Proceedings of the 22nd Annual Conference of the Cogntive Science Society, Philadelphia, August 13-15, 2000
L.R. Gleitman & A.K. Joshi, Eds pp. 322-327. See also: http://www.mic.atr.co.jp/~mlyons

Viewpoint Dependent Facial Expression Recognition Japanese Noh Masks and the Human Face

Michael J. Lyons†, Andre Plante†, Miyuki Kamachi‡, & Shigeru Akamatsu‡
(mlyons@mic.atr.co.jp)
Advanced Telecommunications Research International
†Media Integration & Communications Research Lab
‡Human Information Processing Research Lab
2-2 Hikari-dai, Seika-cho, Soraku-gun, Kyoto, Japan, 619-0288

Ruth Campbell & Mike Coleman
r.campbell@ucl.ac.uk
Dept. of Human Communication Science, University College London
Chandler House, 2 Wakefield St, London WC1 N 1 PG

Abstract

With certain masks used in Japanese Noh drama the apparent facial expression is a function of the vertical viewing angle. Rotation in depth produces changes in the retinal image of the face which viewers may confound with the distortion of features due to muscular action. In particular, as the mask is tilted forward it appears to smile, and as it is tilted backwards it appears sad. We explored this effect in two experiments with a Noh mask and one with a 3-D model of a laser-scanned human face. Separate British and Japanese subject pools were used to investigate cross-cultural effects. The results confirmed a systematic relationship between vertical angle of view and judged affect. For the Noh mask the effect was culturally moderated, whereas for the human face there was no significant effect of of culture. These results are discussed and interpreted in terms of perceptual strategies for processing familiar and unfamiliar faces.

Introduction

A variety of visual cues inform the viewer of affect. Muscular action of the face (Ekman & Friesen, 1978), causing feature displacement and consequent wrinkling of the skin, conveys the most salient information. Visible changes in skin hue caused by modulation of blood flow are also telling signs to internal state. A further source of information about affect is delivered by body posture: positive affect is accompanied by an upright posture with the head held high and negative states may be signaled by a bowed head and crouched posture (Darwin, 1872).

Visual processing of the feature displacement and textural cues to face muscular action requires a representation sensitive to the fine metric properties of the spatial patterns on the surface on the face (Lyons *et al.*, 1999, 2000). Rotation and translation of the head in 3-D space accompanying vertical movements of the head or changes in viewpoint, distort configural relations on the face as they appear in the 2-D projection of the face on the retina. Indeed the signals from facial muscle action and head posture conflict. Affect should be judged relatively positively in a head held up and back. Under frontal viewing conditions, tilting the head backwards reduces upward curvature of the mouth in the 2-D projection of the face, giving the impression of a sadder or negative expression. Tilting the head forward, a negative head posture signal, increases upward curvature of mouth - which usually accompanies the muscular action signaling a smile. It is therefore interesting to ask whether the changes due to rigid displacement of the head interferes with the interpretation of featural and textural cues due to muscular action of the face. Our interest in this question was stimulated when we learned of an illusion of facial expression perception involving masks (figure 1) used in Japanese traditional Noh drama (Komparu, 1983). It has been known for centuries in the Noh theater that certain masks, particularly those used to portray young female roles, appear to change expression as the vertical inclination of the mask changes (figure 2). Tilt the mask for-

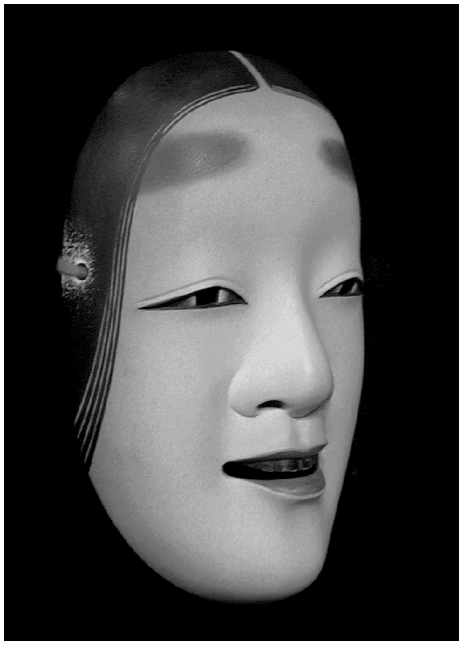

Figure 1: Magojiro mask used in the Japanese Noh drama.

ward and it appears to smile; tilt it backwards and it appears sad.

Is this phenomenon evidence for a lack of invariance of the facial expression recognition system under rigid transformations? Or are special techniques employed by the mask carver and Noh actor to trick the visual system into mistaking a rigid rotation of the mask for a non-rigid distortion of its internal features? We conducted three experiments to investigate the following questions (1) Do changes in vertical inclination in fact generate different perceptions of affect? (2) Is the effect culturally moderated or does it depend on familiarity with the mask? (3) Is the effect particular to the Noh mask or does it generalize to the human face?

Materials and Methods

Stimuli

Stimuli for **Experiment 1** consisted of photographs of a Noh mask (figure 2) at 13 inclinations, from -30° to +30° in equal 5° increments. An antique Magojiro mask, used for young female roles, dating to the Edo period (1600-1868) was photographed on a Noh stage under lighting conditions similar to what would be used during a performance. The mask was photographed from a frontal viewpoint using a digital camera (Kodak Professional DCS 460) from a distance of 7.7m with a 200 mm lens. The 3060x2036 pixel 24-bit color images were cropped and re-sampled to 300x400 pixel tiff images. Stimuli for **Experiment 2** (figure 3) comprised the same images, but cropped so as to emphasize the internal features of the face. Stimuli for **Experiment 3** (figure 4) were derived from the head and face of a 30 year old Japanese female model posing a neutral expression similar to that of the Noh mask. A Cyberware 3030 Color 3-D scanner was used to acquire shape and color information of the model's head. The 24-bit RGB color map was acquired under room light from an overhead fluorescent lamp. Screen captures were taken at 13 (virtual) head inclinations from a reconstructed 3-D model of the head, the face oriented frontally and saved as 24-bit 300x400 pixel tiff images. The vertical viewing angles varied from -30° to +30° in equal 5° increments. Inter-ocular distance and eye position were normalized for each stimulus set and matched across sets.

Experimental Procedure

Experiments were run in separate laboratories in London and Kyoto. In each case the stimuli were displayed on a 17 inch 24-bit color computer monitor in a slightly darkened room. Viewing distance was approximately 60 cm. Following a practice trial, four epochs of all 13 stimuli were presented in succession, with presentation order randomized within each epoch. Presentation order was as follows: fixation point (500 ms) - blank (400 ms) - stimulus (300 ms). Subjects were instructed to respond whether the stimulus face appeared happy or sad by pressing the left or right shift key. Japanese subjects were instructed in Japanese using the terms *"yorokobi"* and *"kanashimi"*. Left/right assignment of response keys was counterbalanced across subjects. The words "happy" and "sad" (in English for both subject groups) appeared on the response-appropriate side of the screen for each subject to maintain correct response orientation. All Japanese subjects were familiar with the English terms. Reaction times and decision type were recorded automatically for each subject for each trial.

Subject Pool

Different subjects were run for each of the three experiments. There were 5 females and 5 males from each cultural group for each experiment, making 60 subjects in total. Subjects were undergraduates, graduates and staff from Doshisha University, Kyoto and University College London. Ages ranged from 18 to 50 years. All had normal or corrected-to-normal vision. All were either native to the country of testing or had first-school education in that country. The Japanese subjects were familiar with Noh masks as images or, occasionally, objects. None of the UK subjects had familiarity with Noh or had visited Japan.

Results

All three experiments had the same mixed, 3-factor repeated measures design. There were 13 levels of the first factor (inclination), which varied within subjects, and one level of the first between-subjects factor (culture) and the second between-subjects factor (gender). Both RTs and response (as the proportion of "happy" responses over 4 trials) were examined. Following initial analysis, in which the full range of scores were examined, only the seven mid-range scores (-15 ° to + 15 °) in the following treatment. Preliminary data analysis confirmed parametric, normal distributions of the scores reported here. There were no significant differences in the pattern of results when the full range was included. However, the full-range analyses may be less reliable due to non-Gaussian distribution of scores at some of the endpoints. RT (medians) showed no systematic relationship to the other variables, and are not reported here. Gender had no effect on any of the analyses and effects of gender were not considered further. Table 1 outlines the significant finding for each experiment. The relevant graphs are shown in figure 5. These analyses show a significant linear relationship between angle of inclination and rated happiness for all three experiments. The Noh mask, but not the scanned face, is classified differently by Japanese and British viewers. Further analyses and their justification in terms of individual experimental hypotheses are reported below.

Experiment 1 - Full Mask

The experimental hypothesis was that the Noh mask would generate changes in perceived affect as a function of vertical angle. The perception of facial expressions is thought to be similar for our two cultural groups (Matsumoto, 1992), hence similar results were expected for the two subject groups.

The results confirmed the predictions in general terms, but with some important deviations. The groups differed in

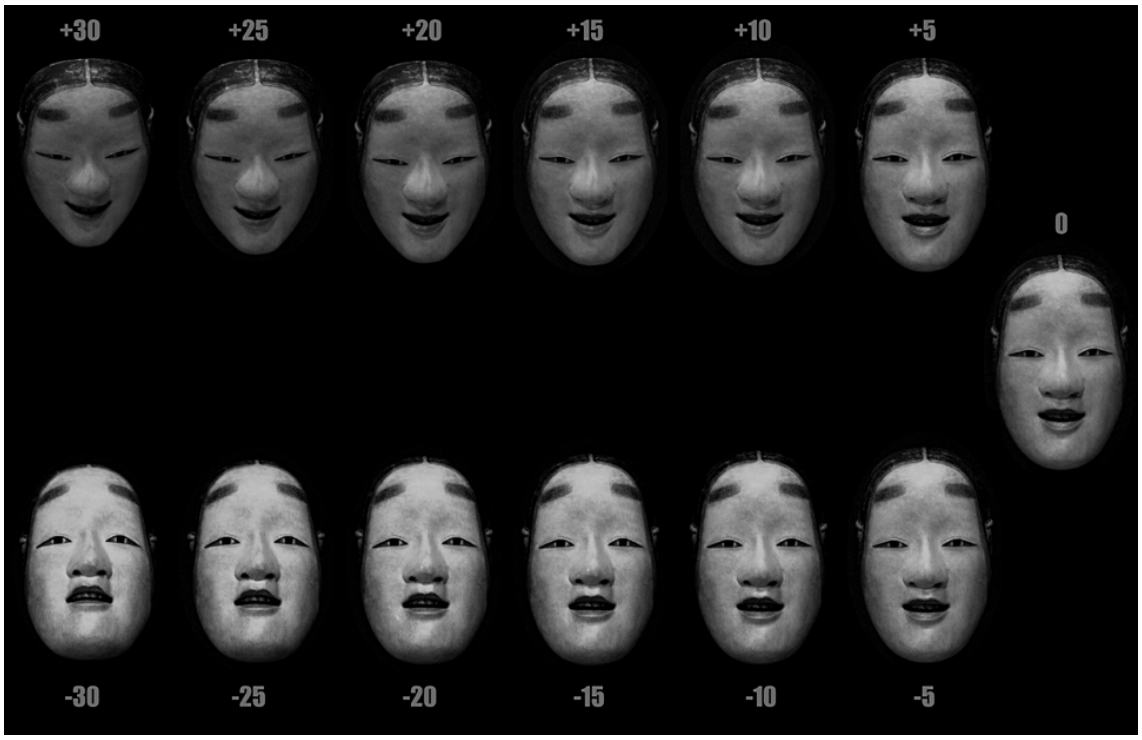

Figure 2: Stimuli for Experiment 1. Edo-period Magojiro mask at 13 different vertical inclinations.

Figure 3: Sample stimuli for Experiment 2. Same images as in figure 2 but with the edges of the mask cropped to highlight internal features. Images from left to right show the mask at inclinations of -15°, 0°, and 15°.

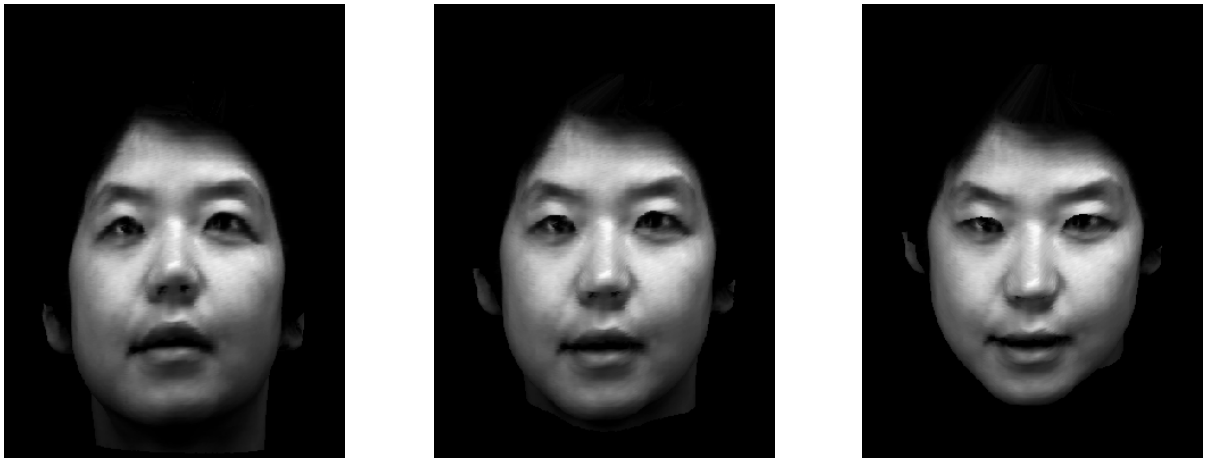

Figure 4: Sample stimuli for experiment 3. Images captured from rendered 3-D model of a Japanese female face obtained using a Cyberware laser scanner. Images from left to right show the face at inclinations of -15°, 0°, and 15°.

the inclination at which a "neutral" expression is perceived. Japanese viewers rated the back-tilted mask more positively than British viewers. This may reflect different boundaries in terms of perceived facial expression on the categories of "happy" and "sad" and their cognates in Japanese or lower rates of sadness recognition for Japanese viewers. We address this further in Experiment 3. A second unexpected finding is that the groups differed in the linearity of the relationship with inclination angle (interaction of group and inclination was highly significant). While the relationship was linear over this range for the British subjects, for the Japanese subjects, the proportion of "happy" responses peaked at 5 °, and then dipped. At 15 °, mean "happy" response was no greater than at -10 °. Why should this change in perceived expression occur? One possibility is that the Japanese viewers were more sensitive to postural cues in the images of the head. A head bowed forwards may be seen as "sadder". Perhaps the two subjects groups weight the posture and internal features cues differently. For the Noh mask images used, pose cues are most visible in the disposition of the top of the head and the chin with change in inclination of the head.

Experiment 2 - Cropped Mask

In this experiment, the face images were cropped to diminish cues to head pose and emphasize internal features of the face (figure 3). The experimental prediction was that this may eliminate the non-linearity in the relationship between perceived expression and vertical inclination in the Japanese viewers. The results supported this. In this study, the "dip" at greater positive inclinations was greatly reduced. Thus it appears that Japanese viewers take account of cues to head pose in ascribing expression to the image of the vertically inclined mask.

Otherwise, experiment 2 replicates the main findings of experiment 1: a linear relationship between angle of inclination and judged expression; as well as a group difference emerged between Japanese and British viewers. The group difference may indicate that the terms "happiness" and "sadness" in English and Japanese do not share similar extensions. This would suggest that Japanese may be more willing than British viewers to ascribe "happiness" (a socially acceptable facial signal) to a relatively "unhappy" face, despite the apparent reversal of this pattern for masks at high forward tilt. If this were so, we would expect a similar disparity between groups to emerge when images of natural faces are perceived. Experiment 3 explores this possibility.

Experiment 3 - Human Face

This experiment used stimuli generated from a 3-D laser scan of a human face (figure 4) to explore the question: do group differences in ascribing expression to a cultural artifact, the Noh mask, extend to natural face images? If they do, we may infer that cultural and linguistic interpretations of facial expression may differ between these groups. If they do not, then the Noh mask may have special perceptual status for Japanese viewers. The findings strongly support the latter conclusion. The relationship between inclination angle and happy-sad judgments was to all purposes identical in both groups. Mann-Whitney non-parametric t-tests explored group differences at each orientation point. None approached significance. We can conclude that the Noh mask effects reported in Experiments 1 and 2, including both the dip in the function at high angles of forward tilt and the "happier" classification at most other angles, reflected a cultural phenomenon - but one related to *perceptual* processing differences between the groups. Though the laser-scanned face did not replicate the lighting conditions of the naturally photographed images used in Experiments 1 and 2, the relationship between inclination and judged expression still held, suggesting that the difference in lighting differences did not contribute notably to the illusion for this set of conditions.

Table 1: Summary of F values, separate ANOVAs for each experiment (SPSS GLM).

	F(6,216) Main effect of inclination	F(1,36) Main effect of group	F(6,216) Group x inclination interaction	F(1,36) Main effect of inclination Linear trend	F(1,36) Group x inclination Quadratic fit
Experiment 1 Full Mask	10.5 $p < 0.001$	NS	4.67 $p < 0.001$	15.33 $p < 0.01$	7.56 $p < 0.01$
Experiment 2 Cropped Mask	23.9 $p < 0.001$	5.27 $p < 0.05$	3.47 $p < 0.01$	55.77 $p < 0.001$	3.47 $p < 0.01$
Experiment 3 Scanned Face	14.96 $p < 0.001$	NS	NS	36.88 $p < 0.001$	NS

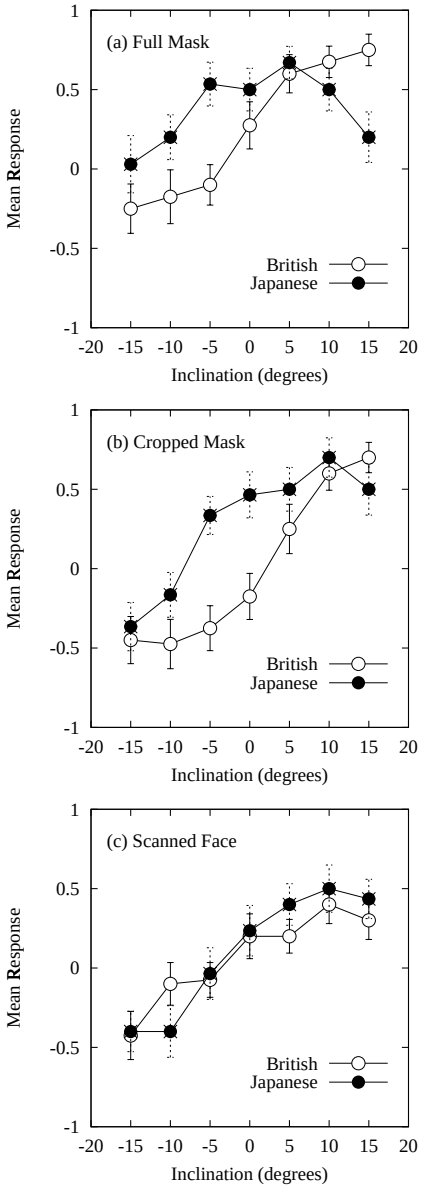

Figure 5: Mean response versus vertical inclination for the 3 experiments.

Discussion

The three experiments confirmed that the angle of vertical inclination of a face profoundly influences a simple expression discrimination task: faces tilted down have a happier cast than those tilted back. This may be understood in terms of the projection of the three-dimensional facial surface onto the image. An earlier study (Cavanagh *et al.*, 1988) noted the effect as an example of the failure of shape constancy under rotation in depth. Another study (Kappas *et al.*, 1994) looked at viewpoint dependence of facial expression recognition using video clips of posed dynamic expressions as well as a schematic wire-frame model of the face, both quite different from the Noh mask stimuli and scanned face used here. However, that work did not attempt to look at facial expression cues separately from pose cues, as in experiment 2, or study different cultural groups.

A surprising but consistent (40 subjects altogether) finding was that the Noh mask elicited different responses in the two cultural groups. The skilled processing of faces has typically been described as configurational (Diamond & Carey, 1986; Young *et al.*, 1987). That is, skilled viewers take account of the various face features and their disposition in coming to a unified account of the identity or reading of the face. Their reading of the face cannot be predicted on the basis of local featural detail. One possibility is that familiarity may have delivered a greater degree of configural processing for the mask in Japanese than British viewers, for the Noh mask occasionally appears in the Japanese media, though an understanding of Noh, or interest in Noh as a tradition is no longer widespread in the Japanese population. One local feature that reliably signals "happy-sad" is the curvature of the mouth. It is possible that British viewers of the Noh mask took account of this feature alone. For Japanese viewers, other aspects of the face may have moderated the effect. Only further experiments will indicate what facial aspects these may be.

At the outset of these studies, we speculated that the three-dimensional structure of the Noh mask and the disposition of the painted features, may be intentionally designed to elicit changes of perceived expression with small changes in pose. Examination of the 3-D structure of the mask showed, for example, that the depth of the mouth region is exaggerated

relative to the human face. Our psychological studies confirmed that small changes in pose of the mask lead to significant changes in perceived affect. A forward tilted mask appeared relatively happy and one tilted backwards, relatively sad. Paradoxically, however, in the stylized use of mask pose in Noh drama, the convention is the opposite to our findings. In one gesture known as *terasu* (shining), signifying a happy state, the mask is turned upwards. In another known as *kumorasu* (clouding), signifying a sad state, the mask is turned downwards (Komparu, 1983).

In this connection, it is notable that Zeami (1363-1443), the most influential early Noh dramatist, ranked *yugen*, or subtle profundity, as the highest aesthetic principle of Noh (Zeami, 1968). In the framework of the Noh world, a joyful pose tempered with a slightly sad mouth may be appreciated as more beautiful than a direct expression of joy. Likewise, sadness or pain masked with a smiling mouth suggests more emotional complexity than an display of pure sadness. A further interpretation is possible, not necessarily in conflict with the above. The psychometric curves (figure 5) show that small changes in inclination angle significantly affect perceived facial expression. Minor movements of the actor's head may trick viewers into thinking that the internal features of the mask are moving non-rigidly as if it were an animated living face. One of the authors (MJL) has observed this effect while watching a Noh play. In a related perceptual effect, a rigid 3-D stick man figure rocked longitudinally back and forth can appear to walk with non-rigid limb movement in 2-D projection (Sinha & Poggio, 1996).

Acknowledgements

We are grateful to Hotaka Komparu for loan of the Magojiro mask and a helpful discussion about Noh drama; the management at Nara Shin Kokai Do for access to the Noh stage; Akiko Tohma and Katsunori Isono for their generous help; Larry Maloney and Patrick Cavanagh for stimulating discussions; and Vicki Bruce for encouraging us to pursue this study.

References

Cavanagh, P., Peters, S. & von Grnau, M. (1988). Rigidity failure and its effect on the Queen. *Perception17* suppl. 27A.

Darwin, C. (1872). *The expression of the emotions in man and animals*. London, John Murray.

Diamond, R. & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *J. exp. Psychol. 115*, 107-117.

Ekman, P. & Friesen, W.V. (1978). *Facial Action Coding System*. Palo Alto, CA, Consulting Psychologists Press.

Kappas, A., Hess, U., Barr, C. & Kleck, R. (1994). Angle of regard: the effect of vertical viewing angle on the perception of facial expressions. *J. Nonverbal Behav. 18*, 263-280.

Komparu, K. (1983). *The Noh Theatre: Principles & Perspectives*. New York & Tokyo, Weatherhill/Tankosha.

Lyons, M.J., Budynek, J., & Akamatsu, S. (1999). Automatic Classification of Single Facial Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 1357-1362.

Lyons, M.J., Morikawa, K., & Akamatsu, S. (2000). A Linked Aggregate Code for Processing Faces. *Pragmatics & Cognition 8*, 63-81.

Matsumoto, D. (1992). American-Japanese cultural differences in the recognition of universal facial expressions. *J. Cross-Cultural Psychol.23*, 72-84.

Sinha, P. & Poggio, T. (1996). Role of learning in three-dimensional form perception. *Nature384*, 460-463.

Young, A.W., Hellawell, D., & Hay, D.C. (1987). Configural information in face perception. *Perception16*, 747-759.

Zeami, M. (1968). *Kadensho*. Kyoto, Doshisha Univ., Sumiya-Shinobe Publ. Inst.