rossi, L. Bovio , Loudness measurement by robust magnitude estimation, to be published in proc. of the 14th IMEKO TC1-TC7-TC13 joint Symposium, "Intelligent Quality Measurement - Theory, Education and Training ", August 31-September 2, 2011, JENA, Germany

[18] G. B. Rossi, F. Crenna, Insight in ratio scales, in A. Bastianelli & G. Vidotto (Eds.) (2010) Fechner Day 2010. Proceedings of the 26th Annual Meeting of the International Society for Psychophysics, Padua, Italy: The International Society for Psychophysics, October 19-22, pp. 215-220

[19] J. Townsend, M J.Wenger, "Compritetional, Geometric and process perspectives an FACIAL Cognition" Seves 2001.

[20] Wenger, M. J., & Townsend Faces as gestalt stimuli: Process characteristics. Chapter in book on face perception and cognition, edited by M. J. Wenger and J. T. Townsend. Erlbaum Press. (2001).

[21] Townsend, J. T., Aisbett, J., Assadi, A., & Busemeyer, J. General recognition theory and methodology for dimensional independence on simple cognitive manifolds. In H. Colonius & E. N. Dzhafarov (Eds.), Measurement and Representation of Sensations: Recent Progress in Psychophysical Theory. (2006).

[22] Berglund B, Rossi G.B., Townsend J.T., Pendrill L.R., eds. "Measurement with persons", Psychology Press - Taylor and Francis, 2012