

Perceived age as a biomarker of ageing: a clinical methodology

David A. Gunn · Peter G. Murray ·
Cyrena C. Tomlin · Helle Rexbye ·
Kaare Christensen · Andrew E. Mayes

Received: 17 January 2008 / Accepted: 4 April 2008 / Published online: 22 April 2008
© Springer Science+Business Media B.V. 2008

Abstract In a previous field-based study, how old one looks for one's age (perceived age) was found to be predictive of mortality in elderly individuals. In conjunction, perceived age is of relevance and interest to the layperson. Here, a clinical methodology for generating perceived age as a biomarker of facial ageing is detailed. The methodology utilises facial photographs of subjects to present images to large numbers of age assessors who are primarily nationals of the country of study origin. In five observational studies in five different countries involving 874 female subjects it was found that subject age and assessor gender, nationality, age and ageing expertise had little effect on the perceived age data generated. However, increasing the numbers of

age assessors up to 50 substantially increased the reproducibility of the mean perceived age for an image and a minimum of 10 assessors were required to give reproducible data. This methodology was also compared to a methodology that utilises passport-type photographs of subjects typically taken in field studies. Although the perceived age data from the two types of images were more similar to each other than to chronological age, there was a marked difference between the two sets of data. Therefore, to allow meaningful comparisons across perceived age studies, the same type of image should be used for the generation of perceived age. In conclusion, the methodology detailed here has demonstrated that perceived age can be a reproducible measure when large numbers of adult age assessors are used and can be utilised globally in studies to investigate facial ageing.

Electronic supplementary material The online version of this article (doi:[10.1007/s10522-008-9141-y](https://doi.org/10.1007/s10522-008-9141-y)) contains supplementary material, which is available to authorized users.

D. A. Gunn (✉) · C. C. Tomlin · A. E. Mayes
Unilever Corporate Research, Colworth Science Park,
Sharnbrook, Bedfordshire MK44 1LQ, UK
e-mail: david.gunn@unilever.com

P. G. Murray
Unilever Statistics and Sensory Science Group,
Colworth Science Park, Sharnbrook, UK

H. Rexbye · K. Christensen
The Danish Twin Registry, Epidemiology Unit, Institute
of Public Health, University of Southern Denmark,
Odense, Denmark

Keywords Perceived age · Ageing ·
Facial ageing · Face · Photography

Introduction

Ageing results in the decline of the structural integrity and functional performance of tissue over time. Measures of these changes are representative of a tissue's biological age and such measures can be targeted to a specific tissue [e.g. cardiovascular tissue

(Bulpitt et al. 1999)] or to the whole body (Borkan and Norris 1980). The usefulness of such a measure can be judged by its ability to predict mortality or age-related disease outcomes (Borkan and Norris 1980; Ingelsson et al. 2007; Seeman et al. 2001). Once an accurate measure of biological age has been established, it can be used to investigate the environmental and genetic factors that influence it and hence, it can be used to study the aetiology of ageing.

Ageing results in dramatic changes to the appearance of individuals' faces (Krejci-Papa and Langdon 2006) and, thus, alters how old an individual looks. The changes to visual features with age are likely to reflect the biological age of the tissue from which they originate. In support of this, perceived age has been shown to be a measure of skin photo-ageing (Warren et al. 1991). Furthermore, perceived age has also been shown to be associated with predictors of mortality (Borkan and Norris 1980), to be predictive of mortality in Danish twins (Christensen et al. 2004), and is associated with lifestyle factors (e.g. smoking) known to affect the incidence of age-related diseases (Rexbye et al. 2006). Perceived age is, therefore, a measure of biological as well as chronological age.

Photographs of subjects used for perceived age studies can be captured at the subjects' home addresses (i.e. a field setting) or in a clinical setting. Non-standardised digital photographs are relatively quick and simple to capture in a field setting for large numbers of subjects. For example, two previously reported studies (Christensen et al. 2004; Rexbye et al. 2006) utilised digital passport-type photographs of nearly two thousand subjects taken in the field for age assessments. Photographs taken in a clinical setting are typically more standardised and have focused on the facial appearance of subjects; for example, photographs of the whole head (Burt and Perrett 1995) and frontal facial photographs (Warren et al. 1991) have both been used in a clinical setting. However, changing the type of image presented to assessors changes the subject features available for age assessment. Therefore, it is unclear how comparable perceived age measures are across studies.

The number of assessors used in perceived age studies has been diverse and two contrasting instances of this are Bulpitt et al. (2001) who used three assessors for age judgments and Fink et al. (2006) who used 430 assessors. Furthermore, the age, gender and ageing expertise of assessors has been

varied; for example, Yamaguchi and Oda (1999) used 'naïve' undergraduate students, Christensen et al. (2004) used older female geriatric nurses, Rikkert (1999) used four experienced geriatricians, and Burt and Perrett (1995) used a combination of old, young, male and female naïve assessors. It is unclear how such variability in the number and composition of assessors affects the reproducibility and absolute values of perceived age data. Reported here is a new methodology for the generation of perceived age which was utilised in five studies run in five different countries consisting of nearly 900 female participants. The primary aims were to determine how different numbers of assessors, their gender, nationality and age and different ways of presenting individuals in images affected the mean perceived age scores of female adult subjects.

Materials and methods

Study designs

A total of five studies were completed in England, Spain, Denmark, Canada and China (Table 1). Recruitment of the subjects was carried out to ensure there was an even spread of subjects across the chronological age range of each study. Subjects were nationals of the study of origin and of Caucasian ethnicity for the European and Canadian studies and East Asian ethnicity for the Chinese study. All individuals were photographed three times using a Fuji digital camera with their hair drawn off their face (see Supplementary Material for further details). Selected photographs consisting of both a frontal and 45 degree angle eyes closed shot were cropped around the face and neck-line then presented to age assessors via a website. In addition to these facial photographs, passport-type photographs of the Danish subjects were taken and presented to assessors as previously reported (Rexbye et al. 2006).

Evaluation of photographs

The cropped facial photographs were assessed by 'naïve' assessors who were primarily employees based at Unilever Research and Development sites. For the English and Danish studies, assessors were mainly British; for the Spanish study only Spanish

Table 1 Summary of data for the age assessment sessions using the cropped facial images for each study

Age assessment sessions summary									
Country of study origin	Number of subjects	Number of assessors	Mean number of assessments per photograph	Subject age range in age assessment sessions	Mean chronological age	Middle age of age range presented	Mean perceived age of subjects	Slope of regression line	Probability that slope is equal to 1
Canada (A)	18	136	134.7	28–42	34.6	35.5	32.4	0.99	N/S
Canada (B)	18	59	58.6	52–68	59.9	60.5	58.1	0.70	N/S
England	162	53	51.0	45–75	59.5	60.5	57.9	0.96	N/S
Spain	233	116	55.7	30–57	42.0	45.5	44.9	1.23	<0.0001
				48–70	60.5	60.5	64.2	1.03	N/S
Denmark	204	102	71.3	59–81	68.4	69.5	67.9	0.72	<0.0001
							73.8 ^a	0.61 ^a	<0.0001 ^a
China	239	77	75.2	25–47	35.2	35.5	36.6	0.79	0.0003
				46–70	59.2	58.0	58.7	0.94	N/S

N/S = Not significant ($P > 0.05$)

^a Danish nurse assessor data

assessors were used; for the Canadian study a combination of US and British assessors were used; for the Chinese study only Chinese assessors were used. For the Danish study, 10 female geriatric nurses rated three types of images; the passport-type photographs, the Fuji camera frontal (pre-cropped, eyes open) photographs, and the cropped facial images (as above); one extra nurse evaluated the passport-type and pre-cropped images.

For the cropped images, to assist assessors in the process of defining an age for a subject, two questions were presented via a website. The assessors were first asked to select from a range of age groups that they thought the subject looked like they belonged to and then the assessors were asked for a precise age within the selected age group. An age assessment session consisted of the presentation of 18 subjects to assessors with a number of sessions required to generate ages for all the subjects in a study (see Supplementary Material for further details). The assessors were unaware that the middle age from the age range presented approximated to the mean chronological age of the subjects. The Canadian, Spanish and Chinese studies were split into separate 'young' and 'old' subject ageing sessions due to the large subject chronological age ranges. Ten of the images from the English study were included in the Spanish age assessment sessions. Gender and age

data were collected from the assessors who completed the Canadian study.

Statistical analysis

Perceived age was calculated as the best linear unbiased predictor of the mean estimated age of the person in the image from a Proc-mixed model (SAS 9.1) with the viewing order as a fixed effect and assessor and image as random effects.

For analysing the reproducibility of the perceived age scores, repeated random samplings from the panel of assessors for particular sample sizes were carried out. The mean age was calculated from these samples of assessors for each image and the maximum age range, the mean inter-quartile range, the mean inter-quartile percent, Cronbach's alpha and Pearson correlation scores were generated from these means (see Supplementary Material for further details). The Canadian age groups were analysed separately to examine differences in assessments of young and old adult subjects.

Results

Influence of the age range presented to assessors

For each of the studies carried out, a minimum of 50 assessments per image were recorded (Table 1). When

judging age, assessors first selected an age group from a range of age groups and then chose a precise age within the selected age group (Supplementary Material Fig. 1). The middle age of the age range presented during the age assessments of the younger Spanish subjects was higher than the mean chronological age in concordance with the resulting mean perceived age. However, such a concordance was not found for the age assessments of the older Spanish subjects or for the Canadian, English or Danish studies (Table 1). Thus, there was no evidence the age range presented to the assessors influenced whether the mean perceived age was higher or lower than the mean chronological age. Assessors had a slight preference for the middle age option within their selected age group (Supplementary Material Table 1).

Regression towards the mean

There is evidence that assessments of age tend towards the mean perceived age (Rexbye et al. 2006). Analysis of the regression of the perceived age data to the chronological age data reveals a regression to the mean for the Danish study (for both sets of assessors) and for the younger age assessment sessions in the Chinese study (Table 1). However, none of the other studies had a significant regression to the mean and the assessment sessions for the

younger subject age range in the Spanish study had a tendency to regress away from the mean.

Reproducibility of perceived age across studies

Increasing the number of assessors used to generate the mean perceived ages increased the reproducibility of the perceived ages, although the incremental gain diminished as more and more assessors were used (Table 2, Canadian Study A). The majority of mean age judgements for an image from groups of 20 assessors were within 3 years of each other for all the studies (Table 2). The age judgements were marginally more similar for the young Canadian subjects than for the older subjects (Table 2, mean inter-quartile). However, this variability represented a slightly higher percentage of the perceived ages of the younger subjects than the older subjects (Table 2, mean inter-quartile percent).

Assessor nationality, gender and age effects on perceived age

There was a strong correlation between the Spanish and British assessor means for the 10 English subject images included in both studies (Table 3) although the Spanish absolute ages were higher (Supplementary Material Fig. 2). Assessors from the Canadian

Table 2 Comparisons of similarity between mean perceived ages given by groups of 20 randomly selected assessors within each study, and data from the Canadian study (A) on the effects

on reproducibility of perceived age by changing the numbers of assessors from 1 to 50

Sampling of assessor data							
Country of study origin	Assessor sample size	Maximum range for an image (years)	Mean inter-quartile (years)	Mean inter-quartile percent	Cronbach's alpha	Mean correlation between iterations	Minimum correlation between iterations
Canada (A)	1	30	7.78	25.14	0.972923	0.674	0.244
	5	14	3.52	10.99	0.992578	0.885	0.755
	10	8.3	1.99	6.04	0.996614	0.949	0.899
	20	5.3	1.32	4.07	0.998703	0.978	0.954
	30	5.2	1.28	3.97	0.999043	0.983	0.959
	50	4.4	1.08	3.35	0.999411	0.990	0.978
Canada (B)	20	7.2	1.70	2.97	0.998159	0.968	0.931
England	20	7.4	1.54	2.76	0.999443	0.990	0.983
Spain	20	16.2	2.95	5.90	0.999027	0.982	0.971
Denmark	20	11.8	2.12	3.16	0.997778	0.960	0.937
China	20	11.9	2.09	4.77	0.999368	0.989	0.976

Table 3 Comparisons between perceived ages resulting from changing the type of image or the composition of assessors for age assessment

Image	Assessors	Correlation, <i>R</i>	<i>P</i> -value
Cropped image	Nurses versus British	0.935	<0.0001
Passport-type versus pre-cropped image	Nurses	0.829	<0.0001
Pre-cropped versus cropped image	Nurses	0.804	<0.0001
Passport-type versus cropped image	Nurses	0.756	<0.0001
Passport-type image versus chronological age	Nurses	0.691	<0.0001
Pre-cropped image versus chronological age	Nurses	0.689	<0.0001
Cropped image versus chronological age	Nurses	0.610	<0.0001
Cropped image versus chronological age	British	0.604	<0.0001
10 English subject images	Spanish versus British	0.965	<0.0001
Canadian study (A and B) images	Male versus Female	0.996	<0.0001
Canadian study (A and B) images	'Old' versus 'Young'	0.994	<0.0001

study were split by gender and age (using 40 years old as a cut-off) to determine whether these factors have an effect on the mean perceived ages generated. There was a very high correlation between the male and female assessor data and between the younger and older assessor data (Table 3).

Assessor expertise and image effects on perceived age

For the Danish study, two photograph formats were acquired and three types of image were assessed by the Danish geriatric nurses (Fig. 1); the cropped facial images were also assessed by British assessors. The correlation between the nurses' and the British assessors' age judgments on the cropped facial images was the highest found (Table 3), although the Danish nurses rated the images older than the British assessors (Table 1). The correlation between the ages from the nurses' judgements on the passport-type images and the pre-cropped facial images was slightly higher than that between their age judgements on the cropped and pre-cropped facial images (Table 3). The lowest correlation for the nurses' data was between the passport-type and cropped facial images, although this was higher than any of the correlations between the perceived and chronological ages (Table 3).

Discussion

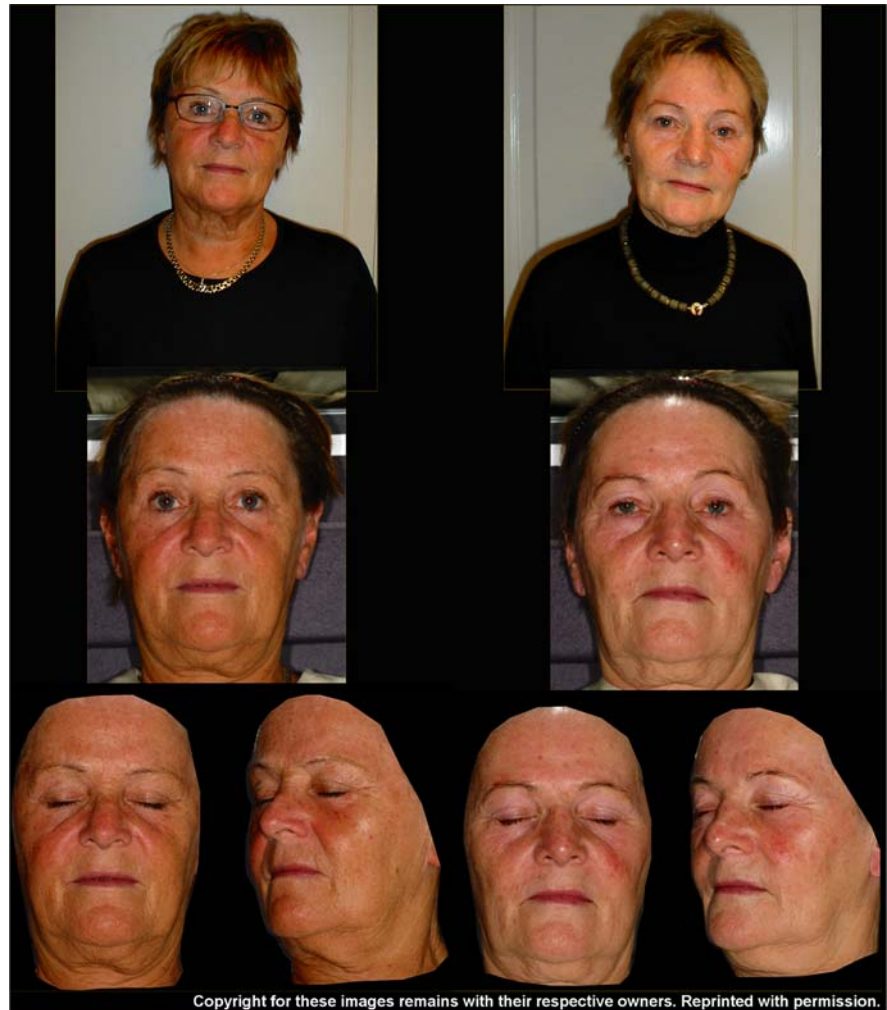
In this study, the way in which various parameters involved in the generation of perceived age affected

the resulting data was examined. As an age range was presented to assessors alongside the cropped facial images, the middle age of the range could have indicated to assessors what the average chronological age was of the subjects. However, whether the middle age of the age range presented to assessors was higher or lower than the subjects' mean chronological age did not influence whether assessors thought the subjects, on average, looked older or younger than their chronological age. In the calculation of the mean perceived age, 'assessor' was fitted as a random effect. This compensates for assessors who judge the differences between images similarly to the panel of assessors but judge the absolute ages consistently either higher or lower than the panel of assessors. This method reduced the variance in the age data (data not shown) and, hence, its use is recommended for perceived age studies.

Previously studies have indicated a tendency towards the mean in perceived age data (Rexbye et al. 2006). Here, there is no indication that this tendency is universal across studies and in the Spanish study there was a tendency away from the mean. Hence, any tendency towards the mean is likely to be driven by the composition of subjects presented to the assessors (for example the presence of very old subjects) rather than by the assessment of age *per-se*.

For all the studies, the perceived ages were generally reproducible to within three years when 20 age assessors were used (Table 2). Although such detailed analysis of the reproducibility of perceived

Fig. 1 Example of an identical twin pair demonstrating differences in their perceived ages depending on type of image presented to the assessors. The top images were judged by the nurses to be 6.9 years different, the middle images 4.5 years and the lower images 0.9 years, with the older looking twin on the right for all images. The British assessors judged the lower images as 0.7 years different, with the older looking twin on the left



age has not been previously reported, similar perceived age measures have been carried out in Scotland (Burt and Perrett 1995), Holland (Rikkert 1999) and Denmark (Rexbye et al. 2006) and all have reported low assessor error. There was a substantial gain in the reproducibility of the mean perceived ages when comparing data from each assessor through to groups of 50 assessors, although there was a decrease in the incremental gain. Based on the data presented here, a minimum of 10 assessors is recommended for perceived age studies and larger numbers are recommended to minimise the probability that the data would not be representative of that from a very large assessor population.

Perceived age data generated by different nationalities on the same images resulted in high correlations

between the data. However, the Spanish assessors rated the 10 English subjects older than the British assessors did and the Danish nurses rated the Danish subjects older than the British assessors did. Thus, although the age differences between the images is maintained, different nationalities have a different perception of the absolute ages of subjects present in cropped facial photographs. This emphasises the importance of comparing perceived age differences within a study population rather than the perceived age scores *per-se*. For example, a subject who looks a year younger than their chronological age should not be considered a 'good' ager if they are part of a large study population whom are judged, on average, to look two years younger than their chronological age. No data was collected on cross-ethnicity age judgements and it is

unclear if this factor affects the variability and reproducibility of age judgements.

There was no evidence for any assessor age effect on the perception of age in the Canadian study. In support of this, Burt and Perrett (1995) also found that there was no difference between using older or younger assessors in the generation of perceived age. No assessor gender effect was observed on the mean perceived ages. There is evidence that the cues used to judge age differ depending on the age group under study; for example, male pattern baldness affects the perceived age of young but not old individuals (Rexbye et al. 2005). However, even if there were different cues used to judge the age of the younger compared to the older Canadian subjects, there were no large differences between these two age groups in the reproducibility of their perceived ages given by groups of 20 assessors. The composition of the assessors, therefore, in terms of gender and age did not have an impact of the resulting perceived ages, and the age of the subjects did not have a large impact on the variability of the perceived age data.

There was a high correlation between the perceived age data from the British assessors and Danish geriatric nurses judging the cropped images. Both groups of assessors also gave a tendency towards the mean. Thus, there was little difference between the perceived age data from a group of assessors experienced in gerontology and from a group of 'naïve' assessors.

Perceived age is generated from age assessments of many visual cues and it is unclear which of these cues are the most important and how they are combined to generate an age score for a particular type of image. The four types of perceived age measures in the Danish study were more closely related to each other than to chronological age, suggesting similar cues are used to judge age in the images and/or there is a correlation between the changes in the appearance of particular cues with age. However, there was a marked reduction in the similarity between the perceived ages when the images were changed. The pre-cropped images were marginally more similar to the passport-type images than the cropped images for the nurses' data. This is despite a greater loss of cues from the passport-type to the pre-cropped images (loss of posture, clothing and grooming behaviour cues) than the pre-cropped to cropped images (loss of eye colour and hair cues).

It is unknown whether it was the loss of the hair and eye information or the addition of the 45-degree angle facial view that drove the variation between the pre-cropped and cropped images. The addition of the 45-degree angle image could have induced a greater influence of face shape to the perceived age measure and, in support of this, this facial feature is thought to have a strong influence on the judgement of age (Krejci-Papa and Langdon 2006). Therefore, the features present in images have a marked effect on the resulting perceived ages and, hence, for comparisons to be made across perceived age studies the type of images used for age assessment should be the same.

In conclusion, the perceived age methodology detailed here can be used globally to generate a reproducible biomarker of facial ageing when large numbers of assessors are used irrespective of subject age and assessor gender, nationality, age and ageing expertise. Further work into the stability of this measure over weeks and years of a subject's life and whether it relates to objective measures of physiological ageing will be required to determine the relevance of this measure in investigating the biological processes underlying facial ageing.

Acknowledgements The authors would like to acknowledge all the individuals involved from Unilever HPC Spain Madrid, Unilever Research and Development Trumbull US, Unilever Research China, and also Frans Van-der-Ouderaa, Michael Catt, Stewart Granger, Gail Jenkins, Amelia Fereday, Martin Green and Jakob Mortensen for making the work possible. The authors would also like to acknowledge the contribution of the assessors and subjects who took part in the studies.

References

- Borkan GA, Norris AH (1980) Assessment of biological age using a profile of physical parameters. *J Gerontol* 35: 177–184
- Bulpitt CJ, Rajkumar C, Cameron JD (1999) Vascular compliance as a measure of biological age. *J Am Geriatr Soc* 47:657–663
- Bulpitt CJ, Markowe HLJ, Shipley MJ (2001) Why do some people look older than they should? *Postgrad Med J* 77:578–581
- Burt DM, Perrett DI (1995) Perception of age in adult caucasian male faces-computer graphic manipulation of shape and color information. *Proc R Soc Lond B Biol Sci* 259:137–143
- Christensen K, Iachina M, Rexbye H, Tomassini C, Frederiksen H, Mcgue M, Vaupel JW (2004) "Looking old for

- your age”: genetics and mortality. *Epidemiology* 15: 251–252
- Fink B, Grammer K, Madsen PJ (2006) Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. *Evol Hum Behav* 27:433–442
- Ingelsson E, Sullivan LM, Murabito JM, Fox CS, Benjamin EJ, Polak JF, Meigs JB, Keyes MJ, O’Donnell CJ, Wang TJ, D’Agostino RB Sr, Wolf PA, Vasan RS (2007) Prevalence and prognostic impact of subclinical cardiovascular disease in individuals with the metabolic syndrome and diabetes. *Diabetes* 56:1718–1726
- Krejci-Papa, Langdon RC (2006) In: Gilchrist BA, Krutmann J (eds) *Skin aging in three dimensions*. Springer-Verlag, Berlin, pp 133–142
- Rexbye H, Petersen I, Iachina M, Mortensen J, Mcgue M, Vaupel JW, Christensen K (2005) Hair loss among elderly men: etiology and impact on perceived age. *J Gerontol A Biol Sci Med Sci* 60:1077–1082
- Rexbye H, Petersen I, Johansens M, Klitkou L, Jeune B, Christensen K (2006) Influence of environmental factors on facial ageing. *Age Ageing* 35:110–115
- Rikkert MGMO (1999) Visual estimation of biological age of elderly subjects: good interrater agreement. *Gerontology* 45:165–167
- Seeman TE, McEwen BS, Rowe JW, Singer BH (2001) Allostatic load as a marker of cumulative biological risk: MacArthur studies of successful aging. *Proc Natl Acad Sci USA* 98:4770–4775
- Warren R, Gartstein V, Kligman AM, Montagna W, Allendorf RA, Ridder GM (1991) Age, sunlight, and facial skin—a histologic and quantitative study. *J Am Acad Dermatol* 25:751–760
- Yamaguchi MK, Oda M (1999) Does cardioid strain change in real front-view facial images tend to change the perceived age? *Electron Comm Jpn* 82:39–48