Synthesis of Facial Caricatures Using Eigenspaces and Its Applications to Humanlike Animated Agents

Masahide Kaneko Mitsuhiko Meguro

Department of Electronic Engineering, The University of Electro-Communications

1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, JAPAN

E-mail : {kaneko, meguro}@ee.uec.ac.jp

Abstract

This paper first summarizes our unique method to synthesize facial caricatures using eigenspaces. This method can handle shape features of each facial part and arrangement features of facial parts independently. Then, in order to provide a dynamic lifelike agent based on a synthesized facial caricature, a method to represent mouth shape changes corresponding to speech and a method to give facial expressions are proposed. Here, the fact that facial features are exaggerated in facial caricatures is considered to synthesize an appropriate animation of facial caricature automatically. Experimental results show the usefulness of the proposed methods.

Key Words : Facial Caricature, Eigenspace, Humanlike Animated Agent, Mouth Shape Changes, Facial Expressions

1 Introduction

Facial caricatures represent features of individual faces compactly and effectively. If one can conveniently use an animated facial caricature of the specific person whom one knows well in computer interface, it will act as a humanlike agent and will provide more familiar, richer and more pleasant interaction environment.

As for the synthesis of facial caricatures by computer, several methods have been proposed. Most of conventional approaches are based on the idea which calculates the differences between an input face and an average face, and exaggerates them by the extrapolation technique [1]. Usually shapes of facial parts and their arrangement are treated together. This may restrict the variety of representation ability and also cause the unfavorable collapse in synthesized images when the degree of exaggeration becomes large.

We have already proposed a method to synthesize a facial caricature using eigenspaces[2]. Eigenspaces are derived for shape of each facial part and arrangement of facial parts, respectively. Exaggeration process is performed through each eigenspace independently. Thus, this method possesses the high flexibility in drawing facial caricatures.

In order to utilize a facial caricature as a humanlike agent, it is required to give motions to it. Especially, it is important to give mouth shape changes corresponding to speech and facial expressions. Synthesis of mouth shape and facial expressions have been studied in the fields of the model-based image coding [3]-[5] and computer animation[6]. In most of these studies, a geometric model (often 3-D) for real head is employed and few studies have been done for the synthesis of animated facial caricature. The major difference between facial caricatures and real facial images is that the facial features (both shapes and arrangement of facial parts) are exaggerated in facial caricatures. Therefore, this paper discusses a method to give proper mouth shape changes and expressions to a facial caricature under considering exaggerated features.

Firstly, the basic idea to synthesize a facial caricature using eigenspaces is introduced. Then, the automatic synthesis of mouth shape changes and dynamic facial expressions are discussed. Experimental results show that the proposed method can provide animations for different feature exaggerated facial caricatures using the same control method.

2 Preprocessing

A three-dimensional (3-D) wireframe model of head consisted with a number of triangular patches is used to handle facial features by computer. A standard wireframe model of head is fitted to an input facial image. Then this model is subjected to the normalizing operation which corrects the displacement of face position and the difference of scaling factor.

An average wireframe model is obtained by averaging the coordinate values of wireframe models collected from facial images of fifty students. Although a wireframe model contains 3-D information, only 2-D information is utilized

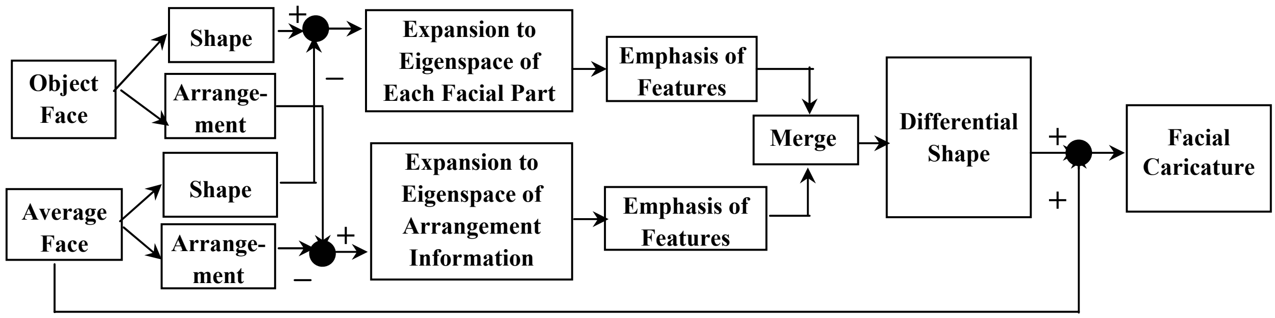

Fig.1 Framework of synthesis of facial caricature using eigenspaces.

at this moment in this paper. Position and contour shape of each facial part is picked up from the 3-D wireframe model of head. As for facial parts, eyebrows, eyes, nose, mouth, and contour of frontal face are considered here.

3 Synthesis of Facial Caricatures using Eigenspaces

Figure 1 shows the framework of synthesis of facial caricatures using eigenspaces[2]. In this method shape of each of facial parts and their arrangement are treated independently. After the emphasizing process for shape features and arrangement feature have been done, their results are merged to yield final facial caricature.

3.1 Processing for Shape of Each Facial Part

Step 1 : Generation of eigenspace for contour shape

A differential shape vector is calculated by subtracting the shape vector given by an average wireframe model of head from that given by wireframe model fitted to each of input facial images. After the variance-covariance matrix is calculated using a number of differential shape vectors, eigenvalues and eigenvectors are derived from this matrix by applying the principal component analysis.

Step 2 : Expansion to the orthogonal bases

A differential shape vector obtained by subtracting the shape vector given by an average wireframe model from that given by wireframe model of an object face is expanded to the orthogonal bases which have been obtained in Step 1. An expansion coefficient r_j for the jth eigenvector is given by the following equation (1).

$$r_j = d \times u_j / \|u_j\| \qquad (1)$$

where d : differential shape vector, u_j : jth eigenvector and $\| \ \|$: norm.

Step 3 : Enhancement of expansion coefficients

A ratio of score is calculated for each expansion coefficient. A magnification factor k for each expansion coefficient is determined based on a ratio of score. For an eigenvector with a larger ratio of score, a larger magnification factor is assigned. For an eigenvector with a small ratio of score, a magnification factor will be set as zero.

Using the eigenvectors multiplied by magnification factors, the exaggerated shape vector P will be obtained by the following equation (2).

$$P = \sum_{j=1}^{m} (k_j \times r_j \times \frac{u_j}{\|u_j\|}) + \overline{P} \qquad (2)$$

where m : number of eigenvectors (here, m = 15), k_j : magnification factor for jth eigenvector and $\overline{P}$: an average shape vector.

3.2 Processing for Arrangement of Facial Parts

Arrangement information of facial parts is represented by the positions of six points; a center point of each eyebrow, a center point of each eye, the bottom of nose and a center of mouth. Procedures to exaggerate the arrangement information of facial parts is basically the same with those for shape information of each facial part. The differences between arrangement information for an average face and that for an object face are processed through an eigenspace for differential arrangement information.

Since the features in arrangement of facial parts are processed independently from shape features, this method possesses the high flexibility in drawing facial caricatures.

3.3 Integration of Exaggeration in Shape Features and in Arrangement Features

By combining the result obtained through the exaggeration in shapes of facial parts and the result obtained through the exaggeration in arrangement of facial parts, a facial caricature, in which both shape features and arrangement features are exaggerated, will be synthesized. Figure 2 shows the example of obtained caricatures. In Fig. 2, (a) and (b) are an input facial image and its initial contour shape, respectively. (c) is an average face. For (d), only shape of each facial part is exaggerated and their arrangement is the same as in an input facial image. For (e), both shapes and arrangement are exaggerated. (f) is a final synthesized caricature with hair. Hair is extracted from an input image and binarized.

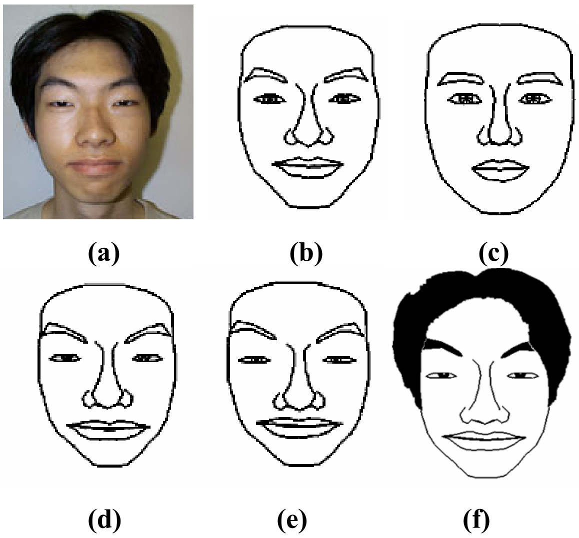

Fig.2 (a) Original image, (b) Initial contour shape, (c) Average face, (d) Only shapes are emphasized, (e) Both shapes and arrangement are emphasized, and (f) Final facial caricature with hair.

4 Lifelike Agent Using Facial Caricature

Since a 3-D wireframe model of head is used in the proposed method to represent the shapes and arrangement of facial parts, this makes it easy to represent the movement of head and further to give facial expressions and mouth shape changes in synthesized facial caricatures. This means that one can synthesize a facial caricature sequence with arbitrary mouth shape changes and expressions, which impersonates someone whom one knows well. It will provide more familiar and pleasant interaction or communication between a user and an agent (caricature) than the case using a ready-made agent of computer generated character.

References [3]-[5] have discussed the methods to give facial expressions and mouth shapes to the synthesized lifelike facial images using real face texture. Since the shapes and arrangement of facial parts are exaggerated in a facial caricature, which are different from those in an original facial image, suitable methods to control expressions and mouth shape are necessary for an exaggerated facial caricature as discussed in the next section.

5 Synthesis of Mouth Shape

5.1 Synthesis of Basic Mouth Shape

In most of conventional studies on the synthesis of mouth shape corresponding to speech, the ordinary, that is non-exaggerated, mouth shape has been considered. For facial caricatures, mouth shapes are usually exaggerated and quite different from that of real faces. For example, very large and wide mouth, small but thick lips, and so on. When the same process to control shapes of ordinary mouth is applied to an exaggerated facial caricature, unnatural facial caricatures may be drawn, for example, mutual interference among a mouth, a nose and a chin may occur. For this problem, the following condition is employed: (a) a mouth is positioned between a nose and a chin, and (b) lips do not go over a jaw.

Furthermore in order to apply the same control method over mouth shape changes to various types of exaggerated mouth shapes, the following method is employed.

(1) Many facial caricatures with the same intensity of exaggeration are collected and ordered considering the width of mouth and thickness of lips. Average and standard deviation are calculated and are used to decide a threshold to classify each of width and thickness of lips into one of three groups, that is, “large”, “medium”, and “small.”

(2) Weighting factor ω_n is determined in consideration of the classification result and the intensity of exaggeration. Modification parameter F is calculated for each of width of mouth, opening of mouth, and thickness of lips so that it magnifies the basis parameter assigned to non-exaggerated mouth shape.

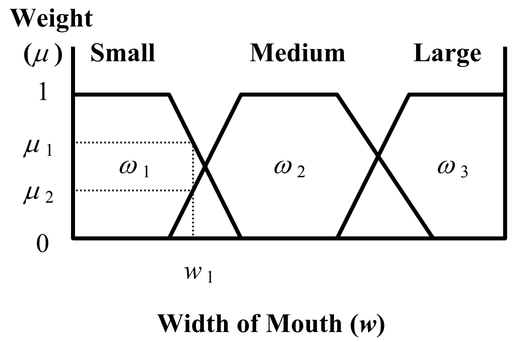

Fig.3 Classification of mouth shape (i.e. width of mouth) and weighting factors.

Figure 3 shows how to classify the width of mouth. ω_1, ω_2, and ω_3 are weighting factors for small, medium and large width of mouth, respectively. When the width of mouth in the original facial caricature is w_1, then weights μ_1, μ_2 and μ_3 are applied to each of ω_1, ω_2, and ω_3. Therefore,

$$F = \frac{\mu_1\omega_1 + \mu_2\omega_2 + \mu_3\omega_3}{\mu_1 + \mu_2 + \mu_3} \quad (3)$$

In the case of Fig.3, weights μ_1, μ_2 and μ_3 are 0.7, 0.3 and 0, respectively.

(3) Initial values of mouth width, opening and lip thickness are calculated using the width of closed mouth under relaxed condition.

(4) Basic shapes of lips are approximated by the combination of quadratic functions. In order to represent the puckered mouth well, opening of mouth and thickness of lips are decreased at left and right sides of mouth for vowels “u” and “o.”

The approximation of mouth shape by quadratic functions does not necessarily represent the individual shape features in an original caricature well. To cope with this problem, difference ΔD between the closed mouth shape in individual caricature and the mouth shape approximated by quadratic functions is calculated. This difference represents the shape feature of individual caricature. After controlling the mouth shape by the above steps (1) – (3), ΔD is added to give individual person’s shape features.

5.2 Sequence of Mouth Shape Corresponding to Speech

Referring the observation of mouth shapes for uttering Japanese vowels and consonants [6], sequence of mouth shapes corresponding to speech is generated as follows:

(1) Mouth shapes corresponding to Japanese five vowels “a”, “i”, “u”, “e”, “o” and closed mouth shape are first generated as basic mouth shapes for a specific facial caricature using a method described in 5.1.

(2) Input text is divided into phonemes and duration time for each phoneme is calculated beforehand.

(3) Sequence of mouth shape is generated by interpolating basic mouth shapes. For a consonant, a mouth shape of succeeding vowel is assigned excepting for the following cases: (a) For labial sounds p, m, and b, lips are first closed, then they are changed into the succeeding mouth shape. (b) For a doubled consonant, the mouth shape just before is kept for a while, then it is changed into the succeeding mouth shape. (c) For a syllabic nasal N, lips are closed for a while, then they are changed into the succeeding mouth shape.

5.3 Example of Synthesized Facial Caricature with Mouth Shape Change

Figures 4 and 5 show the examples of basic mouth shapes synthesized for two types of facial caricatures having different features in mouth shapes. The same control method was applied under considering the exaggeration of features in

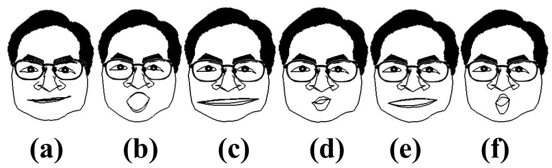

Fig.4 Basic mouth shapes : (a)Input facial caricature, (b) “a”, (c) “i”, (d) “u”, (e) “e” and (f) “o.”

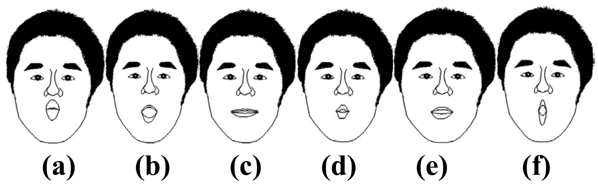

Fig.5 Basic mouth shapes : (a)Input facial caricature, (b) “a”, (c) “i”, (d) “u”, (e) “e” and (f) “o.”

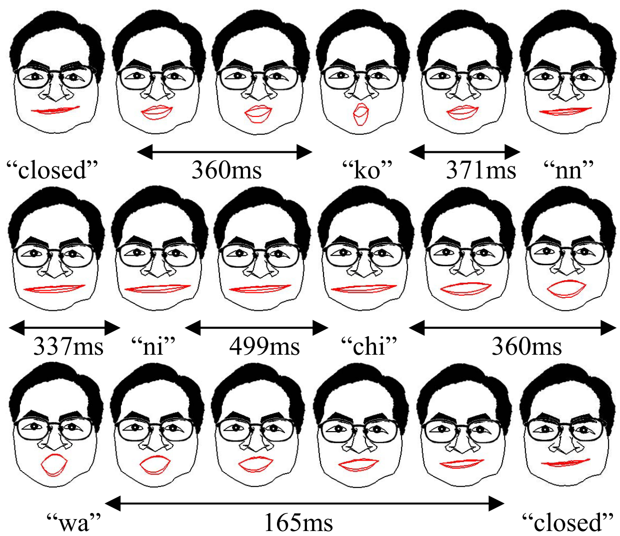

Fig.6 Sequence of facial caricatures uttering Japanese short sentence “konnichiwa.”

mouth shapes. Figure 6 shows the sequence of facial caricatures uttering Japanese short sentence “konnichiwa” for the facial caricature shown in Fig.4. At this moment, teeth and a tongue are not included. However, the proposed method can control the mouth shapes in different types of facial caricatures well as following the same manner.

6 Representation of Facial Expressions

Synthesis of facial expressions has been studied in both real facial images [4][5] and in computer animation [7]. However, few studies have been done for a facial caricature with dynamic facial expressions. The problem, which is particular to a facial caricature, is that facial features (both shapes and arrangement of facial parts) are exaggerated. That is, facial expressions should be properly represented under the condition that facial features are exaggerated and generated facial expressions should match with exaggerated facial features.

6.1 Analysis of Facial Expressions

There are two typical methods to synthesize facial expressions. One is based on Action Units used in FACS [8]. The other is based on the motion analysis of real facial image sequences. This paper employs the latter to make it possible to treat the features of facial expressions for each individual person. 59 feature points are first specified on the contours of eyebrows, eyes, nose, mouth and jaw in the first frame. Motion of each of these points is automatically tracked in consecutive frames as follows:

(1) Facial area is separated from background using color information.
(2) Each of inner corners of right and left eyes is used as a reference point. In order to avoid the effect of motions caused by the movement of the whole head, the position and rotation of face in each of consecutive frames are adjusted by using the positions of reference points. Then we can obtain the facial image sequence in which the position of facial area is fixed.
(3) Motion of each of above 59 feature points is tracked in consecutive frames using a block matching method.

6.2 Average Facial Expressions

In order to represent individual features in facial expressions, it is necessary to analyze the facial image sequence for each person. Since this requires tedious works, it is useful to prepare average motion data for each of typical facial expressions.

Average motion data for each facial expression can be obtained as follows :

Step 1 : <u>Normalization of input facial images</u>

In order to avoid the effect of difference in size and rotation in input images, each of outside corners of right and left eyes and the bottom point of nose are used to normalize input images. Normalization process is carried out by adjusting the area of triangle composed by these three points.

Step 2 : <u>Averaging</u>

For each typical expression, plural frames at the same timing along the time axis are picked up from facial image sequences for different persons; the number of persons is 18 for “happiness”, and 10 for “anger”, “fear”, and “surprise”. Then an average position of each of 59 feature points is calculated.

Step 3 : <u>Facial caricature with average expression</u>

By calculating motion of each of 59 feature points frame by frame, a sequence of average motion vectors for each of facial expressions can be obtained. Figures 7 and 8 show examples of facial caricatures with average “happiness” expression and average “surprise” expression, respectively. Figure 9 shows the sequence with average “surprise” expression applied to the other facial caricature.

6.3 Facial Expressions of Specific Person

When motion data can be obtained for a specific person, it is possible to synthesize facial caricature sequence with his/her own expressions. Since the shape and arrangement of facial parts are exaggerated in facial caricature, motion parameters are also exaggerated according to the intensities of exaggeration for shapes and arrangement, which is specified in the process to synthesize a facial caricature without facial expression.

In addition to above, motion data may be linearly extrapolated along the time axis to exaggerate facial expression. Figure 10 shows the sequence of facial caricatures with exaggerated “happiness” expression. Expression in the rightmost caricature is extrapolated. Figure 11 shows the sequence in which motion data used in Figure 10 was applied to another facial caricature.

Through Figures 7 to 11, it can be seen that the method described here can successfully synthesize several types of facial caricature sequences with appropriate facial expressions.

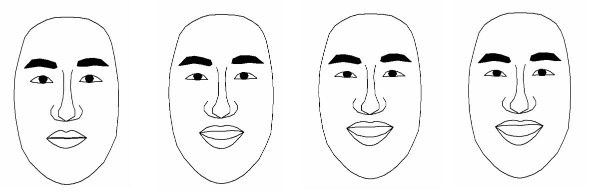

Fig.7 Average "happiness" expression. From left to right, 1st, 8th, 14th, and 21st frames.

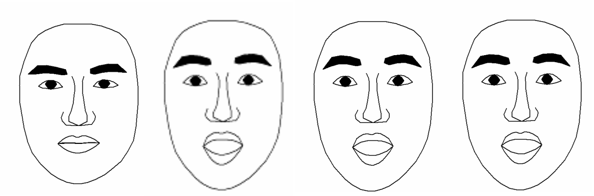

Fig.8 Average "surprise" expression. From left to right, 1st, 8th, 14th, and 21st frames.

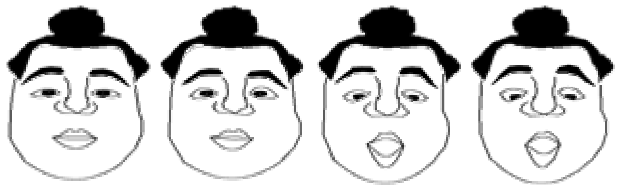

Fig.9 Sequence with average "surprise" expression applied to other facial caricature. From left to right, 1st, 8th, 14th, and 21st frames.

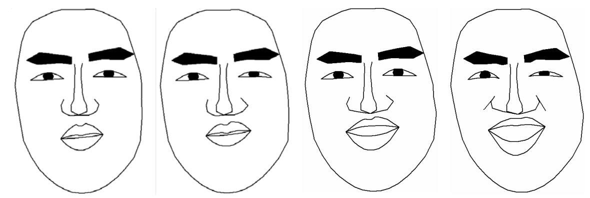

Fig.10 Sequence of facial caricatures with exaggerated "happiness" expression. From left to right, 1st, 10th, 20th, and 36th frames.

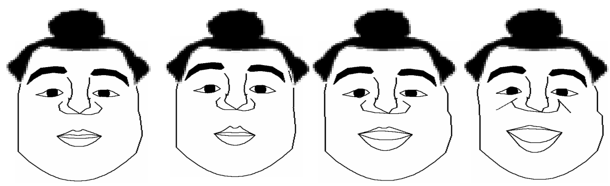

Fig.11 Sequence in which motion data used in Fig.10 is applied to other facial caricature. From left to right, 1st, 10th, 20th, and 36th frames.

7 Conclusion

This paper has first summarized our unique method to synthesize facial caricatures, which can handle shape features of each facial part and arrangement features of facial parts independently. Then, in order to provide a dynamic lifelike agent based on a synthesized facial caricature, a method to represent mouth shape changes corresponding to speech and a method to give facial expressions are proposed. Here, the fact that facial features are exaggerated in facial caricatures is considered to synthesize appropriate animation of facial caricature automatically. Experimental results show the usefulness of the proposed methods.

Combination of facial caricature and body action, especially gesture and 3-D motion of the whole head, is the further research objective to realize more pleasant and richer interaction between humans and humanlike agents.

Acknowledgement

The authors express their deep appreciation to Prof. Akira Kurematsu for his encouragement and advice to this study. They also thank Mr. GuangZhe Xu, Miss Sayaka Yamada and Miss Ling Lu for their great help to this study.

References

[1] H. Koshimizu and K. Murakami : " Facial caricaturing based on visual illusion --- a mechanism to evaluate caricature in PICASSO system," *IEICE Trans*., vol.E76-D, no.4, pp.470-478, April 1993.

[2] GuangZhe Xu, Masahide Kaneko and Akira Kurematsu : " Synthesis of facial caricature using eigenspaces, " *IEICE Trans. D-II*, vol.J84-D-II, no.7, pp.1279-1288, July 2001. (in Japanese)

[3] Masahide Kaneko, Atsushi Koike, and Yoshinori Hatori : " Synthesis of moving facial images with mouth shape controlled by text information," *IEICE Trans. D-II*, vol.J75-D-II, no.2, pp.203-215, Feb. 1992. (in Japanese)

[4] Masahide Kaneko, Atsushi Koike, and Yoshinori Hatori : " Automatic synthesis of moving facial images with expression and mouth shape controlled by text," in *Visual Computing*, ed. T. Kunii, Springer-Verlag, pp.57-75, June 1992.

[5] C.S. Choi, H. Harashima, and T. Takebe : " 3-dimensional facial model-based description and synthesis of facial expressions," *IEICE Trans. A*, vol.J73-A, no.7, pp.1270-1280, July 1990. (in Japanese)

[6] Y. Fukuda and S. Hiki : " Characteristics of the mouth shape in the production of Japanese --- Stereoscopic observation, " *J. Acoust. Soc. Japan* (*E*), vol.3, no.2, pp.75-91, 1982.

[7] Frederic I. Parke and Keith Waters : " *Computer Facial Animation*, " A K Peters, Ltd., 1996.

[8] P. Ekman and W. V. Friesen : "*Facial Action Coding System*, " Consulting Psychologist Press, 1977.