Neuropsychologia 45 (2007) 506–513

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

Are representations of unfamiliar faces independent of encoding modality?

Sarah J. Casey, Fiona N. Newell *

School of Psychology and Institute of Neuroscience, Trinity College, Dublin, Ireland

Received 29 August 2005; received in revised form 20 February 2006; accepted 21 February 2006
Available online 4 April 2006

Abstract

It is well documented that both featural and configural information are important in visual face recognition. Less is known, however, about the nature of the information underlying haptic face recognition and whether or not this information is the same as in vision. In our experiments we found better within modal than crossmodal face recognition performance suggesting that face representations are largely specific to each modality. Moreover, this cost in crossmodal performance was found to be independent of differences in exploratory procedures across the modalities during encoding. We found that crossmodal face perception was most efficient when configural information of the facial features was preserved suggesting that configural information is shared across modalities. Our findings suggest that face information is processed in a similar manner across vision and touch but that qualitative differences in the nature of the information encoded underlies efficient within modal relative to crossmodal recognition.

Keywords: Face recognition; Haptics; Crossmodal; Configural processing

1. Introduction

Although there has been much research into how visual and haptic information is integrated into multisensory object representations in memory (e.g. Ernst & Bülthoff, 2004; James, James, Humphrey, & Goodale, 2005; Newell, 2004) relatively little is known about how information about faces is integrated across the modalities. Faces represent a socially important, highly homogenous set of objects which we are adept at identifying. Moreover, face perception is often considered as 'special' relative to general object perception (e.g. Yin, 1969; Kanwisher, 2000) although others have argued that face perception is similar to object perception but that mechanisms for identifying faces are optimised due to practice (e.g. Diamond & Carey, 1986; Gauthier & Tarr, 1997). This debate, however, does not speak to the issue of the contribution of information from other modalities in face recognition, although it could be argued that face perception is different from general object recognition in that it relies on processes or expertise *within a particular modality*, i.e. vision. We might therefore expect that the contribution of information from another modality, such as touch, would be minimal and that processes involved in the visual recognition of faces are not the same as those involved in haptic face perception.

In terms of general object recognition, previous research suggests that information from multiple sources of sensory inputs contributes to the representation of an object in memory (Easton Greene, & Srinivas, 1997; Newell, Ernst, Tjan, & Bülthoff, 2001; Reales & Ballesteros, 1999). Such multisensory object representations allow for more robust recognition performance (Ernst & Bülthoff, 2004). Given that faces constitute an ecologically important stimulus, the question arises as to whether information from other sensory modalities can contribute to the formation of robust multisensory face representations. It seems likely that voice information can serve as a marker for personal identity as visual and auditory information for person recognition generally occur in close temporal proximity and are thus likely to be integrated to generate a robust representation of the speaker. However, visual and haptic face information do not automatically co-occur in the real world. Therefore substantial learning may be necessary to associate the two types of information. On the other hand, both the visual and haptic modalities can process shape information, and both modalities appear to represent this information in similar formats conducive to crossmodal information

* Corresponding author at: School of Psychology, University of Dublin, Trinity College, Dublin 2, Ireland. Tel.: +353 1 608 3914; fax: +353 1 671 2006.
E-mail address: fiona.newell@tcd.ie (F.N. Newell).

0028-3932/$ – see front matter
doi:10.1016/j.neuropsychologia.2006.02.011

sharing (e.g. Newell et al., 2001). We might therefore expect that haptic information may contribute to visual information to achieve robust face recognition performance. For this to happen, however, face information should be easily shared across modalities. Recent studies have suggested that haptic information can contribute to the recognition of a face. Casey and Newell (2005) and Kilgour and Lederman (2002) have found that unfamiliar faces can be successfully matched across modalities. Perceptual matching, however, may rely on information processing that is not specific to face perception per se but involves more generic, image-based processing. The question arises as to whether or not face representations in memory are modality independent or are more dependent on modality specific information.

Functional models of visual face recognition (e.g. Bruce & Young, 1986), suggest that structural information from a face is encoded and represented in face memory to uniquely identify that face. Since haptics can also encode structural information (e.g. Lederman & Klatzky, 1990) then, in principle, this information could be shared across the senses in face memory. Structural information is generally assumed to be a feature-based description together with a representation of the spatial arrangement or configuration of those features. Many studies have provided evidence for the importance of both featural (e.g. Collishaw & Hole, 2000; Schwaninger, Lobmaier, & Collishaw, 2002) and configural information (e.g. Tanaka & Farah, 1993; Tanaka & Sengco, 1997; Young, Hellawell, & Hay, 1987) in visual face recognition. In the haptic encoding of faces, the encoding of facial information may be limited to a mainly feature-based description since the global face cannot be encoded directly (e.g. Loomis & Lederman, 1986). However, Kilgour, de Gelder, & Lederman (2004) found evidence for a cost in haptic face recognition for inverted relative to upright face masks. This finding suggests that, like vision, the haptic system is also capable of processing and representing faces in terms of the configuration of facial features (c.f. Bartlett & Searcy, 1993; Lewis & Glenister, 2003; Freire, Lee, & Symmons, 2000).

Face recognition, however, may rely more on subtle visual differences between faces. For example, visual features such as pigmentation (Yip & Sinha, 2002), and contrast polarity (e.g. Kemp, Pike, White, & Musselman, 1996) can affect recognition. Furthermore, configural information may be easier to derive from a visually encoded face than one encoded through touch. In haptic perception configural information is indirectly perceived because of the serial nature of feature encoding through touch. Consequently, visual processing of a face may be more conducive to efficient face recognition than haptic processing of a face. This would suggest that face perception may be more modality-dependent than was previously found for the recognition of other types of objects.

In this paper, we investigated the extent to which face representations in memory are shared across modalities. In Experiment 1 we investigated recognition memory performance for a set of unfamiliar faces within and across modalities. We hypothesised that if memory representations contain information that is multisensory then performance within and across modalities should be equivalent. However, if face representations are more modality-specific, then we predicted a cost in crossmodal relative to within modal recognition, due to the lack of transfer of sensory specific face information. In Experiment 2 we controlled for differences in exploratory procedures between the modalities during encoding of face information. In Experiment 3 we investigated whether featural or configural information was shared across modalities by constraining the type of information available for recognition.

2. Experiment 1

In this experiment we examined crossmodal face recognition and predicted equivalent crossmodal and within modal performance if visual and haptic face information is shared within a common structural description in memory.

Given the relative lack of expertise at recognising faces through touch, we attempted to equate performance in the visual and haptic modalities by reducing the number of haptic relative to visual face targets and by allowing longer learning times for haptic relative to visual learning of the faces.

2.1. Method

2.1.1. Participants

Twenty-four undergraduate students from Trinity College, Dublin (mean age 20.7 years, S.D. = 2.18, 10 female) participated in this experiment for course credits or nominal payment. Participants reported normal, or corrected-to-normal, vision and none reported any tactile impairment. This study was approved by the School of Psychology Ethics Committee, Trinity College, Dublin. Written consent was obtained from all participants prior to the experiment.

2.1.2. Materials and apparatus

Twenty-four plaster face models were used as haptic stimuli in this experiment (see Casey & Newell, 2005) for details of the casting of these face models). Colour images of the live faces of the individuals from whom the face models were cast were used as visual stimuli. The individuals used as models in our experiment were previously unfamiliar to participants. Visual stimuli were presented using Microsoft PowerPoint on a Macintosh G3 PowerBook. The images subtended a visual angle of approximately 10° vertically and 7.5° horizontally. Haptic stimuli were presented behind an occluder to the left of the monitor and faced the participant. Visual stimuli were presented in a 3/4 view to maximise structural information.

2.1.3. Design

This experiment was based on a mixed design with learning modality (vision or haptics) as the between subjects factor and recognition task (within or across modalities) as the repeated measure. Participants were randomly assigned to either the visual or the haptic learning modality.

Prior to the experiment, the haptic face stimuli were randomly assigned to one of three target face groups and the visual face stimuli were divided into two target groups. Each participant was assigned one of the haptic face groups or one of the visual face groups as target faces. The remaining faces served as non-target stimuli.

The experiment involved two sessions; a learning block followed by a recognition session. The experiment was based on an old/new recognition paradigm. Unimodal and crossmodal trials were blocked, and these blocks were counterbalanced across participants. Each block comprised 24 trials: eight target and 16 non-targets in the haptic block; 12 targets and 12 non-targets in the visual block. Although we reduced the number of targets in the haptic block to aid memory, we maintained the total number of face stimuli to ensure that each face stimulus was presented the same number of times across modalities. The order of the trials was randomised across blocks and participants.

2.1.4. Procedure

During the learning block, participants were required to learn the name and occupation of each target face and to subsequently identify each face. The twelve visual target images were presented on laminated cards and participants were required to study these images for 5 min in total. Each haptic target face was presented for 1 min until all faces were perceived. This sequence was repeated four times. We set a learning criterion of 7/8 correct responses for haptics and 11/12 correct responses for vision. The learning block was repeated until this response criterion was achieved. On average the learning of the haptic targets took about 50 min to complete whereas learning of the visual targets took approximately 7 min.

The recognition experiment directly followed the learning block. Here participants were required to perform an old/new face recognition task either within the same modality as learning or in a different modality. Participants were informed of the nature of the experimental block (within- or crossmodal) before testing. During the visual block the participant pressed a space bar to begin each trial. A fixation cross was presented for 1 s followed by a visual face image for another second. In a haptic recognition trial a face model was presented behind an occluder and the participant was instructed to palpate the face for 1 min. Exploration time was recorded as soon as the participants' hands came into contact with the face model. After presentation of the stimulus participants were instructed to respond as quickly and as accurately as possible by circling whether the face was old or new on a response sheet.

2.2. Results

Given that the probability levels for the visual and haptic familiarisation conditions differed, we compared performance (i.e. hits and correct rejections) in these conditions to chance levels separately. Although performance in the crossmodal condition was relatively poor, we found that it was significantly better than chance for both the visual [$\chi^2 = 21.87$, d.f. = 11, $p < 0.03$] and haptic [$\chi^2 = 54.70$, d.f. = 11, $p < 0.001$] learning conditions.

A' scores were calculated based on the mean hits and correct rejections for each of the participants. We used A' rather than hits and false positive rates because A' provides a criterion-free measure of participants' ability to discriminate between target and non-target faces. Fig. 1 depicts the mean A' scores across each of the within and across modalities face recognition conditions. These data were analysed using a 2 × 2 mixed design ANOVA with learning modality (vision or haptics) as the between subjects factor, and recognition task (within or across modalities) as the repeated measure. We found no effect of learning modality [$F(1, 22) = 0.94$, n.s.]. However, there was a main effect of recognition modality [$F(1, 22) = 30.23$, $p < 0.001$]. A Newman–Keuls test showed significantly better performance for within relative to crossmodal recognition ($p < 0.001$). There was no significant interaction between the factors [$F(1, 22) = 0.71$, n.s.].

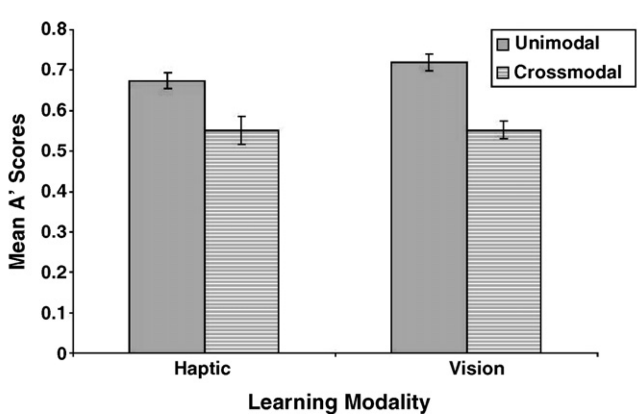

Fig. 1. Plot showing face recognition performance, measured as the mean A' scores, for the within and across modalities conditions in Experiment 1. Error bars represent ±1 standard error of the mean.

2.3. Discussion

We found a cost in crossmodal face recognition relative to the unimodal recognition of the same faces and this cost was incurred independently of the learning modality. Although our participants were able to recognise the faces across modalities, performance was relatively inefficient. These results suggest that face recognition memory is not underpinned by a single multisensory representation.

Our results may indicate that the observed cost in recognition performance reflects modality-specific face information processing. For example, pigmentation information in vision, or surface texture in haptics may promote better recognition performance within each modality. Recognition of faces may therefore be dependent not only on information common to both senses, such as information about face structure, but also on subtle differences in modality-specific face information.

In this experiment the cost in crossmodal recognition performance suggests that a multisensory face representation is not automatically generated during unimodal encoding. On the other hand, if both visual and haptic face information was available during encoding, then subsequent recognition performance in either modality alone may have been improved due to the combination and integration of redundant sensory information in memory (e.g. see Ernst & Bülthoff, 2004). Our preliminary investigations into the role of multisensory interactions in face recognition suggest that this is not the case: nine participants were required to learn 12 target faces using vision and touch together (i.e. bimodal learning) and were subsequently tested either unimodally (i.e. vision or haptics) or bimodally. Although performance was greater than chance in all conditions, haptic recognition was generally poor relative to visual or bimodal performance [$\chi^2 = 0.94$, d.f. = 2, $p < 0.003$], with no difference in performance between these latter conditions. These preliminary findings suggest that visual capture occurs during face recognition, even when haptic information is available. Therefore, unlike the recognition of general objects (cf. Lehmann & Murray, 2005), multisensory encoding does not appear to offer any recognition benefits for faces relative to unimodal encoding.

In Experiment 1, the manner in which face information was encoded was specific to each modality. For example, vision can process all aspects of an image in parallel, so that local facial features and their global configuration can be rapidly processed (Tanaka & Sengco, 1997). Haptic encoding, on the other hand, is limited to serial exploration of an object which probably places demands on working memory due to the integration of object information over time (Loomis & Lederman, 1986; Loomis, Klatzky, & Lederman, 1991). These differences in encoding procedures may have affected representations in memory. In the next experiment we attempted to eliminate encoding differences during familiarisation, in order to investigate whether or not encoding impacted on recognition accuracy in the crossmodal conditions.

3. Experiment 2

In this experiment we predicted that if differences in the manner in which faces are encoded affect the representations of faces in memory, then by making encoding in vision and haptics more similar we expected no cost in crossmodal relative to within modal recognition performance. Here we presented visual face images in parts during learning to promote more serial encoding as in haptics.

3.1. Method

3.1.1. Participants

Twenty-four undergraduate students (mean age 22.25 years, S.D. = 3.05, 13 female) participated in this experiment for pay. Participants had normal or corrected-to-normal vision and no tactile impairments. None took part in the previous experiment.

3.1.2. Materials and apparatus

Sixteen of the plaster face models and corresponding full-face colour images used in Experiment 1 were randomly selected and used as stimuli in this experiment. All face stimuli were unfamiliar to the participants. Using Adobe Photoshop 5.5 for Macintosh, we divided each visual face image into four parts comprising nose area, eye and brow area, mouth and chin area, and external features (see Farah, Drain, & Tanaka, 1995 for a similar procedure). The visual images were displayed on a Macintosh using Microsoft PowerPoint. Each face part was presented against a black background and was positioned in the correct relative facial location (i.e. the mouth and chin area was presented at the bottom of the display). These part-based face images were used in the learning session of the experiment only. The whole face image was presented during the recognition task.

3.1.3. Design

The experimental design was identical to that of Experiment 1 except that participants were familiarised with only eight target faces in either vision or haptics during learning. During the recognition session, the within modal and crossmodal trials were blocked and each comprised 16 randomised trials: eight target trials and eight non-target trials per block.

3.1.4. Procedure

During the learning phase, each participant learned a set of eight faces through either vision or touch. In the visual learning modality a target face was presented as four face parts and each part was randomly presented for 5 s within a single trial. Target faces were repeated four times during visual learning to ensure equivalent performance between the visual and haptic conditions. Haptic target faces were each presented for 1 min during learning and repeated four times. We set a learning criterion of 7/8 correct responses in both learning modalities before the participant could proceed to the experiment. On average, learning of the haptic stimuli took about 50 min whereas learning of the visual stimuli took approximately 30 min. Visual learning took longer in this experiment than in the previous experiment as participants had to learn each face from a collection of face parts which proved a more difficult task.

The task during the recognition phase was to indicate whether a face was old or new by circling the appropriate answer on a response sheet as quickly and as accurately as possible.

3.2. Results

Although crossmodal performance was poor in this experiment, it was nonetheless greater than chance [$\chi^2 = 72.66$, d.f. = 23, $p < 0.001$]. The mean A' scores for each condition in the recognition task are illustrated in Fig. 2. The A' scores were analysed using a 2×2 mixed design ANOVA with learning modality (vision or haptics) as a between subjects factor and recognition task (within or across modalities) as a repeated measure. We found no effect of learning modality [$F(1, 22) = 2.01$, n.s.]. However, there was a main effect of recognition task [$F(1, 22) = 26.11$, $p < 0.001$] with better performance in the within modal than the crossmodal recognition task. There was no interaction between the factors, [$F(1, 22) = 1.22$, n.s.].

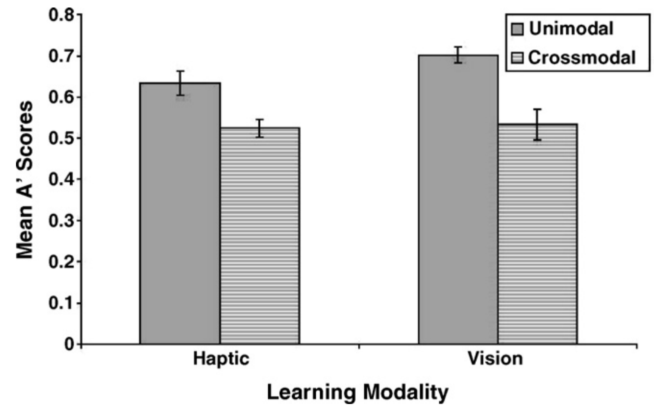

Fig. 2. Plot showing face recognition performance, measured as the mean A' scores, for the within and across modalities conditions in Experiment 2. Error bars represent ± 1 standard error of the mean.

3.3. Discussion

Our results suggest that encoding differences across the modalities did not account for the cost in crossmodal recognition relative to within modality face recognition. Encoding differences across vision and haptics were rendered equivalent by limiting encoding in vision to a feature-by-feature procedure. These results indicate that the cost in crossmodal recognition of faces is due to the nature of the information represented by each modality which is more conducive to within modal than crossmodal recognition.

Faces were, nevertheless, recognised across modalities. Therefore the next question we addressed concerned the nature of the face information that is shared across modalities.

4. Experiment 3

Given that haptic face stimuli are encoded in a serial fashion, it could be assumed that haptic face representations might be more feature-based than configural. However, Kilgour and colleagues (Kilgour, 2003; Kilgour et al., 2004) have shown that configural information can also be represented through touch provided adequate exploration time is permitted. This finding suggests that similar face information can be processed across vision and haptics, however, it is not known whether this information is easily shared across modalities. The following experiment was designed to investigate whether configural face information, or featural information alone, is shared across vision and haptics.

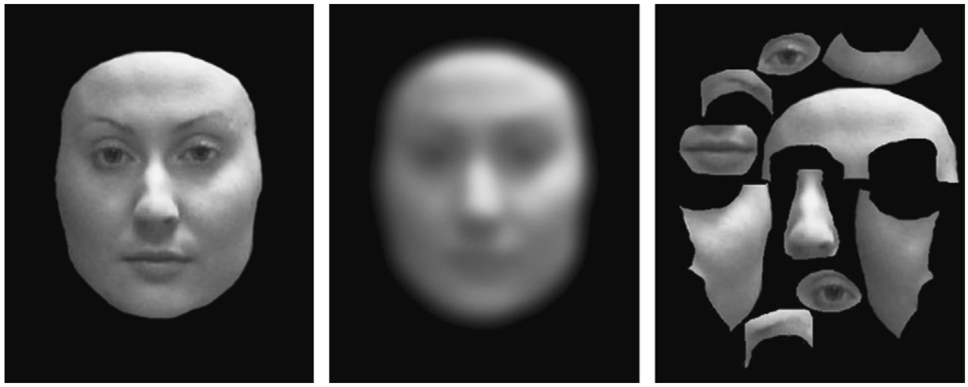

Fig. 3. An illustration of the type of visual face stimuli used in Experiment 3. An example of an intact, blurred, and scrambled image are presented from left to right.

The task in the following experiment was to match a haptic face to a subsequent visual face image. The visual stimuli comprised intact, blurred, and scrambled images (see Fig. 3). Collishaw and Hole (2000) found that recognition performance for scrambled and inverted images was impaired to the same degree in terms of both accuracy and response times, relative to the recognition of upright intact face stimuli. Therefore, in vision, it appears that scrambled images are no more difficult to recognise than inverted images. In both instances recognition is disrupted due to an inability to process appropriate configural face information. In this experiment, we substituted scrambled faces for inverted images, so that featural information in all image types was consistent (i.e. upright). By blurring face images we removed detailed visual information about facial features. The role of featural and configural face information could therefore be studied independently (Collishaw & Hole, 2000; Schwaninger et al., 2002). We hypothesised that if configural face information is shared across modalities then haptic stimuli would be more efficiently matched to intact and blurred images than to scrambled images. If, on the other hand, featural information alone is shared then performance to the blurred images should be reduced relative to the other image types.

4.1. Method

4.1.1. Participants

Twenty-four undergraduate students (mean age 24.29 years, S.D. = 9.08, 18 female) participated in this experiment for research credits. None had participated in the previous experiments. All participants reported normal, or corrected-to-normal vision, and no tactile impairments.

4.1.2. Materials

The models in our experiment were unfamiliar to the participants. Sixteen plaster face models and corresponding colour images were used as face stimuli. The intact images were copied and, using Adobe Photoshop 5.5, we created scrambled and blurred versions of each face. Scrambled stimuli were created by isolating the features and rearranging them into random configurations (see Schwaninger et al., 2002 for a similar procedure). We created blurred images by applying a Gaussian blur of a radius of eight pixels, previously determined in a pilot experiment, to each face image.

The experiment was programmed using SuperLab Pro 1.77. The images subtended a visual angle of approximately 10° vertically and 7° horizontally. Participants responded to each visual or haptic face by pressing the appropriate key on a button box placed just in front of them.

4.1.3. Design

This experiment was based on a one-way, repeated measures design with visual face match (intact, scrambled, and blurred) as the factor. Participants completed a haptics-to-vision, delayed match-to-sample task. Trials were blocked according to each of the three visual image types and participants were informed of the nature of each block. Block order was counterbalanced across participants. Each experimental block comprised 16 crossmodal trials of which there were eight same and eight different pairs of stimuli and trials were randomised across blocks and participants. The participants could take a self-timed break between the experimental blocks. A practice trial was given at the beginning of each block to familiarise the participants with the experimental procedure. Unlike in the previous experiments, responses here were always to visual stimuli thus allowing us to measure performance in terms of both matching A' and reaction times.

4.1.4. Procedure

In any one trial, the participant had to first feel a haptic face stimulus that was presented behind an occluder until a tone sounded 1 min later. The participant was then required to look at the computer monitor. A fixation cross appeared in the centre of the screen for 1 s. Depending on the experimental block, either an intact, scrambled, or blurred face image was presented after fixation. The participants' task was to decide as fast and as accurately as possible whether the visual face image was the same person as, or different to, the haptic face. The visual image remained on the screen until the participant responded. 'Same' or 'different' responses were given by pressing the left or right key on a button box. The position of the 'same' match key was counterbalanced across participants.

4.2. Results

In this experiment we found matching performance (measured as hits and correct rejections) was greater than chance [$\chi^2 = 245.31$, d.f. = 71, $p < 0.01$]. The mean A' scores (±1 standard error of the mean in parentheses) for the intact, blurred and scrambled face types respectively were 0.55 (0.10); 0.56 (0.07) and 0.55 (0.11). The A' data were analysed using a one-way repeated measures Friedman's ANOVA which revealed no effect of image type [$\chi^2 = 2.11$, d.f. = 2, n.s.].

Prior to analysis, response times ±2.5 S.D. from the mean for each participant were removed. In total, 1.9% of responses were excluded. Reaction times to the correct trials only (i.e. hits plus correct rejections) across each of the visual face types are shown in Fig. 4. A one-way repeated measures ANOVA revealed a main effect of image type [$F(2, 46) = 22.43$, $p < 0.01$]. A post hoc, Newman–Keuls analysis revealed that matching times to both intact and blurred images were significantly faster than to

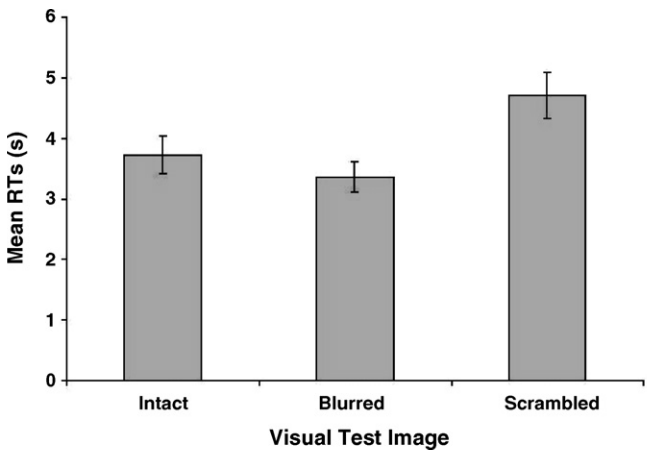

Fig. 4. Plot showing the mean reaction times to the correct responses (hits plus correct rejections) across the matching conditions reflecting the different visual face types. Error bars represent ±1 standard error of the mean.

scrambled images (both $p < 0.01$). There was no difference in response times between the intact and blurred images.

4.3. Discussion

Since matching haptic face stimuli to either intact or blurred visual face stimuli was faster than matching to scrambled face images, our results suggest that configural face information is processed by the haptic system and that configural face information underpins more efficient face matching across the modalities.

The representation of configural face information relies on the encoding of spatial information, and whilst vision is a highly adept spatial processor the precision with which the haptic system processes spatial information is substantially lower than that of vision (Loomis, 1985, 1990; Loomis et al., 1991). Although Kilgour (2003) found evidence suggesting that the haptic system can represent faces configurally, it was not known whether representations based on configural information were shared across modalities. Our results suggest that despite differences in the ability to encode spatial information across vision and haptics, configural face information facilitates crossmodal face matching.

5. General discussion

The aim of our study was to investigate whether the visual and haptic modalities encode similar information to allow for efficient recognition of faces across these modalities. In summary, we found that face information represented in memory is largely modality-specific although some information can transfer across modalities (Experiment 1). We found that these modality specific face representations were maintained even when differences in exploratory procedures during encoding were controlled (Experiment 2). In Experiment 3 we investigated the type of structural descriptions that can be shared across the modalities. Our data indicate that information about the configuration of features mediates more efficient crossmodal face recognition than featural information alone. Taken together these results suggest that face information is processed in a similar way across the visual and haptic modalities, but that qualitative differences in the nature of the information encoded promotes more robust within-modal relative to crossmodal face recognition.

Since both modalities can encode information pertaining to structure (Lederman & Klatzky, 1990) it might be assumed that a common, multisensory representation of face structure might facilitate crossmodal face recognition. Indeed we found that faces could be recognised across modalities indicating that representations of facial structure are common to both modalities. However, crossmodal recognition was poor relative to unimodal recognition, implying that structural information alone is not sufficient for efficient crossmodal face recognition. Indeed, crossmodal recognition performance was just better than chance performance indicating that this task was very difficult for our participants to perform. Our results therefore suggest an important role for modality-specific information during face recognition.

Modality-specific information may include colour or pigmentation information in vision or texture and skin compliance in haptics. Colour information, for example, has been found to interact with shape information such that it facilitates recognition of structurally similar objects (Price & Humphreys, 1989; Tanaka, Weiskopf, & Williams, 2001). Since faces represent a highly homogeneous stimulus category, it is not surprising that variations in colour and pigmentation have been found to affect better face recognition performance (e.g. Bruce & Langton, 1994; Vuong, Peissig, Harrison, & Tarr, 2005; Yip & Sinha, 2002). In our experiments, as in natural face recognition tasks, colour and pigmentation information was included and it is likely that this information provided rich cues for visual recognition. However, if colour was a cue to face recognition then crossmodal face recognition performance might be reduced since colour is specific to vision alone. It is possible that if the visual face images were uniformly coloured (i.e. images of our facemasks) better crossmodal performance may have been achieved. However, we decided against using such images because we wanted to examine the extent to which natural face representations are shared across modalities. The reduction of modality-specific face information in any stimulus set could conceivably occur ad infinitum, at least until within modal and crossmodal recognition performance was equivalent, but this would tell us little about the relative importance of modality-specific information for face recognition.

For haptics, material properties such as surface texture and compliance are found to play an important role in face recognition tasks (Kilgour & Lederman, 2002). In our experiments, such material properties were non-diagnostic of face identification since all our haptic face stimuli were made of the same material. Instead, other haptic-specific cues may have been encoded (possibly relating to the three-dimensional structure of each face stimulus), thus rendering within modal haptic recognition more efficient than crossmodal recognition.

As mentioned earlier, it is argued that faces represent a special class of objects in that certain types of information processing are specific to face perception (e.g. Kanwisher, 2000). For example, inverted faces are more difficult to recognise than inverted objects, suggesting that configural information of facial features is more important for faces than other objects. On the other hand, findings from Diamond and Carey (1986), among others (e.g. Gauthier & Tarr, 1997), suggest that face specific processing, and in particular configural processing of facial features, is a matter of expertise. Although this debate was not the focus of our study, our results may nevertheless lend indirect evidence to the idea that information processing is face-specific. In particular, the results of Experiment 3 suggest that haptics, like vision, encodes information regarding the configuration of facial features. Since haptic face perception is not a task at which most people are expert, it could be argued that configural processing occurred independently of expertise and as such this result reflects face-specific rather than general object information processing per se.

Neuroimaging studies have suggested multisensory interactions in object recognition in general (e.g. Murray, Foxe, & Wylie, 2005), and a benefit for multisensory over unimodal encoding (see e.g. Lehmann & Murray, 2005). However, our pilot study indicated that when faces were encoded bimodally, unimodal visual recognition performance was equivalent to bimodal recognition and significantly better than unimodal haptic recognition. Such results are indicative of visual capture during multisensory encoding. Expertise in visual face perception and face information processing may affect a modality encoding bias, resulting in more efficient processing of face information within this modality. Note that this process may not be the same for object recognition in general since many objects can be highly familiar within the haptic modality. Further research is required to elucidate the role of a modality encoding bias in face recognition and the cortical mechanisms underlying this recognition versus the recognition of other familiar classes of objects.

Finally, our findings are interesting in light of a recent fMRI study investigating areas of cortical activation during the haptic recognition of face stimuli and non-face objects (Pietrini et al., 2004). Although previous neuroimaging studies have found that the fusiform face area (FFA) in the right fusiform gyrus responds preferentially to visually presented faces, possibly mediating visual face recognition (e.g. Haxby, Hoffman, & Gobbini, 2000; Hoffman & Haxby, 2000; Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997; Sergent, Ohta, & McDonald, 1992), haptic face recognition was found to activate a proximal but non-overlapping area of ventral temporal cortex (Pietrini et al., 2004). Furthermore, Kilgour and colleagues reported that haptic face perception activated the fusiform gyrus in the left hemisphere rather than the right which is typically active during visual face recognition (Kilgour, Kitada, Servos, James, & Lederman, 2005). Given the findings of Experiments 1 and 2, it may be possible that separate cortical areas are necessary to cope with the different computational demands associated with the processing of modality-specific visual and haptic face information. However, face structure can be readily encoded by both modalities so the possibility remains that at least this aspect of face recognition may be multisensory and subserved by a single cortical area, perhaps in early visual cortex (e.g. Pietrini et al., 2004).

Acknowledgements

This research was funded by a Higher Education Authority, PRTLI grant, awarded to the Institute of Neuroscience, Trinity College, Dublin, of which FNN is a member.

References

Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology*, *25*, 281–316.

Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognising the sex and identities of faces. *Perception*, *23*, 803–822.

Bruce, V., & Young, A. W. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305–327.

Casey, S. J., & Newell, F. N. (2005). The role of long-term and short-term familiarity in visual and haptic face recognition. *Experimental Brain Research*, *166*(3–4), 583–591.

Collishaw, S. M., & Hole, G. J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception*, *29*, 893–909.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*, 107–117.

Easton, R. D., Greene, A. J., & Srinivas, K. (1997). Transfer between vision and haptics: Memory for 2D patterns and 3D objects. *Psychonomic Bulletin and Review*, *4*, 403–410.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169.

Farah, M. J., Drain, H. M., & Tanaka, J. W. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 628–634.

Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception*, *29*, 159–170.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring the face recognition mechanism. *Vision Research*, *37*, 1673–1682.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–232.

Haxby, J. V., Horowitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., & Grady, C. L. (1994). The functional organisation of human extrastriate cortex: A PET-fCBF study of selective attention to faces and locations. *Journal of Neuroscience*, *14*, 6336–6353.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, *3*(1), 80–84.

James, T. W., James, K. H., Humphrey, G. K., & Goodale, M. A. (2005). Do visual and tactile object representations share the same neural substrate? In M. A. Heller & S. Ballesteros (Eds.), *Touch and blindness: Psychology and neuroscience*. Mahwah, NJ: Lawrence Erlbaum.

Kanwisher, N. (2000). Domain specificity in face recognition. *Nature Neuroscience*, *3*, 759–763.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialised for face perception. *Journal of Neuroscience*, *17*, 4302–4311.

Kemp, R., Pike, G., White, P., & Musselman, A. (1996). Perception and recognition of normal and negative faces: The role of shape-from-shading and pigmentation cues. *Perception*, *25*, 37–52.

Kilgour, A.R. (2003). Haptic face recognition: Behavioural, neuropsychological, and neuroimaging perspectives. Doctoral dissertation, Queen's University at Kingston, Canada. Dissertation Abstracts International, AAT NQ81001.

Kilgour, A. R., de Gelder, B., & Lederman, S. J. (2004). Haptic face recognition and prosopagnosia. *Neuropsychologia*, *42*, 707–712.

Kilgour, A. R., Kitada, R., Servos, P., James, T. W., & Lederman, S. J. (2005). Haptic face identification activates ventral occipital and temporal areas: An fMRI study. *Brain and Cognition*, *59*(3), 246–257.

Kilgour, A. R., & Lederman, S. J. (2002). Face recognition by hand. *Perception and Psychophysics*, *64*, 339–352.

Lederman, S. J., & Klatzky, R. L. (1990). Haptic classification of common objects: Knowledge-driven exploration. *Cognitive Psychology*, *22*, 421–459.

Lehmann, S., & Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination. *Cognitive Brain Research*, *24*, 326–334.

Lewis, M. B., & Glenister, T. E. (2003). A sideways look at configural encoding: Two different effects of face rotation. *Perception*, *32*, 7–14.

Loomis, J. M. (1985). Tactile recognition of raised characters: A parametric study. *Bulletin of the Psychonomic Society*, *23*, 18–20.

Loomis, J. M. (1990). A model of character recognition and legibility. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 106–120.

Loomis, J. M., & Lederman, S. J. (1986). Tactual perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance* (pp. 31/1–31/41). New York: Wiley.

Loomis, J. M., Klatzky, R. L., & Lederman, S. J. (1991). Similarity of tactual and visual picture recognition with limited field of view. *Perception*, *20*, 167–177.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, *9*(5), 605–610.

Murray, M. M., Foxe, J. J., & Wylie, G. R. (2005). The brain uses single trial multisensory memories to discriminate without awareness. *NeuroImage*, *27*, 473–478.

Newell, F. N. (2004). Crossmodal object recognition. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes*. London: MIT Press.

Newell, F. N., Ernst, M. O., Tjan, B. S., & Bülthoff, H. H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological Science*, *12*, 37–42.

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W. H., Cohen, L., et al. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proceedings of the National Academy of Science of the United States of America*, *101*(15), 5658–5663.

Price, C. J., & Humphreys, G. W. (1989). The effects of surface detail on object categorisation and naming. *Quarterly Journal of Experimental Psychology*, *41A*, 797–828.

Reales, J. M., & Ballesteros, S. (1999). Implicit and explicit memory for visual and haptic objects: crossmodal priming depends on structural descriptions. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *25*, 644–663.

Schwaninger, A., Lobmaier, J. S., & Collishaw, S. M. (2002). Role of featural and configurational information in familiar and unfamiliar face recognition. *Lecture Notes in Computer Science*, *2525*, 643–650.

Sergent, J., Ohta, S., & McDonald, B. (1992). Functional neuroanatomy of face and object processing: A positron emission tomography study. *Brain*, *115*, 15–36.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, *46A*, 225–245.

Tanaka, J. W., & Sengco, J. (1997). Features and their configuration in face recognition. *Memory and Cognition*, *25*, 583–592.

Tanaka, J. W., Weiskopf, D., & Williams, P. (2001). The role of colour in high-level vision. *Trends in Cognitive Sciences*, *5*(5), 211–215.

Vuong, Q. C., Peissig, J. J., Harrison, M. C., & Tarr, M. J. (2005). The role of surface pigmentation for recognition revealed by contrast reversal in faces and Greebles. *Vision Research*, *45*, 1213–1223.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141–145.

Yip, A. W., & Sinha, P. (2002). Contribution of colour to face recognition. *Perception*, *31*(8), 995–1003.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configural information in face perception. *Perception*, *16*, 747–759.