



Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

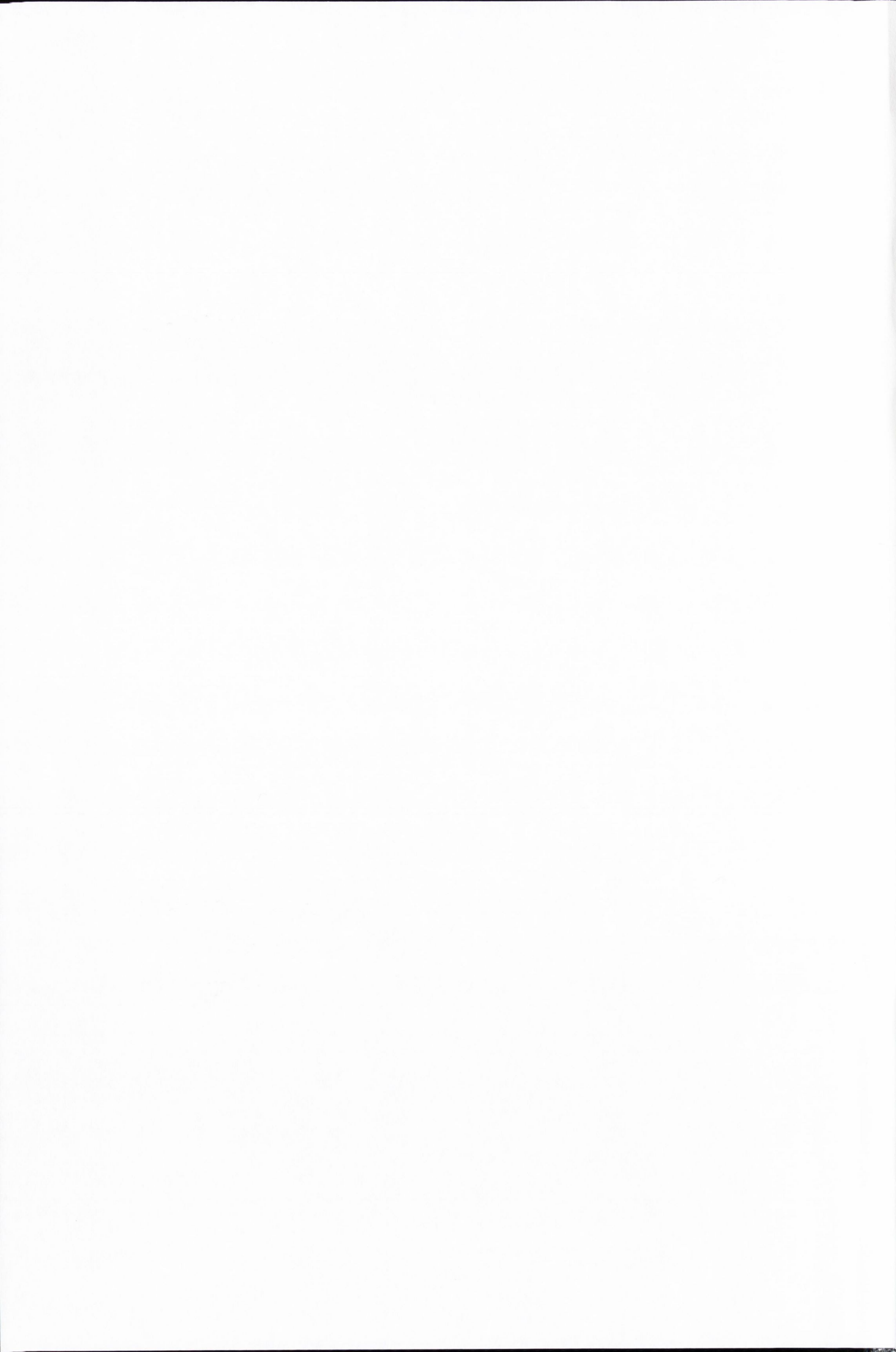
Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.



Crossmodal Face Recognition

by

Sarah Jane Casey

A dissertation submitted for the degree of Doctor of Philosophy of the University of Dublin, Trinity College, Dublin 2, Ireland.

This research was conducted in the School of Psychology and the Institute of Neuroscience.

May, 2006

Summary

Although all faces share the same basic configuration of features, we are capable of distinguishing between, and recognising hundreds, and even thousands of faces with remarkable skill and precision. The perceptual processes, cognitive mechanisms, and neural substrates underpinning visual face recognition have been studied extensively. However, it is now acknowledged that the formation of coherent percepts of objects in the environment benefits from the merging of complementary sources of sensory information about that object. Therefore, researchers have been inspired to venture beyond vision, and examine how information from other sensory modalities can facilitate face recognition.

Recently, it has been discovered that the haptic system is capable of representing and recognising faces. Moreover, one study has provided evidence that information about human faces can be shared between the visual and haptic systems. This thesis has sought to explore further the extent to which face information is shared across the visual and haptic modalities. The primary aims of this thesis were to gain a better insight into visual and haptic processing similarities that can facilitate crossmodal face recognition, as well as to gain insight into the nature of face information that can be represented in memory across vision and touch.

In this thesis I have used a number of simple variations of the basic crossmodal matching and recognition paradigms to investigate crossmodal visual and haptic face recognition. The findings of this body of research demonstrate that face information encoded in one modality can indeed be readily shared with the other modality for the purposes of face recognition. Furthermore, this sharing of face information is affected by face familiarity, and configural face information appears to mediate the efficient sharing of face information across the modalities. The experiments reported in this thesis also provide behavioural evidence to suggest that face recognition is not a completely multisensory phenomenon. It appears that encoded modality-specific face information may play an important role in optimal within modal face recognition. However, the results clearly indicate that enough face information is multisensory to allow successful crossmodal recognition. Although direct evidence of shared memory

representations of face shape was not obtained, the experiments herein do provide indirect evidence to suggest that crossmodal face recognition may be mediated by multisensory representations of face shape. The theoretical implications of these research findings are discussed in Chapter 5.

Crossmodal face recognition is an emerging area of research with huge potential and scope for future work. It is clear that a number of exciting challenges lie ahead for future researchers of crossmodal face recognition.

Acknowledgements

Throughout the course of this thesis I have been fortunate to receive guidance, friendship, support, and assistance from the most wonderful group of people. Therefore I would like to extend my warmest thanks to the following, who have helped me so much over the last few years:

Firstly, I would like to thank my supervisor, Dr. Fiona Newell. She has been a constant source of encouragement and advice. I have benefited greatly from her support and guidance, and under her watchful eye I have developed a number of skills, which will no doubt stand me in good stead for many years to come.

Secondly, I would like to thank my *Viva Voce* Examiners Prof. Susan Lederman and Dr. Hugh Garavan, and Dr. John Davenport, Chair, for providing a challenging yet friendly environment for my thesis defence. They put me at ease from the very first moment which made it one of the most pleasant and rewarding exams I have ever taken!

In addition, I extend my thanks to my appraisers, Drs. Stuart Smith and Hugh Garavan, for their enthusiastic ideas, insights, and advice.

I would like to thank our lab group (Jason, Kylie, Andy, Achille, & Gary) for listening patiently to my various ideas, and providing me with helpful comments and feedback. A special thanks goes to Andy, who had the interesting experience of sharing an office with me. He has been a good friend (even on the very rare occasions when I was grumpy!), and under his tutelage I have become very proficient in Photoshop - so everybody watch out! ☺

Three years ago I embarked on this journey with a little group of fellow explorers. To Fiadhnaít, Clare, Julie, and Michelle, thank you for sharing the ups and downs of PhD life with me.

A big thank you goes to all of the Psychology gang for helping to make my time in Trinity enjoyable and fun. Thank you also to the staff of the School of Psychology, and Institute of Neuroscience, who have always been most generous with their time and resources.

Thanks are also owed to all those who volunteered their faces for the purposes of scientific discovery. Without you this research would not have been possible! Thank you to Tanja Khosrawi for her assistance in creating the haptic face stimuli.

Outside of Trinity I have been blessed with the friendship and support of many friends. Thank you to everyone, but most especially to Claire, Fiadhnaid, Caroline, Orla, Sinéad, Ciara, and Ciara C., for their friendship, loyalty, support, patience, and optimism.

Finally, I would like to thank my family, and my parents Michael and Mary. They have worked tirelessly to support me in every possible way, and for that I am most grateful. Furthermore, they have always believed me capable of achieving my goals, and have instilled in me the motivation and determination to see them through.

This research was funded by a Higher Education Authority, PRTL I grant awarded to the Trinity College Institute of Neuroscience.

Publications arising from the present work

Published Papers

Casey, S.J. & Newell, F.N. (2006). Are representations of unfamiliar faces independent of encoding modality? *Neuropsychologia*. In press. [Chapter 3].

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), 583-589. [Chapter 2].

Casey, S.J. & Newell, F.N. (2003). Haptic own-face recognition. *Proceedings of EuroHaptics 2003*, Dublin, Ireland, 6th-9th July 2003. (pp. 424-429, peer reviewed).

Published Abstracts

Casey, S.J. & Newell, F.N. (2005). Modality-specific and modality-independent processing in face recognition. *The Irish Psychologist*, 32 (2), p.119.

Casey, S.J. & Newell, F.N. (2005). Face information promoting efficient crossmodal matching. *The Irish Psychologist*, 31(9), p.255.

Casey, S.J. & Newell F.N. (2005). Evidence of configural face processing in crossmodal face matching. *Acta Neurobiologiae Experimentalis*, 65, (Supplement), 32.

Casey, S.J. & Newell F.N. (2005). What information promotes efficient crossmodal face matching? *Proceedings of the WorldHaptics Conference*, Pisa, Italy, 18th-20th March 2005.

Casey, S.J. & Newell, F.N. (2004). Familiarity and crossmodal face recognition. *Proceedings of the Touch, Temperature, Pleasure, Pain (TTPP)*, Manchester, UK, 1st-2nd September 2004.

- Casey, S.J. & Newell, F.N. (2004). The role of familiarity in visual and haptic face recognition. Proceedings of the 5th *International Multisensory Research Forum*, Sitges, Barcelona, 2nd-5th June 2004.
- Casey, S.J. & Newell, F.N. (2004). Crossmodal face recognition. Proceedings of the *Experimental Psychology Society London Meeting*, London, UK, 5th-6th January 2004.
- Casey, S.J. & Newell, F.N. (2003). Haptic own-face recognition. Proceedings of the 4th *International Multisensory Research Forum*, Hamilton, Canada, 13th-17th June 2003.

Table of Contents

Declaration	i
Summary	ii
Acknowledgements	iv
Publications	vi
Table of Contents	viii
List of Figures	xi
List of Tables	xiii
Chapter 1: Introduction	I
1.1 Combining the senses	2
1.1.1 <i>Combination and integration of sensory information</i>	3
1.1.2 <i>The perception of objects using vision and haptics</i>	5
1.2 Visual and haptic face recognition	10
1.2.1 <i>Is face recognition special?</i>	11
1.2.2 <i>Configural Processing in object and face recognition</i>	12
1.2.3 <i>Neuropsychological case studies</i>	15
1.2.4 <i>Face-specific neural substrates</i>	16
1.3 Models of face recognition	19
1.3.1 <i>Cognitive models of face recognition</i>	20
1.3.2 <i>A cortical model of face perception</i>	25
1.3.3 <i>Multisensory person recognition</i>	26
1.4 Crossmodal face recognition	30
1.4.1 <i>Familiarity and crossmodal face recognition</i>	30
1.4.2 <i>The representation of faces by vision and haptics</i>	31
1.4.3 <i>Configural processing across modalities</i>	34
1.4.4 <i>Multisensory structural face representations</i>	34
Chapter 2: Familiarity and Crossmodal Face Recognition	36
2.1 Introduction	36
2.2 Experiment 1	39
2.2.1 Method	39
2.2.1.1 <i>Participants</i>	39
2.2.1.2 <i>Apparatus and Stimuli</i>	39
2.2.1.3 <i>Design</i>	41
2.2.1.4 <i>Procedure</i>	41
2.2.2 Results	42
2.2.3 Discussion	43
	viii

2.3	Experiment 2	46
2.3.1	Method	46
2.3.1.1	<i>Participants</i>	46
2.3.1.2	<i>Apparatus and Stimul</i>	47
2.3.1.3	<i>Design</i>	47
2.3.1.4	<i>Procedure</i>	49
2.3.2	Results	49
2.3.3	Discussion	50
2.4	General Discussion	51
2.5	Conclusions	53
Chapter 3:	Are representations of faces independent of encoding modality?	55
3.1	Introduction	55
3.2	Experiment 3	58
3.2.1	Method	58
3.2.1.1	<i>Participants</i>	58
3.2.1.2	<i>Apparatus and Stimuli</i>	58
3.2.1.3	<i>Design</i>	59
3.2.1.4	<i>Procedure</i>	59
3.2.2	Results	60
3.2.3	Discussion	61
3.3	Experiment 4	62
3.3.1	Method	62
3.3.1.1	<i>Participants</i>	62
3.3.1.2	<i>Apparatus and Stimuli</i>	62
3.3.1.3	<i>Design</i>	63
3.3.1.4	<i>Procedure</i>	63
3.3.2	Results	64
3.3.3	Discussion	65
3.4	Experiment 5	65
3.4.1	Methods	67
3.4.1.1	<i>Participants</i>	67
3.4.1.2	<i>Apparatus and Stimuli</i>	67
3.4.1.3	<i>Design</i>	67
3.4.1.4	<i>Procedure</i>	68
3.4.2	Results	68
3.4.3	Discussion	60
3.5	General Discussion	70
3.6	Conclusions	73
Chapter 4:	Priming face recognition across modalities	74

4.1	Introduction	74
4.2	Experiment 6	76
4.2.1	Method	77
4.2.1.1	<i>Participants</i>	77
4.2.1.2	<i>Apparatus and Stimuli</i>	77
4.2.1.3	<i>Design</i>	77
4.2.1.4	<i>Procedure</i>	78
4.2.2	Results	79
4.3	General Discussion	82
4.4	Conclusions	86
Chapter 5:	General Discussion	87
5.1	Summary of findings	87
5.2	Interpretation of the cumulative results	89
5.2.1	<i>Processing similarities facilitating crossmodal face recognition</i>	90
5.2.2	<i>Crossmodal face representations</i>	92
5.3	Implications of results	96
5.3.1	<i>Domain-specificity versus expertise</i>	96
5.3.2	<i>Neural substrates for crossmodal face recognition</i>	97
5.3.3	<i>Models of face recognition</i>	101
5.4	Limitations and directions for future research	102
5.5	Concluding remarks	107
References		108
Appendix A:	Ethical permission	133

List of Figures

Figure 1.1: A graphical depiction of the Bruce and Young (1986) functional model of face recognition.	22
Figure 1.2: An illustration of the IAC model of face recognition, after Bruce and Young (1993).	24
Figure 1.3: The distributed neural system for face perception after Haxby et al. (2000).	26
Figure 1.4: Models of crossmodal face recognition.	29
Figure 1.5: An illustration of the general match-to-sample crossmodal paradigm adopted in this thesis.	30
Figure 1.6: Model of convergence (Millar, 2000).	33
Figure 2.1: A sample of the haptic face stimuli used in all of the experiments in this thesis.	40
Figure 2.2: Mean percentages of correct recognition of one's own face across the haptic and visual conditions.	43
Figure 2.3: Mean percentages of correct 'hit' responses in the crossmodal matching task.	50
Figure 3.1: Plot showing the mean percentage of correct responses for face recognition both within and across modalities.	61
Figure 3.2: An illustration of the type of visual stimuli used in the learning and test phases of Experiment 4.	63
Figure 3.3: Plot showing the mean percentage of correct responses for the within modal and crossmodal recognition of faces.	65
Figure 3.4: An illustration of the type of visual face stimuli used in Experiment 5.	66
Figure 3.5: Plot showing the mean reaction times for the correct responses to each of the visual face types.	69
Figure 4.1: Plot showing the mean percentage of correct 'familiar' responses for the crossmodal and within modal conditions and the correct 'unfamiliar' responses to unfamiliar face stimuli.	81

Figure 4.2: Plot showing the mean reaction times for 'familiar' responses for the crossmodal and within modal conditions and the correct 'unfamiliar' responses to unfamiliar face stimuli. 82

Figure 5.1: Millar's (2000) model of convergence as a model of crossmodal face representation 95

List of Tables

Table 5.1: Talairach coordinates (Talairach and Tournoux, 1988) as reported in the text (mean $\pm$ standard deviation).	100
---	-----

Chapter 1

Introduction

In terms of evolution, the ability to process faces is an important skill, not only for humans, but for other species as well. Faces manifest a vast array of informative signals; they index emotion, facilitate social communication, and are a marker of personal identity. For example, direction of eye gaze is indicative of social attention, facilitating interaction and communication between individuals (Perrett, Rolls, & Caan, 1982). Emotive facial expressions can tell us if a person is agreeable or poses a threat. They also warn if something in the environment might be harmful, for instance a grimace of disgust might indicate that available food has spoiled. Perhaps most importantly, each face also conveys personal identity, allowing us to discern friends from foes. A face can be classified into a variety of categories that correspond to diverse properties of the individual - age, race, sex, etc. Moreover, all faces share the same basic configuration of features, yet we are capable of distinguishing between, and recognising, hundreds or even thousands of faces with remarkable skill and precision. The ability to detect the minute and subtle differences that allow us to identify one face from another is a testament to the formidable capacity of the human brain. This ability appears to be acquired very early during development and does not seem to require formal training or cultivation through special techniques (Sergent and Signoret, 1992).

Although the perceptual processes and cognitive mechanisms underpinning face recognition have been studied extensively, virtually all of these experimental investigations have, until recently, been conducted within the visual domain. However, it is now acknowledged that the formation of a coherent percept of an object in the environment benefits from the merging of complementary sources of sensory information about that object (for instance, Ernst and Bühlhoff, 2004; James, James, Humphrey, & Goodale, 2005; Newell, 2004; Stein and Meredith, 1993). This has inspired a number of researchers to venture beyond vision, and examine how information from other sensory modalities can facilitate face recognition. Recently, it

has been demonstrated that the haptic system^{1.1} is capable of representing and recognising faces (Kilgour, de Gelder, & Lederman, 2004; Kilgour and Lederman, 2002). This thesis is concerned with how faces are processed and recognised across the visual and haptic modalities. Specifically, I am interested in whether the same information is processed for recognition by both the visual and haptic systems, and if so, whether a single common memory representation underpins recognition in both modalities. In this chapter I will set the context for my experimental work by outlining the benefits of crossmodal recognition relative to unimodal recognition, and by providing a brief review of the central findings from visual and haptic face recognition research.

1.1 Combining the senses

Traditional approaches to the study of sensory perception stemmed from a belief that sensory processing is highly modular. Thus, a wealth of literature exists in which the processes of the visual, auditory, and haptic systems have been studied in isolation. Although each sense acquires information that differs in terms of its physical properties (i.e. light waves in vision, sound waves in audition, and mechanical vibrations and kinesthetic input in haptics are each detected as modality-specific sensory stimuli), our experience of objects or events in the environment generally does not comprise a collection of separate sensations (Meredith, 2002). A single event can evoke a number of spatially and temporally congruent sensory stimuli (e.g. a bouncing ball) and these sensory stimuli are processed simultaneously by different sensory receptors, yet we perceive a coherent event. This implies that our experience of the world reflects the combination and integration of multiple sources of sensory information in the brain, resulting in a unified, robust, perception of objects or events (Calvert, Bramner, & Iversen, 1998; Driver and Spence, 2000; Ernst and Bühlhoff, 2004; Stein and Meredith,

^{1.1} The haptic modality (system) is the information processing system that uses both tactile and kinesthetic inputs (Loomis and Lederman, 1986). Tactile, or cutaneous, input is encoded through receptors in the skin and kinesthetic input refers to information about the position and movement of limbs in space encoded through receptors in the tendons, muscles, and joints. The inclusion of the kinesthetic component reflects the importance of active exploration in the haptic perception of objects (Gibson, 1966).

1993). The ability to integrate information across the senses is advantageous as it allows us to be more certain about objects or events in the environment, thus influencing our subsequent responses. For instance, identifying an object in a dimly lit room is difficult and error prone if using vision alone, however, combining the impoverished visual information with information obtained by actively exploring the object using haptics increases the likelihood of accurately identifying the object. In other words, information from both modalities helps to minimise the uncertainty in the information associated with either modality alone, to yield the best approximation of the object (Driver and Spence, 2000; Ernst and Bühlhoff, 2004).

1.1.1 Combination and integration of sensory information

Researchers have turned to investigating the convergence of sensory information in the brain with a view to understanding the mechanisms underpinning the integration of sensory information into coherent percepts. A number of studies investigating brain activity in humans in response to different sensory inputs have found areas of common brain activation when information from different sensory sources relates to the same object or event, such as perceiving speech, or recognising objects using vision and haptics (e.g. Amedi, Malach, Hendler, et al., 2001; Calvert, Campbell, & Brammer, 2000; Sathian, Zangaladze, Hoffman, & Grafton, 1997; James, Humphrey, Gati, et al., 2002; Prather, Votaw, & Sathian, 2004). However, it is not clear whether these areas of activation reflect populations of neurons that respond to multiple types of sensory input, or simply an intermingling of different types of sensory cells within the same cortical area. Animal studies have revealed that architecturally, convergence can occur in three ways (Meredith, 2002; 2004). Areal convergence refers to the convergence of inputs of two or more different modalities in the same brain region, but without information from these modalities terminating on a single neuron. Neuronal convergence on the other hand, refers to the convergence of input from multiple modalities on a single neuron. The resultant multisensory neuron can respond to each modality in turn, and also integrate simultaneous input across the modalities. In this way, multisensory neurons are attuned to features of objects or events encoded through any modality. The final type of convergence is a combination of the first two, and most neural areas related to multisensory perception tend to incorporate both unimodal and multisensory neurons (Meredith, 2002). Feedforward and feedback connections serve to relay information

between unimodal and multimodal areas of cortex (e.g. Clavagnier, Falchier, & Kennedy, 2004; Foxe and Schroeder, 2005).

One can extrapolate from animal models of convergence and integration to explain how humans generate multimodally derived percepts at a very low level through excitatory and inhibitory activation in neuronal circuits (Meredith 2002). However, in many ways it is difficult to understand multisensory qualia simply by investigating patterns of neuronal responses. For instance, does activation in a multisensory neuron underpin a veridical or an illusory percept? Insights into the nature of multisensory perception have come from behavioural studies examining the perception of illusions that rely on input from at least two modalities, for example, the McGurk effect, the ventriloquism illusion, or illusory flashing, as they can be systematically induced and studied in a controlled environment (Bertelson, 1999; McGurk and MacDonald, 1976; Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Munhall, Gribble, Sacco, & Ward, 1996; Shams, Kamitani, & Shimojo, 2002; Shimojo and Shams, 2001). From these studies it is clear that two factors in particular influence multimodal perception, namely the spatial co-location and close temporal proximity of two or more sensory events. If these criteria are met then sensory stimuli are generally perceived as emanating from a common source (e.g. Bertelson and Radeau, 1981). For instance, the ventriloquism illusion arises when the perceived location of a sound shifts towards a visual stimulus at a slightly different position. This illusion is commonly perceived at the cinema; the sound of a voice is perceived as coming from an actor on the screen, when it is actually coming from a speaker mounted on the wall. The temporal proximity of audio and visual stimuli is crucial to the success of this effect, as the effect disappears when the audio-visual asynchrony exceeds approximately 300 ms (Slutsky and Recanzone, 2001).

The ventriloquist and flashing illusions also provide a very good demonstration of sensory capture. Sensory capture refers to the tendency for one sense to dominate over another so as to shape the perception of a multimodally defined event and has largely been studied in terms of the visual capture of other sensory information (e.g. Hay, Pick, & Ikeda, 1965; Rock and Harris, 1967). Capture is thought to occur in the direction of the modality that provides the most accurate information about the event. So, with reference to the ventriloquism effect, vision can bias ('capture') the perceived location of sounds because vision provides more accurate spatial location information (e.g. Welch and Warren, 1980; 1986). In his 'fuzzy logic' model of perception, Massaro

(1999) proposed that sensory information is combined in an optimally efficient manner, that is, one modality can be weighted more heavily in the combination if it produces a greater reduction in uncertainty. In other words, the more informative modality guides perception, which may result in a capture effect. Using psychophysics, Ernst and Banks (2002) expanded on this idea of weighted sensory input. Their data indicated that integration is a bottom-up process and that the direction of sensory capture can be predicted according to the informativeness of the encoded information. They proposed that the nervous system combines information (in this instance visual and haptic information) in a statistically optimal manner similar to a maximum likelihood integrator. Thus, in a task wherein a judgement must be made about the height of an object, vision dominates perception when the variance associated with the noise in the visual signal associated with height is less than that associated with the noise in the haptic estimation of height. This statistical account of integration has become known as the maximum likelihood estimation (MLE) model, and has been used to account for other multisensory phenomena (e.g. Bresciani, Ernst, Drewing et al., 2005).

Although it is possible to combine sensory information from a number of senses when perceiving objects and environmental events, the visual and haptic modalities are particularly compatible for 2- and 3-dimensional object recognition. This compatibility stems from the fact that they are the only modalities capable of encoding information about geometrical shape. Therefore, each modality can provide information about the shape of an object, as well as contributing unique sensory information, to provide a rich description of the object for recognition or subsequent action (e.g. taking more care when handling the blue glass bauble than when handling the red wooden one). Indeed, Ernst and Bühlhoff (2004) distinguished between 'sensory combination' and 'sensory integration' processes during multisensory perception. Sensory combination refers to the interactions between sensory information that are modality specific (e.g. colour in vision, or texture in haptics) to provide a richer description of an object than could be gained through either modality alone. Sensory integration, on the other hand, refers to the interactions between redundant signals from both modalities (e.g. the geometrical structure of the object) to provide the best approximation of object shape. This bimodal redundant information is that which may be integrated according to the MLE model.

1.1.2 The perception of objects using vision and haptics

It has been known for some time now that the haptic system is capable of recognising 3-dimensional objects with ease and speed (Klatzky, Lederman, & Metzger, 1985). As in vision, shape information is a diagnostic property for object recognition (Lederman and Klatzky, 1990). However, there are also a number of differences between the visual and haptic modalities in terms of how they respectively process object related sensory input, and this may come to bear on how bimodal visual-haptic percepts are generated. Of course such disparities need not necessarily have a negative impact on object perception and recognition, indeed they often result in a richer, more robust object percept.

Although fine-grained object information is obtained by rapidly scanning across the object's surface in a series of foveal saccades (e.g. Yarbus, 1967), the visual system is capable of simultaneously carrying out a coarser analysis using peripheral receptors in the retina, thus vision is capable of encoding the global structure of an object quickly and easily. In contrast, unless an object falls within the scope of the fingertip, the haptic system must encode an object in a serial fashion over time, and integrate this serially acquired information to form a coherent object percept, placing a considerable demand on working memory (Loomis and Lederman, 1986)^{1,2}. Furthermore, the haptic system appears to employ a series of specialised exploratory procedures (EPs) to encode object information (for a more complete description refer to Lederman and Klatzky, 1987). Each EP is associated with a particular object property, and if the appropriate EP is not performed, then the associated property is of minimal use for recognition (e.g. Lederman, Klatzky, & Reed, 1993). EPs differ in terms of measurement precision and the speed in which data is acquired, for instance, there are two EPs used for encoding shape information: enclosure, and contour following. Enclosure is used to gain a quick but rough estimate of shape and size, and is executed by moulding the hands to the surface of the object. A more detailed description of shape can be obtained through contour following, that is, tracing the contour of the object with the finger(s), and

^{1,2} One study does suggest that objects can be identified from a "haptic glance" (Klatzky and Lederman, 1999). A haptic glance is defined as a brief, spatially constrained haptic exposure involving little or no movement of the fingers (akin to a brief enclosure EP). However, the notion of a haptic glance was tested only for objects differing at the basic-level (e.g. paperclip vs. button, bowl vs. plate). It is not clear how informative a haptic glance might be for the recognition of highly similar objects differing in terms of subordinate level information (i.e. faces).

although more precise, this EP requires the integration of local feature information over time.

While the geometrical properties of objects are important for recognition in both vision and haptics, the processing of the material properties of objects (e.g. surface texture, compliance, and temperature) has also been shown to be important for haptic object recognition (e.g. Klatzky et al., 1985; 1993; Klatzky, Lederman, & Reed, 1987). Indeed, each material property has its own specific EP (Lederman and Klatzky, 1987). For instance, Klatzky, Loomis, Lederman, et al. (1993) found that recognition performance, as a result of exploring objects through heavy gloves, was significantly enhanced when the tips of the gloves were cut off to expose the fingers to the object's surface texture, divined through 'lateral motion' of the fingers across the surface.

The differences in the spatial bandwidth of sensory receptors in vision and haptics suggest that a modality encoding bias exists, with vision preferring to encode more global shape properties, and haptics preferring local shape information. However, although a greater importance for local features is evident early in haptic object processing, global features become more important as processing time increases (Lakatos and Marks, 1999). Also, when the cognitive demands of haptic perception are contrived in vision by restricting the visual field of view so that object properties must be encoded in a serial manner, recognition performance in vision has been found to be similar to haptic performance (Loomis, Klatzky, & Lederman, 1991). Thus, despite different receptor capacities, the visual and haptic modalities appear to be functionally equivalent in terms of representing and recognising objects.

There is some debate as to how objects, including faces, are actually represented in the brain. One account posits that representations are viewpoint-independent (e.g. Biederman and Gerhardstein, 1993; Marr, 1982), such that they are stored as structural descriptions of objects similar to 3D computer models. Recognition of an object is a reconstructive process: it begins by extracting local features and progressively combining them into more complex structural descriptions. However, it is difficult to generate a 3D representation of some (e.g. unfamiliar) objects given their 2D images on the retina. Indeed, even familiar objects are not recognised equally well across all viewpoints (e.g. Bruce, Valentine, & Baddeley, 1987). Therefore, a second account has been proposed which suggests that objects are represented as a collection of view-dependent images (for a review see Tarr and Bülthoff, 1999). In this instance

recognition is based on matching the retinal image to the corresponding stored view of the object. It is not necessary to store every conceivable view, in many cases a single 'canonical' view can contain sufficient information to allow matching to a number of different retinal images, such as the 3/4 view of a face (Bruce et al., 1987).

This thesis is not concerned with the exact nature of the representations underpinning crossmodal face recognition, rather, whether or not information can be shared between the two modalities for recognition^{1.3}. As both vision and haptics can encode the geometrical properties of faces and this information is stored regardless of whether the representation is viewpoint-independent or viewpoint-dependent, it is likely that both modalities share the same representational format. For instance, efficient recognition of objects in both modalities appears to be viewpoint-dependent (e.g. Edelman and Bülthoff, 1992; Harmann and Humphrey, 1999; Newell, Ernst, Tjan, & Bülthoff, 2001; Newell and Findlay, 1997)^{1.4}.

An increasing number of studies are investigating the way in which information from the haptic and visual modalities is integrated to form coherent percepts of objects in the environment. Behavioural evidence has found that crossmodal transfer of shape information can occur between the visual and haptic systems, but that this sharing of information is viewpoint dependent (Newell et al., 2001). Physical limitations on manual exploration predispose the haptic system to encoding the back of objects, whilst the visual system encodes the front. Thus to maximise recognition across modalities, it is necessary to rotate the object so that the test modality encodes the object from its preferred view. However, in terms of representing a whole object in memory, both modalities provide complementary information that, when combined, serves to increase the information content regarding that object, allowing the formation of a richer object representation (cf. Ernst and Bülthoff, 2004).

Given that information regarding object shape is necessary for object recognition in both vision and haptics, and that geometric information encoded in one modality can

^{1.3} Although the term 'structural' is typically associated with the viewpoint-independent model, in this thesis the words structural and structure are used to connote the geometrical shape of objects and faces, which can be represented regardless of whether the underlying representation is 3-dimensional or image-based.

^{1.4} Viewpoint-dependence for haptically explored objects occurs for those objects that are fixed into position. It is of course possible that for palpable objects, haptic representations are omnidirectional.

be utilised for recognition in the other, it is possible that the two modalities share the same underlying representation of object shape. A number of studies have investigated the existence of common geometrical object representations using a crossmodal priming paradigm (Easton, Greene, & Srinivas, 1997; Reales and Ballesteros, 1999). Priming refers to the facilitation of a response (in terms of accuracy or speed) to a stimulus to which an individual has recently been exposed (Tulving and Schacter, 1990). In a crossmodal priming task objects are primed in one modality and then identified in another modality. Evidence from the visual literature suggests that object priming is mediated by structural descriptions of objects (e.g. Cave, Bost, & Cobb, 1996; Cooper, Schacter, Ballesteros, & Moore, 1992). Therefore, if visual and haptic object recognition is underpinned by shared structural object representations, crossmodal priming effects should be similar in magnitude to within modal priming effects. Data supporting this prediction has been provided by both Easton et al. (1997) and Reales and Ballesteros (1999). It could be argued that such shared representations comprise redundant shape information encoded by vision and haptics.

In a natural progression, research has turned to identifying the neural substrates that might underpin visual and haptic object recognition. The question arises as to whether a single brain area might subserve object recognition in both modalities. A number of functional magnetic resonance imaging (fMRI) studies have identified the lateral occipital complex (LOC) as a possible substrate of not only visual, but also crossmodal object recognition. Previously implicated in visual object recognition (Grill-Spector, Kourtzi, & Kanwisher, 2001; Kanwisher, Chun, McDermott, & Ledden, 1996; Malach, Reppas, Benson, et al., 1995), a region of the LOC has been found to respond when participants recognise objects through touch (Amedi, Malach, Pender, et al., 2001; Grill-Spector, Kushnir, Hendler, et al., 1998; Grill-Spector, Kourtzi, & Kanwisher, 2001; Lerner, Hendler, Harel, & Malach, 2001). Also, haptics-to-visual priming of novel objects has been found to elicit activation in the LOC that is equivalent to that of vision-to-visual priming of the same objects (James, Humphrey, Gati, et al., 2002). Furthermore, evidence suggests that haptically induced activation of the LOC is not due to verbal mediation or visual imagery (see James et al., 2005). These studies suggest that visual and haptic object representations share the same neural substrate.

In summary, it is clear from the evidence outlined above that the visual and haptic systems are both capable of representing 3-dimensional information about general

objects, and, what's more, this information can be shared efficiently between the two modalities. This sharing is possibly mediated by a common object representation subserved by shared neural substrates.

1.2 Visual and haptic face recognition

It is argued that one advantage of generating multisensory object percepts is that it allows us to use sensory information interchangeably so that object recognition can be maintained should one sense be lost (Calvert, 2001). Given that a face constitutes an ecologically important stimulus, the question arises as to whether person recognition can occur outside of the visual domain. It seems likely that voice information can serve as a marker of personal identity in the absence of vision. Visual and auditory information for person identification generally occur in close temporal proximity, that is, auditory speech and face movements occur as temporally synchronous information inputs. Therefore, this information is likely to be integrated and generate a robust multisensory representation of the speaker. In contrast, visual and haptic face information do not automatically co-occur in the real world, therefore it may take substantial learning to associate these two types of face information. On the other hand, the visual and haptic systems are both capable of processing shape information, therefore the haptic system may well be capable of recognising faces and representing face information in a format conducive to crossmodal face recognition.

The first thorough investigation of haptic face recognition was conducted by Kilgour and Lederman (2002). In a matching-to-sample task, blindfolded participants were required to match a haptically explored face to one of three exemplars. Matching accuracy was surprisingly accurate, and performance in the haptic alone condition was as good as when the target stimulus was initially explored using vision and haptics together. In addition, performance accuracy declined when participants were tested on clay models of the live faces indicating that, as for non-face objects, the encoding of material properties of faces (e.g. skin texture and compliance) is necessary in addition to shape for optimal haptic face recognition. Crossmodal face matching was also assayed and, irrespective of learning modality, crossmodal matching performance was better than chance. This study thus demonstrated that the haptic system is indeed capable of representing and recognising faces, and moreover, that face information can be shared between the visual and haptic modalities for matching purposes. This study raised a

number of new, interesting questions, specifically, how are faces processed by the haptic system, and are these processes similar to those underpinning visual face recognition? Subsequent studies have drawn on the visual face recognition literature to inform both their conceptual and methodological approach to investigating haptic face recognition.

1.2.1 Is face recognition special?

One debate that recurs time and again throughout visual face recognition literature centres on the notion that faces constitute a 'special' class of objects. On the one hand, face processing can be considered domain-specific, that is, faces are processed by mechanisms separate from those mediating the recognition of all other objects (e.g. Bentin and Deouell, 2000; Kanwisher, 2000; Kanwisher, McDermott, & Chun, 1997; Moscovitch, Winocur, & Behrmann, 1997). On the other hand, face recognition may be considered a function of expertise, so that processes that appear to be specific to faces are actually employed in the processing of any object class with which an individual is expert (Diamond and Carey, 1986; Gauthier and Tarr, 1997; Gauthier, Williams, & Tarr, 1998; Gauthier, Tarr, Moylan, et al., 2000a). Evidence from behavioural, neuropsychological, and imaging studies has been gathered in support of each perspective. A detailed account of this debate is beyond the scope of this thesis. It is sufficient to note that regardless of whether face recognition is innately specified or due to expertise, it is clear that face processing is facilitated by configural face information, is orientation sensitive, and mediated primarily by cortical structures in the right hemisphere.

Although the haptic system is much practised in terms of processing and recognising common objects, for instance when moving through a darkened room or searching for an item at the bottom of a bag, it is generally not used for recognising faces. Therefore investigations of haptic face recognition prove interesting in light of the specificity debate. Given that face recognition is non-expert in the haptic modality, if haptic face recognition involves similar processes and utilises the same neural substrate as visual face recognition, then it could be argued that face recognition processes are indeed underpinned by a domain-specific face mechanism. An overview of the general findings concerning visual face processing is provided in the following sections, and these findings are discussed in relation to recent evidence from haptic face recognition studies.

1.2.2 Configural processing in object and face recognition

Biederman (1987) proposed that objects are represented and recognised in terms of their constituent parts (termed "geons"), and that the spatial arrangement of those parts is unimportant. Although featural information is important for face recognition (e.g. Collishaw and Hole, 2000; Schwaninger, Lobmaier, & Collishaw, 2002), we appear to have a preference for processing faces in terms of configural information^{1.5}. For instance, in a study by Young, Hellawell, and Hay (1987), the top and bottom halves of different famous faces were combined and participants were presented with the new combinations in a face recognition task. They found that when the two halves were closely aligned participants had great difficulty in naming the famous individuals to whom the top halves of the composites belonged. However, identity performance for each face half was much better when the halves were not closely aligned. This finding supports a configural processing strategy, since by closely aligning the two halves of the faces, this effectively created a 'new' face with different configural relationships which interfered with the ability to identify the top halves of the original faces.

For adults, the recognition of individual faces is more affected by inversion than the recognition of any other type of object (Yin, 1969, 1970; Diamond and Carey, 1986). The 'face inversion effect' refers to the impaired ability to recognise inverted faces relative to other inverted non-face objects or upright faces, and the effect manifests as reduced accuracy scores and increased response latencies. A plethora of experiments using varied methodologies have demonstrated that inversion disrupts configural face processing (e.g. Bartlett and Searcy, 1993; Diamond and Carey, 1986; Freire, Lee, & Symmons, 2000; Leder and Bruce, 2000; Lewis and Glenister, 2003), as when

^{1.5} Three different types of configural processing have been distinguished (Maurer, Le Grand, & Mondloch, 2002). Faces are typically distinguished from other objects on the basis of first-order configural relations, that is all faces comprise two eyes above a nose, above a mouth. Differentiation of individuals therefore requires the processing of second-order configural relations or the spatial distances among facial features (Diamond and Carey, 1986). The final type of configural processing is termed holistic processing. Generally, facial features and the spatial relations among them are processed interactively (i.e. faces are processed as a gestalt) (Hole, 1994; Tanaka and Farah, 1993; Tanaka and Sengco, 1997; Young et al., 1987). In this thesis 'configural' refers to the spatial arrangement of and relations between discrete facial features.

individuals are presented with inverted faces they can readily detect manipulations affecting recognition at the featural level, but they are largely insensitive to changes in configural information. For instance, the Thatcher Illusion (Thompson, 1980) is often used to discern the occurrence of configural processing. In this illusion, configural face information is manipulated such that the eyes and the mouth are inverted giving the face a grotesque appearance. However, when inverted, the face no longer looks abnormal, presumably due to the disruption of configural processing. This illusion has also been used to demonstrate that configural information is gradually degraded during the rotation of a face from upright to inverted (Lewis, 2001; Lewis and Glenister, 2003). Although it is not clear to what extent configural information is disrupted (e.g. Murray, 2004), the processing of featural information does not appear to be particularly affected by inversion (Valentine, 1988). As inversion does not usually affect the recognition of non-face objects (presumably because the processing of objects is thought to be feature-based, e.g. Biederman, 1987), configural processing and the inversion effect are largely accepted as indicators of a face-specific mechanism (Moscovitch, Winocur, & Behrmann, 1997). However, as a small number of studies have also found an effect of inversion for mono-oriented non-face stimuli such as cars, even for non-experts (e.g. Gauthier, Skudlarski, Gore, & Anderson, 2000b), it is important to note that the inversion effect by itself cannot be taken as conclusive evidence of face-specificity, but rather must be interpreted in unison with evidence from neuropsychological and imaging studies, and behavioural studies using alternative paradigms, such as blurring, for investigating configural face processing (Maurer et al., 2002).

Expertise with visual objects leads to well established behavioural effects that are similar to those obtained with faces, especially in the use of configural cues for recognition (Diamond and Carey, 1986; Gauthier and Tarr, 1997; Gauthier et al., 1998; 2000a). For instance, a novel set of stimuli were created by Gauthier and her colleagues, named 'greebles', which were constrained to be similar to faces along several dimensions (e.g. canonical orientation, bilateral symmetry, etc.). Participants were trained to become expert at recognising individual greebles, and their performance in recognition tasks was compared to that of novices who had no experience with the greeble stimuli. It was demonstrated that extensive practice with greebles could lead to some of the recognition effects, especially that of inversion (Gauthier and Tarr, 1997; Gauthier et al., 1998).

It seems from the evidence described that the processing of configural information is important for face processing in vision, regardless of whether it arises due to domain-specificity or expertise. The question for haptic researchers was whether the haptic system processes faces in terms of featural or configural information. There are a number of reasons why the haptic system might be expected to use featural rather than configural information for face processing. Firstly, face stimuli fall outside the area confined to the fingertip, thus haptic face encoding must proceed in a serial fashion over time, and this may influence the manner in which faces are represented in memory. For instance, visual face stimuli that were encoded in a piecemeal manner were found to be less prone to the inversion effect than faces encoded holistically, implying a shift towards more featural processing for these stimuli (Farah, Drain, & Tanaka, 1995). This suggests that haptic face processing may rely more on individual face features for identification due to encoding differences across the modalities. Secondly, the haptic modality is non-expert at processing faces therefore a demonstration of the behavioural effects (i.e. configural processing and the inversion effect) associated with expert object recognition should not be obtained.

The inversion paradigm has been used to assay the capacity for configural face processing in haptic face recognition studies (Kilgour, 2003; Kilgour, de Gelder, & Lederman, 2004). In haptic face matching tasks configural face information was manipulated by presenting haptic face stimuli upside-down. Unexpectedly it was found that matching performance for inverted facemasks was significantly poorer than that for upright facemasks, indicating a haptic face inversion effect similar to that obtained in vision. Moreover, haptic matching performance for inverted non-face objects (clay teapots differing only in terms of second-order relational information) did not exhibit an inversion effect relative to upright objects. Given a lack of expertise in haptic face processing, these results lend credence to the domain-specificity account of face recognition. The findings imply that the haptic system also appears to dissociate face and object recognition in using configural processing more for faces than for non-face objects. Of course, as in vision, the demonstration of a haptic face inversion effect is not in itself sufficient evidence of configural processing and face-specificity in the haptic modality (cf. Maurer et al., 2002). Thus researchers extended their investigations of haptic face recognition, incorporating neuropsychological case studies and brain imaging techniques to substantiate their findings using the inversion paradigm.

1.2.3 Neuropsychological case studies

In neuropsychological research, double dissociations of function after focal brain damage offers the strongest evidence that face and object recognition are mediated by separate mechanisms and distinct neural pathways. Prosopagnosia is a neuropsychological disorder and refers to the inability to recognise familiar faces following cerebral injury, usually to the right temporal lobe (Young, 1998). Prosopagnosics can distinguish faces as a category quite easily from other visual objects but are unable to identify specific faces (Bodamer, 1947; Bruyer, Laterre, Seron, et al., 1983). Although a number of prosopagnosics have demonstrated a loss of face processing capabilities, they have been found to have intact object processing abilities (Farah, Levinson, & Klein, 1995; McNeil and Warrington, 1993). In contrast, other studies have reported incidents of object agnosia, where patients who are perfectly able to identify faces have great difficulty recognising common objects (Moscovitch et al., 1997). This indicates that the two processing systems are functionally distinct and anatomically segregated; focal brain damage can selectively damage either one whilst leaving processing intact in the other system (Farah, Rabinowitz, Quinn and Liu, 2000).

As prosopagnosics can distinguish faces correctly without being able to identify them, it is argued that prosopagnosia reflects an inability to process faces configurally beyond the level of first-order relational information (that is, a face comprises eyes above a nose, above a mouth) (e.g. Moscovitch et al., 1997). Therefore it might be assumed that prosopagnosics perform to the same degree of accuracy with upright and inverted face stimuli. However, studies involving prosopagnosic LH have found that he actually tends to show a face inversion superiority effect, that is, he processes inverted faces better than upright faces (de Gelder, Bachoud- Levi, & Degos, 1998; de Gelder and Rouw, 2000; Farah, Wilson, Drain, & Tanaka, 1995). This suggests that configural processing, though disrupted, is still present to a lesser degree and interferes with identification-by-features in upright faces. Thus when configural processing is removed through inversion, a feature-based analysis can be conducted more easily (de Gelder and Rouw, 2000).

A recent study assessed whether LH might retain the ability to discriminate faces through touch (Kilgour, de Gelder, & Lederman, 2004). Despite intact sensorimotor hand function, LH performed at chance when matching two upright faces explored

haptically, and his performance was significantly worse than that of neurologically intact controls. However, LH's matching of upright non-face objects (clay teapots) was significantly greater than chance, and neither he nor the controls demonstrated an inversion effect with the non-face objects. Interestingly, as in vision, LH was found to demonstrate a paradoxical inversion effect in the haptic domain. His matching performance for inverted faces was significantly above chance. The results of this study are important for a number of reasons. Firstly, it appears to be the first documented case of haptic prosopagnosia. Secondly, the results have implications as to the neural substrate underpinning visual and haptic face recognition. It appears that damage to the cortex that precluded face processing in the visual modality may also have resulted in a deficit in face processing in the haptic modality. Moreover, LH's haptic deficit mirrored his visual deficit in terms of inversion superiority effects being present in both modalities. These findings suggest that visual and haptic face processing may be underpinned by the same area of cortex. Finally, LH's intact ability to haptically discriminate between both inverted and upright non-face objects suggests that a dissociation might exist between face and object processing in the haptic modality as has been evidenced in vision.

1.2.4 Face-specific neural substrates

Prosopagnosic case studies have indicated that face processing deficits typically occur after damage to the right hemisphere. A number of imaging techniques have been used in an attempt to pinpoint a more circumscribed region of cortex in the right temporal lobe that might act as neural substrate for face recognition in neurologically intact individuals. These include positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and the recording of event-related potentials (ERPs) from the scalp.

A study investigating event-related potentials (ERPs) to face stimuli in normally sighted individuals recorded a negative potential peaking at 170 ms from the stimulus onset (N170) (Bentin, Allison, Perez, et al., 1996). The N170 responded preferentially to human faces and isolated human eyes, but not to human hands, animal faces, furniture, cars, and a range of other non-face stimuli. The N170 was distributed over a relatively bounded region of the temporal lobes, and was greater at right than at left hemisphere sites. Although electrophysiological techniques have a temporal resolution in the order

of milliseconds, the spatial resolution is poor, so PET and fMRI studies have been designed and executed in order to identify specific regions of activation in response to face stimuli. Several studies have shown that the cortical region in which most participants show a significantly greater activation for visually presented faces than objects is the right fusiform gyrus of the inferior temporal lobe (Haxby, Horowitz, Ungerleider, et al., 1994; Kanwisher, McDermott, & Chun, 1997; Puce, Allison, Gore, & McCarthy, 1995; McCarthy, Puce, Gore, & Allison, 1997; Sergent, Ohta, & McDonald, 1992), coinciding with the location of greater N170 activity. One of the seminal works in this area is an fMRI study conducted by Kanwisher, McDermott and Chun (1997), demonstrating strong evidence of cortical specialisation in the fusiform gyrus obtained by testing the responsiveness of the same region of cortex with many different stimulus comparisons^{1.6}. The right fusiform gyrus region does not simply respond to any animal or human images of body parts, but instead, generalises to respond to images of faces taken from different viewpoints. Kanwisher and her colleagues have dubbed this area the "fusiform face area (FFA)". A number of subsequent studies have implicated the FFA in the recognition of facial identity (e.g. Hoffman and Haxby, 2000; Sergent et al., 1992).

A recent fMRI experiment demonstrated that when participants are experts with "greebles", identity matching of upright minus inverted greebles, as well as passive viewing of greebles minus objects, both produce focal activation in the FFA, comparable in spatial extent to that obtained with faces (Gauthier et al., 2000a). So, the combination of categorisation level and expertise appears to provide a plausible explanation for the strong focal activation obtained when faces are compared to non-face objects. Therefore it is possible that there is nothing unique about faces activating the fusiform gyrus as many other tasks also do so, indeed this suggests that the FFA subserves identification at the subordinate level rather than at the category level (Gauthier et al., 2000a, 2000b). Faces may elicit more activation simply because of greater processing experience.

The prosopagnosic patient LH, whose face processing deficit has been found to extend into the haptic modality, sustained damage to the right temporal lobe, among other areas (Etcoff, Freeman, & Cave, 1991; Kilgour et al., 2004). It may be possible

^{1.6} The area of the right fusiform gyrus maximally responding to faces in this study is represented by the Talairach coordinates (Talairach and Tournoux, 1988) x40, y-55, z-10.

that the cortex in the right temporal lobe subserving visual face recognition, specifically the FFA, also facilitates haptic face processing. To date, only two fMRI studies have examined cortical activation for haptically explored face stimuli (Kilgour, Kitada, Sevros et al., 2005; Pietrini, Furey, Riccardi et al., 2004).

Although Pietrini and colleagues did not investigate haptic face processing *per se*, they were interested in whether patterns of category-related cortical activation for objects in visual extrastriate cortex were specific to the visual modality or represented multisensory object form, and whether visual experience was necessary for such category-related activations to occur. Their fMRI results indicated that visual experience was not necessary for the development of category-related responses in the ventral visual stream as category related activation was also found for haptically explored objects and facemasks in blind individuals. For both sighted and blind participants, visual and haptic patterns of cortical activity elicited by bottles and shoes clearly differed from patterns of activation evoked by face stimuli (images or facemasks), but were difficult to distinguish from each other. Also, in sighted participants, tactile recognition of objects evoked patterns of response in the ventral visual pathway that were closely related to the responses evoked during visual recognition of the same objects, indicating that representations of manmade objects in visual extrastriate cortex may indeed be multisensory (see also e.g. Amedi et al., 2001; James et al., 2002). Contrary to expectations, haptic recognition of facemasks did not evoke responses in the same area of inferior temporal cortex as visual recognition. On the other hand, collapsing across all conditions, earlier visual areas in the occipital cortex were activated during haptic recognition, indicating that perhaps only higher-order processing occurs in different areas during haptic and visual face recognition. For instance, it has been suggested that the posterior face regions that fall within the region of LOC may serve to encode the structure of faces, whereas more anterior parts of the FFA subserve face recognition (e.g. Courtney, Ungerleider, Keil, & Haxby, 1997; Sergent et al., 1992).

A subsequent study by Kilgour et al. specifically investigated haptic face recognition in normally sighted participants. Prior to imaging, participants were trained for up to 12 hours to recognise a set of facemasks and a set of control objects that were of similar size and complexity, with 100% accuracy. During scanning, participants felt each stimulus and completed a silent identification task designed to encourage the haptic

processing of the facemasks explicitly as faces (i.e. it required processing of face information at the subordinate level, which would presumably encourage configural rather than just featural processing). Typically imaging studies report face-specific activation in the right fusiform gyrus. In contrast, this study found that haptic face recognition activated the left fusiform gyrus, right hippocampal/ parahippocampal regions, left cingulate, left inferior parietal lobe and clusters in the right cerebellum.

As acknowledged by the authors, it is not possible to completely dissociate activation due to memory retrieval and face processing; the activation pattern rather reflects the neural network underpinning haptic face recognition. However, the right hippocampal and parahippocampal regions, the cingulate gyrus, and the cerebellum have previously been implicated in visual face memory (Kim, Andreasen, O'Leary et al., 1999; Leveroni, Seidenberg, Mayer, et al., 2002).

Kilgour et al. proposed that left fusiform activation may indicate the processing of sequentially explored facial features due to serial haptic exploration. It is argued that ventral visual areas mediate integration within objects (Humpherys, 1999a; 1999b). Indeed, it appears that activation of the left fusiform gyrus is associated with integrative processes in visual object recognition, that is, integrating object parts into perceptual wholes (Gerlach, Aaside, Humphreys, et al., 2002). Moreover, this integration process seems to be modulated by stored knowledge about objects. For the highly familiar facemasks it might be assumed that a representation of global shape is stored in memory (cf. Lakatos and Marks, 1999). However, encoding differences between vision and haptics may lead to haptically encoded faces being processed by the left rather than the right fusiform gyrus, in a manner more similar to visual object recognition^{1.7}. In any case, haptic face recognition appears to recruit cortical structures that subserve visual face and object recognition.

1.3 Models of face recognition

The literature outlined thus far indicates that the visual and haptic modalities process faces in a similar manner (i.e. they both appear to utilise configural face

^{1.7} It is noteworthy that, although visual imaging studies emphasise the right fusiform gyrus for face processing due to significantly greater cortical activation (some specifically selecting the right fusiform as a region of interest), many studies do report smaller levels of cortical activity in the left fusiform gyrus.

information, and both are subject to inversion effects), albeit initial imaging studies suggest that these processes are subserved by slightly different cortical areas in each modality. It is now pertinent to ask how faces are recognised, and if possible, at what level haptic face information might be incorporated into extant models of visual face recognition.

1.3.1 Cognitive models of face recognition

It is assumed that people process familiar and unfamiliar faces differently due to the finding of double dissociations for the two types of processes after brain damage to the visual system (Malone, Morris, Kay, & Levin, 1982). Successful recognition requires complete identification and consequently a range of information is required, for example, a visual representation of the person as well as some biographical information about them (i.e. a name or occupational information). Therefore, functional models of face recognition serve primarily to outline the recognition of familiar faces (e.g. Bruce and Young, 1986; H. Ellis, 1987; Hay and Young, 1982). As the name suggests, such models can be broken down into a number of different processing modules, or stores, each of which have distinct functions. These modules can be eliminated, isolated, or independently manipulated during experiments, and face recognition deficits with varying symptoms can arise due to brain damage to different modules (Bruce and Young, 1986).

The most widely cited model in the visual literature is that devised by Bruce and Young (1986). In this model, person recognition is thought to occur independently of the processes involved in the recognition of facial expressions and the analysis of facial speech information (e.g. Campbell, Brookes, de Haan, & Roberts, 1996; Etcoff, 1984; Young, McWeeny, Hay, & Ellis, 1986; but see also Calder and Young, 2005). Figure 1.1 illustrates the sequence of stages involved in face recognition under this model. The further one has to progress through the sequence of stages, the longer it takes to access the desired information about the individual.

When a face is perceived, a perceptual description is formed which characterises that face's appearance. This is known as structural encoding. In order for recognition to occur, a face must activate a visual representation of that face held in memory. These representations are stored in face recognition units (FRUs), with each FRU

corresponding to a particular familiar face. The FRUs comprise structural face information that is sufficiently abstract so as to allow recognition of faces across multiple viewpoints and viewing conditions^{1.8}. Indeed, studies using recognition priming have demonstrated that priming effects for familiar faces are strongest when the prime and test views are the same, but are still present even when different views of the faces are used in the priming and test phases (A. Ellis, Young, Flude, & Hay, 1987).

^{1.8} As mentioned earlier, the exact nature of the representation is not the focus of this thesis. Although originally the FRUs were considered to comprise structural representations (Bruce and Young, 1986), they could also conceivably comprise collections of canonical face views for individuals (i.e. front, profile and 3/4 views). Therefore, these cognitive models serve as a guide as to how the actual process of recognition occurs, rather than being largely concerned with how faces are represented in memory. Thus, the FRUs remain the first point in the recognition process, regardless of whether they comprise abstract 3-dimensional representations or are image-based.

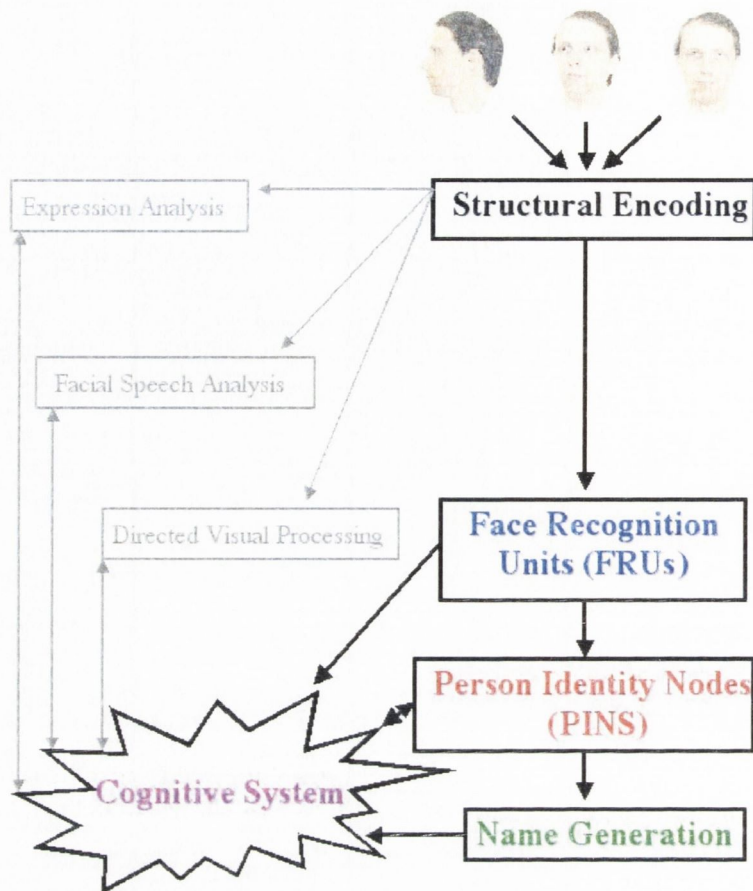


Figure 1.1. A graphical depiction of the Bruce and Young (1986) functional model of face recognition. The modules in bold are involved in the face recognition process.

The person identity nodes (PINs) contain identity-specific information such as semantic information about the person (e.g. occupation, nationality), and can be considered the gateway to person recognition. The PINs can be activated by any person-related information: by faces, voices, or names. When faces or voices are encoded, names are only accessible through the PINs. It is argued that a name cannot be put to a face without having some other information available about the person. For instance, in one diary study, Young, Hay, and Ellis (1985) found evidence of a dissociation between name and semantic information; participants remembered an individual's name only when cued with other autobiographical information, but could remember autobiographical information without the name. Associative (semantic) priming is also

influential on familiar face recognition (Bruce, 1983; Bruce and Valentine, 1986) and is thought to operate via the PINs. The basic effect is that recognition of a famous face (e.g. Hillary Clinton) is faster if it is preceded by an image of a related person (e.g. Bill Clinton), than if it is preceded by an unrelated or unknown face. Whereas repetition priming occurs only in response to perceptually matched person information, that is, face images prime other images but not written names and vice versa (Bruce and Valentine, 1985; A. Ellis et al., 1987; A. Ellis, Flude, Young, & Burton, 1996), associative priming can occur across any domain (Burton, Bruce and Hancock, 1999).

This model has since been developed into the interactive activation competition (IAC) model, adopting a connectionist approach (e.g. Burton and Bruce, 1993; Burton et al., 1999; Burton, Bruce and Johnston, 1990) (see Figure 1.2). The basic modules for face recognition have been retained, with the addition of semantic information units (SIUs). The SIUs code information about known individuals, such as name, occupation, and nationality, and often many of these SIUs will be shared among familiar individuals. This is different from the Bruce and Young model in that both name and biographical information are stored in the SIUs.

Activation passes along the links between the modules and these links are bi-directional. As before, person identity occurs at the PINs, wherein all informational domains of person-related information converge. Unlike in the Bruce and Young model wherein familiarity decisions were made at the level of the FRUs, here familiarity decisions occur at the PINs and can be signalled by a name, face, or semantic information, as well as other information such as voice information (Burton et al., 1999).

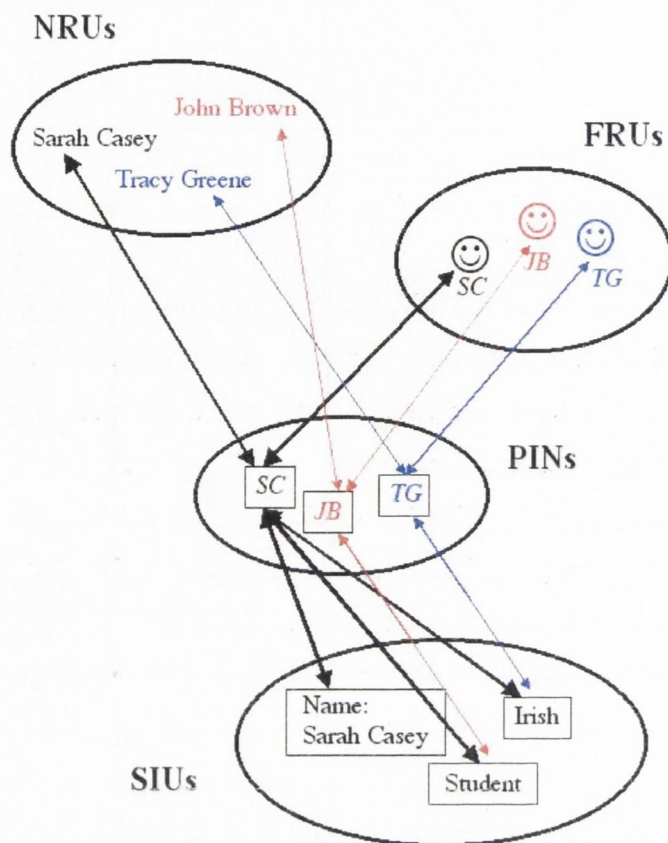


Figure 1.2. An illustration of the IAC model of face recognition, after Burton and Bruce (1993). All domains of person-related information can be seen to converge at the PINs.

The model has been applied to priming effects that have been found with faces. Repetition priming is characterised by a strengthening of a link between a module of information and the PINs. For instance, when a familiar face is encoded its respective FRU and PIN become active, thus the link between the FRU and the PIN becomes strengthened. When that face is subsequently encountered, it takes a shorter time for its PIN to become active. Repetition priming is domain specific (e.g. name priming arises due to the strengthening of links between the NRUs and the PINs). For associative priming, the first face evokes activation through the model to the SIUs, which feed back activation to the PIN of that face and related faces so that they are activated to a level just below threshold. This activation speeds up the familiarity decision for the second face. As PINs are activated by names also, the same priming effect should be found

when a name is followed by a related face. Associative priming thus allows priming to cross domains (faces can prime names and vice versa) and even modalities. For instance, H. Ellis and colleagues found that associative priming facilitated face-voice priming of familiar individuals (H. Ellis, Jones, & Mosdell, 1997).

1.3.2 A cortical model of face perception

A great deal of emphasis has been placed on the fusiform face area (FFA) as the possible neural substrate of visual face recognition. However, the FFA is not the only cortical region found to respond to visual face information. For instance, regions of face-responsive cortex have been located in the lateral inferior occipital gyri (IOG) and in the posterior superior temporal sulcus (STS) (e.g. Hoffman and Haxby, 2000; Kanwisher et al., 1997). The STS is thought to be involved in the perception of facial movement and static images of changeable aspects of the face, such as expression, and the direction of eye gaze (e.g. Calvert, and Campbell, 2003; Hoffman and Haxby, 2000; Perrett and Mistlin, 1990). The IOG abut the lateral fusiform region ventrally and the STS region dorsally, and it is argued that they may input information into both of these cortical areas (Hoffman and Haxby, 2000). These regions may be involved in the early processing of face shape, as the IOG have been found to respond preferentially to images of whole faces, and less so to scrambled face images (see Grill-Spector, Kourtzi, & Kanwisher, 2001). Bearing in mind the cognitive model of Bruce and Young (1986), Haxby and his colleagues have developed a model of the distributed neural system that mediates face perception (Figure 1.3) (Haxby, Hoffman, & Gobbini, 2000). This model, like the cognitive models, is hierarchical in nature, with a core system providing a visual analysis of faces, and an extended system for the further processing of the meaning of the facial information (e.g. semantic information about a person) in collaboration with other neural systems (e.g. memory). Face recognition requires the concerted effort of a number of different regions that perform different components of that function (i.e. the inferior occipital areas, the fusiform gyri, and anterior temporal regions). In this manner, representations of faces are reflected by a distinct pattern of response across a wide area of ventral temporal cortex (Haxby, Gobbini, Furey, et al., 2001).

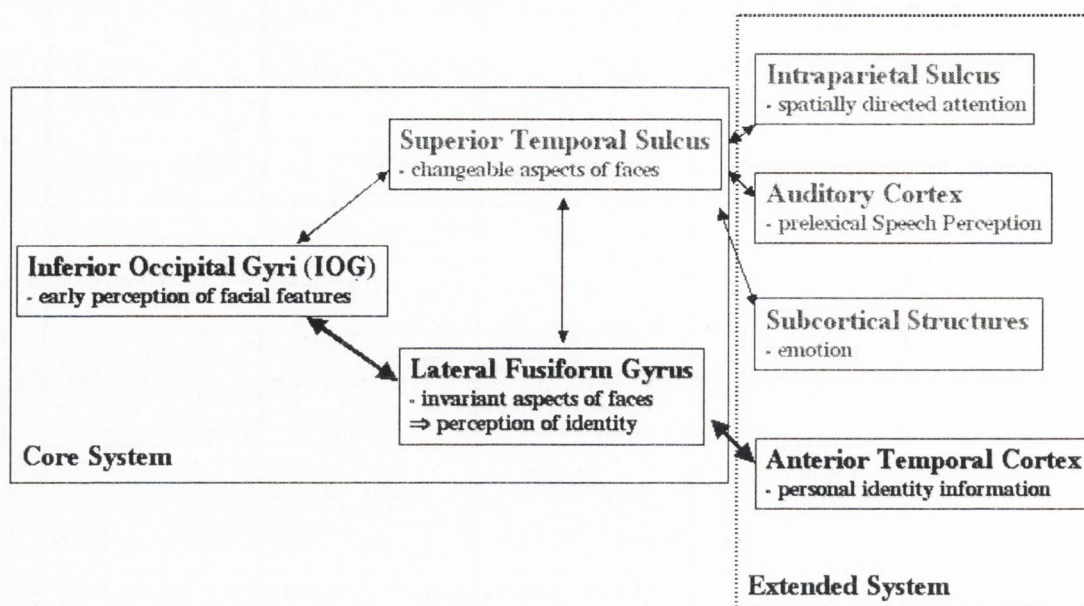


Figure 1.3. The distributed neural system for face perception after Haxby et al. (2000). The core system is concerned with the visual analysis of face information whilst the extended system is involved in the further processing of face information in unison with other neural systems for other cognitive processes. The regions involved in face recognition are represented in heavy bold font.

1.3.3 Multisensory person recognition

There has been little research investigating how the models just described can account for person-related information encoded through other sensory modalities than vision. A number of studies have looked at how voice information contributes to person recognition. Repetition priming has been demonstrated in voice recognition tasks in which participants decided whether a speech sample of neutral content was spoken by a famous or unfamiliar speaker (Schweinberg, Pfütze, & Sommer, 1995). Moreover, priming was found even when the primes and targets were different speech samples of the same person. The priming effects in voice recognition have been attributed to a mechanism similar to that in vision (H. Ellis et al., 1997; Schweinberger, Herholz, & Sommer, 1997). There has been a failure to demonstrate crossmodal repetition priming for faces and voices, presumably because the information is represented differently in both modalities (geometric structure versus phonological information), however associative priming of faces and voices does occur (H. Ellis et al., 1997). These results have implications about how voice information might be represented in terms of the functional models outlined above. For instance, H. Ellis and colleagues have proposed a

separate module of voice recognition units (VRUs), linked to the PINs, which in turn permit crossmodal associative priming of faces. Separate VRUs and FRUs would explain why disorders of visual and auditory person recognition are often dissociated (e.g. Van Lanker, Kreiman, & Cummings, 1982). Indeed, prosopagnosics often use voice information as a compensatory route to person recognition (Pallis, 1955).

A recent fMRI study investigated crossmodal effects for faces and voices, examining whether cortical activity would reflect modality specific modules (i.e. activation of the FRUs and VRUs) and if crossmodal effects would be mediated by separate neural substrates that might reflect the function of the PINs (von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005). Voice recognition elicited the expected responses in the superior temporal sulcus (STS), an area of the cortex previously shown to be involved in voice recognition (e.g. Belin, Zatorre, & Ahad, 2000). However, crossmodal responses to voices of people familiar to the participants was also found in the fusiform face area when the task emphasised speaker recognition over recognition of the verbal content. The area of the fusiform gyrus activated by familiar voices was found to be very close to, and in some cases overlapping, the area that responded to the corresponding face images. Furthermore, an investigation of the functional connectivity between the FFA and the STS showed that the two regions were linked only during familiar speaker recognition. These results are interesting in terms of the functional models of face recognition outlined above. If it is assumed that the fusiform face area is the neural analogue of the FRUs and the superior temporal sulcus is the locus of the VRUs then this study indicates that information may be shared directly between the FRUs and the VRUs without being relayed through the PINs. Therefore, a feeling of familiarity can be shared crossmodally at the sensory level, which is more in keeping with the original Bruce and Young (1986) model of face recognition.

These studies investigating face and voice recognition can inform us about how haptic face information might possibly be represented in models of face recognition (see Figure 1.4 for two different suggestions). Recall that face recognition units based on visual information are domain-specific, that is, they respond only to faces. The geometrical properties of a face remain the same, whether encountered visually or haptically, and in both instances shape properties are the most diagnostic for recognition (Lederman and Klatzky, 1991). Therefore, it is possible that face recognition units might be multisensory (Figure 1.4a), facilitating face recognition regardless of whether

structural encoding occurs through the visual or haptic modalities. Such multisensory FRUs predict successful crossmodal recognition, and therefore account for the ability of participants to successfully match faces across modalities (Kilgour and Lederman, 2002). Moreover, prosopagnosia is thought to arise due to damage to the FRUs (see de Gelder and Rouw, 2000) thus, damage to multisensory FRUs would explain prosopagnosic LH's haptic face processing deficit. However, it is necessary to account for differential neural activity for visual and haptic face recognition. It has been demonstrated that successful within modal visual and haptic face recognition can occur on the basis of face shape alone (e.g. Bruce, Healey, Burton, et al., 1991; Hill, Schyns, & Akamatsu, 1997; Kilgour and Lederman, 2002), however, these studies also found that recognition performance improved with the inclusion of other sensory information (e.g. pigmentation in vision or surface texture in haptics). Therefore it is possible that the FRUs code information about face shape that is common to both modalities, but also information that is more modality-specific and not easily shared across modalities. Therefore FRUs may represent information in a fashion similar to Millar's (2002) model of convergence, discussed further in section 1.4.2.

The distributed neural system for face recognition may also accommodate both the behavioural and imaging results of haptic face recognition studies to date (Figure 1.4b). Haxby and his colleagues (2000) identified the lateral inferior occipital gyri (IOG) as the locus of early visual processing in their face recognition model. Pietrini et al. (2004) found that haptic recognition of all stimulus categories activated the ventral occipital cortex, an early visual processing area that includes the IOG. This implies that these structures may also mediate early haptic processing of face shape. These areas fall within the region of damage sustained by LH (Kilgour, de Gelder, & Lederman, 2004), possibly explaining his inability to recognise faces in either the visual or the haptic modalities. However, this model also suggests that visual prosopagnosia may occur independently of haptic prosopagnosia and vice versa, depending on the focus of sustained brain damage. For instance, if the IOG mediate multisensory representations of face shape, then damage to the FFA may impair visual face recognition, whilst haptic face recognition would remain intact through the haptic face recognition pathway. However, damage to the IOG necessarily precludes face recognition in either modality. Clearly further prosopagnosic case studies are required to investigate possible dissociations of visual and haptic processing deficits.

Behavioural evidence suggests that, although shape alone is sufficient for face recognition in either modality (e.g. Bruce et al., 1991; Hill et al., 1997; Kilgour and Lederman, 2002), within modal recognition improves significantly with the inclusion of modality-specific face information. Presumably visual and haptic modality-specific information is subject to differing computational demands, necessitating the inclusion of a separate area of cortex for higher-order haptic face processing. The modality-specific areas could receive sensory-specific information via direct pathways from primary sensory cortices, which is then bound with shape information processed in IOG, resulting in optimal within modal face recognition.

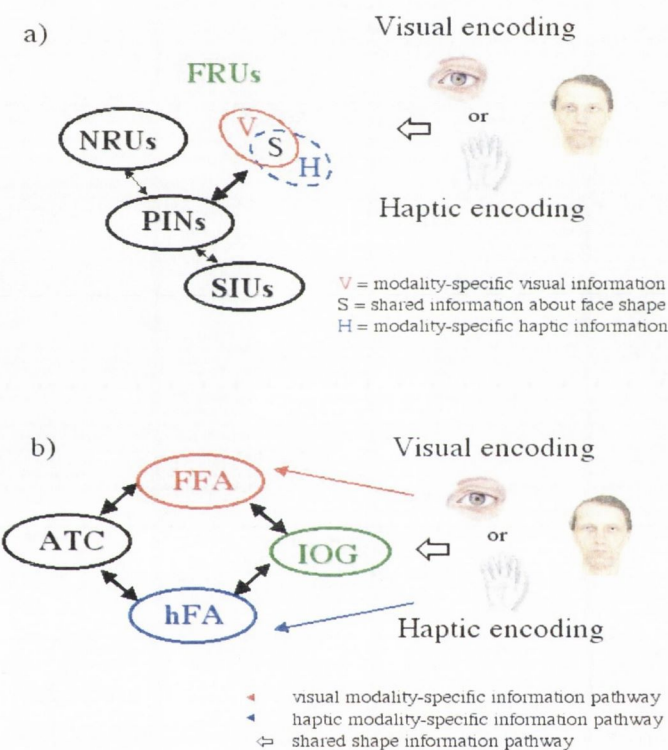


Figure 1.4. Models of crossmodal face recognition. This figure illustrates two possible ways for incorporating haptic information into extant face recognition models. In model a) the FRUs comprise overlapping representations of visual and haptic face information. In model b) visual and haptic shape information is independently processed in the lateral inferior occipital regions (IOG). Separate cortical areas may process modality-specific face information (visually in the FFA and haptically in a haptic face area). In this model, differential activation throughout the distributed system can result in either crossmodal face recognition, or within modal visual or haptic face recognition.

1.4 Crossmodal face recognition

This thesis endeavours to contribute to both the crossmodal and face recognition literatures. The aim of the experiments in the following chapters is to study further the processes underpinning visual and haptic face recognition, and examine how visual and haptic face representations may be stored in memory. The use of crossmodal experimental designs (see Figure 1.5), whereby faces are encoded in one sensory modality and recognised in another, should clarify what processes and types of face information can be utilised by both modalities for the purposes of face recognition. Thus, these experiments extend the work of previous researchers who have studied face recognition in one sensory modality only (i.e. vision or haptics), as the multisensory nature of face recognition can be observed explicitly, rather than being inferred.

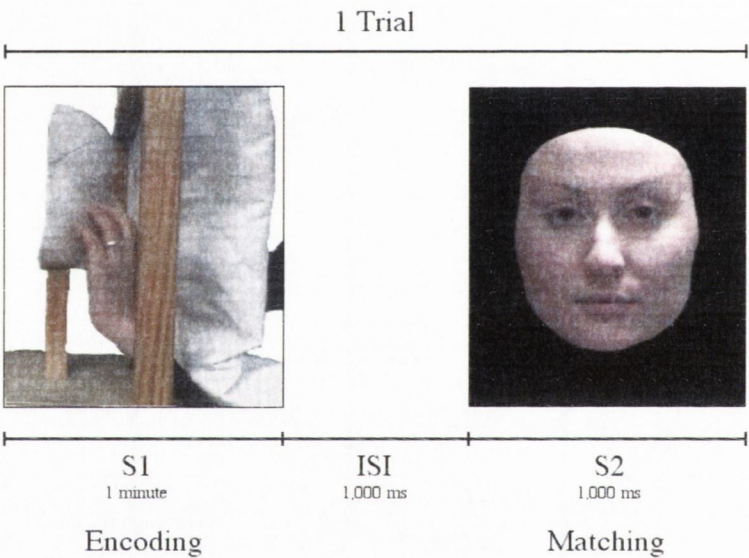


Figure 1.5. An illustration of the general match-to-sample crossmodal paradigm adopted in this thesis. In this illustration, the crossmodal task is in the direction of haptics-to-vision. The haptic face stimulus (S1) is encoded for a specified amount of time. After a short interstimulus interval (ISI), the visual face stimulus (S2) is presented. Participants must then make a response. For instance, in a matching task, the participant must decide if the visual stimulus represents the same person as the haptic stimulus by responding 'same' or 'different'.

1.4.1 Familiarity and crossmodal face recognition

The Bruce and Young and the IAC models of face recognition provide a coherent account of the various kinds of information about familiar faces. However the account

given of the processing of unfamiliar faces is much less detailed. A recent review by Hancock, Bruce, and Burton (2000) discussed how unfamiliar face recognition is akin to general pattern recognition: the recognition of unfamiliar faces can be easily disrupted by variations such as lighting, viewpoint, external features, facial expression, and image quality. However recognition of familiar faces can be effortlessly carried out under such conditions. Indeed, it has been shown that there is a large discrepancy between recognition rates for familiar and unfamiliar faces (Bruce, Henderson, Greenwood, et al., 1999). When faces in a face matching task are unfamiliar to the participants, performance is surprisingly poor even with no time pressure or memory load in the task. Such impoverished performance for unfamiliar faces may rely on the fact that the FRU for an unfamiliar face relies on just one encoded view, or experience of, that face. Thus, the representation may not be robust enough to facilitate very accurate recognition across multiple viewpoints and varying conditions. Although it has been demonstrated that unfamiliar faces can be matched across modalities (Kilgour and Lederman, 2002), changing modality can be considered like a change in view condition, thus it would be expected that familiar faces would be recognised across modalities with significantly greater accuracy than that observed for unfamiliar faces. In Chapter 2, the modulatory effects of familiarity are examined with respect to two different kinds of familiar stimuli: highly familiar face stimuli (participant's own faces; Experiment 1), and recently familiarised stimuli (Experiment 2).

1.4.2 The representation of faces by vision and haptics

In addition to vision and haptics both having the capacity to encode information about face shape, each modality also encodes sensory specific information, so that through the processes of sensory combination and integration (Ernst and Bühlhoff, 2004) a robust description of an individual face can be built and maintained in memory. The question arises as to how visual and haptic face information might be represented in the brain so as to facilitate such multisensory representations. A number of candidate models of crossmodal representation have been proposed and can generally be described as follows: a) all information is sensory specific without any crossmodal sharing, b) all representations are visual and information encoded haptically is recoded into vision, c) all information is multisensory and information contributes to the percept according to its reliability, d) information is partly shared and information contributes to the percept

according to its reliability. In terms of the models of crossmodal face recognition proposed in the previous section, these candidate models would describe how information would be represented at the level of the FRUs (Figure 1.4a) or processed by the IOG (Figure 1.4b).

Scenario a) is unlikely because, as outlined above, there is evidence to suggest that face information is shared across the modalities (Kilgour and Lederman, 2002). The second model, b), proposes that haptic information is mediated by the visual system (e.g. Zhang, Weisser, Stilla, et al., 2004). According to the image mediation model, haptic information is translated into a visual image which is subsequently processed and represented visually (Klatzky and Lederman, 1987; Lederman and Klatzky, 1990; Lederman, Klatzky, Chataway, et al., 1990). Although appealing, there is however no ready analogue for some haptic information, such as texture, temperature, and compliance, which contribute to haptic face recognition (Kilgour and Lederman, 2002). Furthermore, Kilgour and Lederman proposed that visual mediation of crossmodal face matching would be evident if a negative correlation was found between scores on the Vividness of Visual Imagery Questionnaire (VVIQ, a low score indicates high ability to use visual imagery) and both recognition speed and accuracy in a haptic face recognition task. Such negative correlations were not obtained.

Model c) is possible, but it is unlikely that the visual and haptic modalities share one overarching amodal face representation to which they both have equal access, as each modality encodes sensory specific information that contributes towards recognition. For example, face recognition performance is affected by attributes specific to vision such as pigmentation (Yip and Sinha, 2002), and contrast polarity (e.g. Kemp, Pike, White, & Musselman, 1996), as well as properties specific to the haptic modality such as skin texture and compliance (Kilgour and Lederman, 2002).

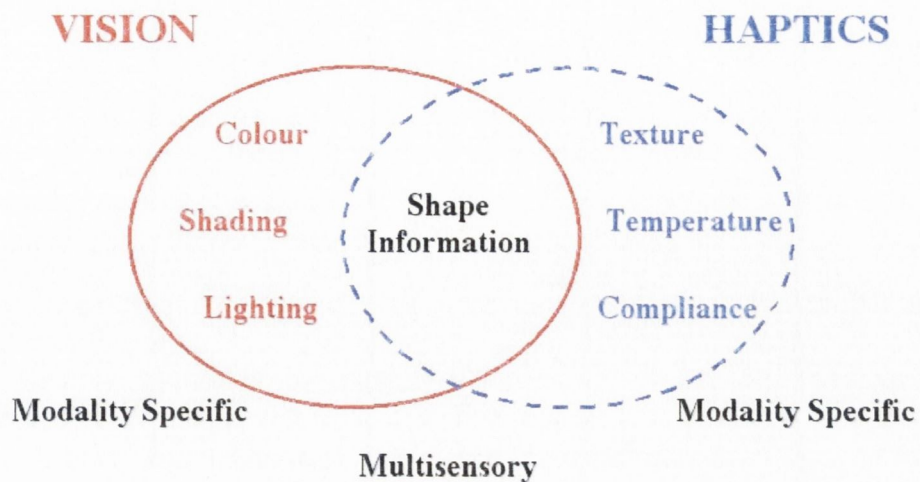


Figure 1.6. Model of convergence (Millar, 2000). Information common to both modalities is represented in the intersection of the diagram, whereas modality-specific information is represented on either side.

Lastly, option d) is a model of multisensory representation wherein information from different sensory sources (such as vision and haptics) partially overlaps. Instead of either wholly separate modules, or a single amodal system, the senses converge to share information that is represented in the same manner in both modalities (see Figure 1.6 for an illustration of this model) (Millar, 2000). For instance, vision and haptics can both encode geometrical properties of objects such as size and shape, and in both instances, geometric properties are the most diagnostic attributes for recognition (e.g. Lederman and Klatzky, 1990). Therefore, crossmodal face recognition may be mediated by a representation of facial shape common to both vision and haptics. The overlap of information comprises redundant information about facial shape that is encoded by both modalities, and this information is integrated (possibly according to the MLE model) to provide the best estimate of facial structure. Further to this, modality-specific face information can be combined to provide a rich description of the face for recognition through either or both modalities. This model of convergence would correspond well to the representation of face information in the FRUs of the model of crossmodal face recognition proposed in section 1.3.3.

The nature of the memory representations underpinning crossmodal visual and haptic face recognition is examined in explicit recognition tasks in Chapter 3 (Experiments 3 and 4). These experiments should provide insight into the relative

importance of encoded modality-specific information for subsequent face recognition across modalities.

1.4.3 Configural processing across modalities

Structural face descriptions stored in the FRUs are thought to comprise information about facial features and the configuration in which those features are arranged (Bruce and Young, 1986). Evidence suggests that each modality can process faces configurally, however, it is not clear whether this information is directly compatible. As faces fall outside the receptor bandwidth of the haptic modality, the haptic encoding of faces is limited to the serial encoding of face parts over time. Although there may be a shift towards a more global representation of a face as exploration time increases (cf. Lakatos and Marks, 1999), thus facilitating configural processing, the integration of serially acquired information can be error prone (cf. Loomis, Klatzky, & Lederman, 1991). Moreover, the representation of haptic space has been found to be distorted compared to visual representations, such that haptic spatial information is represented in a non-Euclidean manner (Kappers and Koenderink, 1999). This implies that the configural representations underpinning visual and haptic recognition of the same face might actually differ in a number of ways that may pose problems for information-sharing across the modalities. In Chapter 3 I investigate whether configural information can be shared directly between the visual and haptic modalities, and whether configural or featural information more readily facilitates crossmodal face matching (Experiment 5). Again, this experiment will provide an idea of how face information might be coded in memory, providing further insight into the mechanisms underlying crossmodal face recognition.

1.4.4 Multisensory structural face representations

Recognition accuracy has been found to decline significantly for objects recognised explicitly when the sensory modality changes between study and test, presumably reflecting the role of modality-specific information in within modal recognition (Easton et al., 1997; Reales and Ballesteros, 1999). However, crossmodal object priming experiments have demonstrated that priming effects are of similar magnitude within and across modalities, suggesting that visual and haptic object recognition is mediated by shared structural representations (Easton et al., 1997; Reales

and Ballesteros, 1999). Therefore, objects at least tend to follow the model of convergence proposed by Millar (2000). Whereas the experiments in Chapter 3 examine the importance of modality-specific information for face recognition, the aim of Chapter 4 is to investigate whether crossmodal face matching and recognition is mediated by structural representations of faces common to vision and haptics on an implicit level. A crossmodal priming task is used to test whether visual and haptic priming can elicit equivalent performance in a visual familiarity-decision task.

In the General Discussion (Chapter 5) the findings of these experiments are discussed collectively in terms of the multisensory and face recognition literatures. Both crossmodal recognition models and current models of face recognition are considered in light of the results. The implications of these behavioural data for the face specificity debate are briefly mentioned. Other issues related to the integration of sensory face information are raised and future avenues for crossmodal face recognition research are contemplated.

Chapter 2

Familiarity and crossmodal face recognition

Abstract

Recent studies have suggested that the familiarity of a face leads to more robust recognition at least within the visual domain. The aim of this chapter was to investigate whether face familiarity resulted in a representation of faces that was easily shared across the sensory modalities. In Experiment 1 I tested whether haptic recognition of a highly familiar face (i.e. one's own face) was as efficient as visual recognition. The observers were unable to recognise their own face models from tactile memory alone but were able to recognise their faces visually. However, haptic recognition improved when participants were primed by their own live face. In Experiment 2 I found that short-term familiarisation with a set of previously unfamiliar face stimuli improved crossmodal recognition relative to the recognition of unfamiliar faces. The findings suggest that familiarisation provides a strong representation of faces but that the nature of the information encoded during learning is critical for efficient crossmodal recognition.

2.1 Introduction

The various facets of visual face processing and recognition are well documented (e.g. Bruce and Young, 1998; Kanwisher and Moscovitch, 2000; Peterson and Rhodes, 2003; Rakoner and Cahlon, 2003; Tarr and Cheng, 2003). However, very little is known about our capability for processing and recognising faces outside the visual domain. Recently, Kilgour, Lederman and colleagues have embarked on a series of studies aimed at furthering our understanding of face recognition in the tactile domain (Kilgour, de Gelder, & Lederman, 2004; Kilgour and Lederman, 2002). In their initial study, Kilgour and Lederman (2002) demonstrated that haptic input allowed for the successful tactile matching of unfamiliar faces. Furthermore, they found that participants were able to match unfamiliar faces across sensory modalities suggesting that face information can be represented and shared across vision and haptics.

The studies reported by Kilgour et al. suggest that the haptic and visual systems both have the capacity to process faces and that face information can be shared across

sensory modalities, possibly involving the same neural substrate (Kilgour, de Gelder, & Lederman, 2004). Given this evidence, it might be reasonable to assume that other factors affecting visual face recognition will similarly influence face recognition through touch. In this chapter I investigated whether or not familiarity with a face leads to a better recognition of that face across modalities. In particular, I was interested in whether long-term familiarity, such as familiarity with one's own face, results in robust recognition in both vision and haptics. Furthermore, I tested whether short-term familiarity, such as recent exposure to a face, results in better crossmodal recognition.

The degree to which a face is familiar to an observer has been found to influence the way in which that face is processed and recognised through vision. Familiarity with a face can confer recognition advantages, in terms of accuracy and response latencies. For instance, when deciding whether a face is intact or scrambled, familiarity facilitates quicker responses to intact stimuli (Bruce, 1986). A large discrepancy has also been found between recognition rates for unfamiliar and familiar face images captured from video. Face matching performance was poor when face images were unfamiliar, despite unlimited time and reduced memory load (Bruce, Henderson, Greenwood, et al., 1999). However, even in poor quality CCTV images, participants were surprisingly good at recognising highly familiar targets from video footage. Performance however dropped to chance levels if participants were unfamiliar with the targets (Burton, Wilson, Cowan, & Bruce, 1999). Hancock, Bruce, and Burton (2000) argued that the recognition of unfamiliar faces is not robust to changes in image-based factors such as changes in lighting (Johnston, Hill, & Carman, 1992) viewpoint (Hill, Schyns, Akamatsu, 1997; Newell, Chiroro & Valentine, 1999; O'Toole, Edelman, & Bülthoff, 1998), or facial expression (Bruce, Henderson, Newman, & Burton, 2001). Recognition of familiar faces, however, can be effortlessly carried out despite such variations (Hancock et al., 2000).

Although both long and short term familiarity have been shown to lead to more robust visual recognition it is not clear whether the effect of familiarity is sensory-modality specific or independent of sensory modality. On the one hand, familiarity with a face may promote a more multisensory representation to allow for better crossmodal recognition of the face. Indeed it has been shown that the recognition of highly familiar objects generalises across modalities (e.g. Reales and Ballesteros, 1999) and if faces are considered as a type of object class (e.g. Tarr and Gauthier, 2000) then we might expect

that the recognition of familiar faces also generalises across modalities. On the other hand, faces represent a set of highly similar stimuli. Studies of crossmodal object recognition using highly similar stimuli found that crossmodal recognition was impaired relative to unimodal recognition (Newell, Ernst, Tjan & Bühlhoff, 2001). The participants were familiarised with these novel objects in one modality only during the course of the experiment. Therefore, it is possible that crossmodal object recognition benefits from more long-term familiarity with objects. The aim of this study, therefore, was to investigate the role of familiarity in the recognition of faces across vision and haptics.

Vision researchers have studied the effects of familiarity on face recognition performance by comparing responses to unfamiliar faces with responses to either newly learned faces (short-term familiarity e.g. Katanoda, Yoshikawa, & Sugishita, 2000), or highly familiar or famous faces (long-term familiarity e.g. Bruce, 1986; Bruce and Valentine, 1985). Furthermore, faces varying in familiarity may undergo differential processing in terms of the facial information used, or the neural substrate involved (e.g. Clutterbuck and Johnston, 2002; Leube, Erb, Grodd, et al., 2003). In this study I was concerned with the role of both long-term and short-term familiarity in multisensory face recognition. I investigated these effects first by testing self-face recognition across modalities (Experiment 1), and then by testing crossmodal face recognition to a recently familiarised set of faces (Experiment 2).

In Experiment 1 I examined haptic recognition of one's own face relative to visual recognition. Given the long-term familiarity of our own faces, we should expect to have a robust representation of our face in visual memory, as was reported in previous studies using famous faces. However, we also have a lifetime's experience with our own face through touch (e.g. from grooming, etc.) and it may be possible that we use this haptic information along with visual information to construct a multisensory representation of our faces in memory. Currently there is no direct evidence that such a representation exists, however the results of a recent study on reach and grasp movements suggest that a haptic representation of our own face is available in memory, at least at the featural level (Edwards, Wing, Stevens, & Humphreys, 2004). If familiarity with a face facilitates multisensory recognition, it was hypothesised that self-face recognition performance would be similar across the visual and haptic modalities.

In Experiment 2 I investigated whether familiarity obtained due to brief exposure to previously unknown faces promotes more efficient crossmodal face matching or if the benefits of short-term familiarity are modality specific. It was hypothesised that if familiarity confers more multisensory representations of faces, better crossmodal matching performance should be found for familiar than unfamiliar faces, regardless of whether they were familiar in vision or haptics.

2.2 Experiment 1

Experiment 1 examined whether or not individuals are able to recognise their own face through touch. Our own face is highly familiar to us visually and possibly through touch, suggesting that both a visual and a tactile representation of our own faces may reside in memory. Kilgour and Lederman (2002) have shown that practice with even unfamiliar faces allows for sufficient haptic recognition of faces. It was predicted therefore, that long-term familiarity would facilitate self-face recognition, with participants easily able to recognise models of their own faces through touch, at least as well as through vision.

2.2.1 Method

2.2.1.1 Participants

Thirty-two undergraduate and graduate students (16 male, mean age 22.88 years, $SD = 4.78$) from the School of Psychology, Trinity College Dublin participated in this study for research credits. Testing procedures of this and the following experiment met with the standards delineated by the Ethics Committee of the School of Psychology, Trinity College Dublin, and all participants gave written consent prior to performing this experiment. All participants reported normal or corrected-to-normal vision and none reported any tactile impairment.

2.2.1.2 Apparatus and Stimuli

Stimuli comprised 32 plaster models, one of each participant's face. To create these face models each participant's hair was first covered with a swimming-cap and then a thin layer of petroleum jelly applied to their eyebrows and eyelashes. A thick layer of Accu-cast 880TM, an alginate casting substance, was applied to the face leaving an opening for the nostrils and was allowed to set for 8 minutes. A layer of gypsum (i.e.

plaster-of-paris) bandages was then applied over the alginate and allowed to harden for 10 minutes before it was removed from the participant's face. A thin layer of casting plaster was poured into each face mould and allowed to stand for 20 minutes. A wooden mount was inserted into the base of the mould consisting of a square pole attached to a base. Each face model was positioned approximately 11 cm above the base of the mount so that participants could rest their elbows comfortably on the table whilst palpating the facemodels. The face models were left to dry in the mould for 24 hours and when removed were allowed to dry for a further 24-48 hours. Small imperfections were corrected using very fine-grain sandpaper. An example of a male and female face model along with their live counterparts is illustrated in Figure 2.1. A trial consisted of a 'line-up' of face models including the target face and 7 non-target face models, all of which were positioned side-by-side on a workbench.

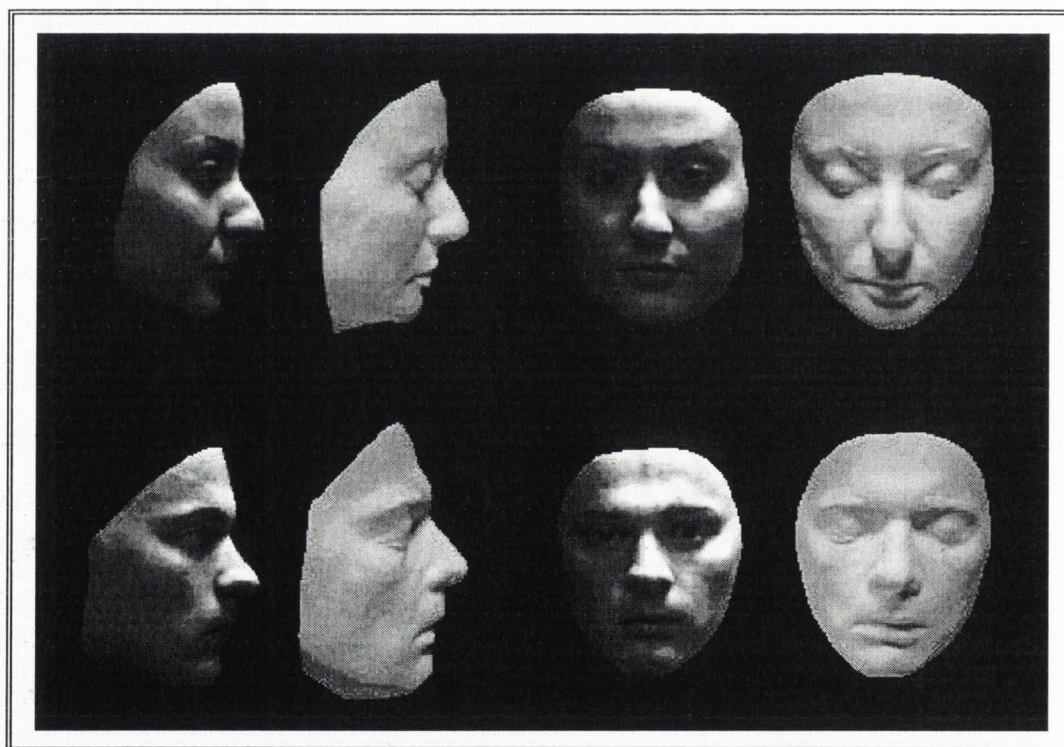


Figure 2.1. A sample of the haptic face stimuli used in all of the experiments in this thesis. On the top row a female face is compared with its plaster model in both profile and portrait views. The bottom row shows the same detail for a male face. The haptic models provide faithful representations of the structure of the live faces.

2.2.1.3 Design

The experiment was based on a one-way repeated measures design with recognition modality as the factor. This factor comprised three levels; two haptic and one visual. The two haptic conditions were haptic recognition from memory (haptic memory) and haptic recognition after priming (haptic prime). The visual condition was recognition from memory (visual memory). There were two trials to the haptic memory condition (namely forward-facing and backward-facing, based on the direction of the face model with respect to the participant)^{2.1} and one trial each in the haptic prime and visual memory conditions. In each trial the target face had to be recognised from among seven distractors in a 'line-up' scenario. For each participant the same distractor faces were used in all four line-ups. Distractor faces were randomised across participants. For all participants, the forward-facing and backward-facing face stimuli were counterbalanced across trials in the haptic memory condition and the position of the target face stimulus in the line-up was randomised across trials. The trials always followed the same sequential order; haptic memory, haptic prime and visual memory. This was necessary in order to avoid learning effects or effects of visual imagery on performance in the haptic memory task. There were 2 trials in the haptic memory condition, and 1 trial each in the haptic prime and visual memory conditions. Performance was measured in terms of recognition accuracy.

2.2.1.4 Procedure

Approximately one week after the casting process, participants were recalled to complete the face recognition task. The task for each participant was to select his or her own face from a line-up comprising eight faces which always included the target face and seven distractor faces. In the haptic trials participants could feel each stimulus for an unrestricted period of time but were not permitted to return to any previously explored face once they had moved on in order to promote a recognition-from-memory strategy rather than a strategy involving feature comparisons. Participants were blindfolded for the first two haptic memory trials. In the haptic prime trial the blindfold was removed but the face model was positioned such that it faced away from the participant and was

^{2.1} Both the forward and away facing orientations of the face models relative to the observer were included in order to avoid any advantages or disadvantages of orientation presentation (see e.g. Newell, Ernst, Tjan & Bühlhoff, 2001).

thus obscured from view. In this trial participants were instructed to first feel their own face for 10 seconds before feeling each face model.

Ten seconds were necessary so that the participant had enough time to feel all of their face and not just one or two features (cf. Lakatos and Marks, 1999). In the visual memory trial, participants were presented with the line-up of face models and required to visually identify their own face from among the distractor face models by indicating its position in the line-up.

2.2.2 Results

The data were first analysed across the trials within the haptic memory condition only in order to determine if there was any benefit for one particular orientation of the face model over the other with respect to the observer. A Wilcoxon Matched-pairs test was used and found no effect of face orientation [$T(32) = 11$, n.s.]. For subsequent analyses each participant's average performance for the haptic memory condition was used.

The mean percentages of correct recognition responses for the different conditions are illustrated in Figure 2.2. A Friedman's ANOVA was performed on recognition accuracy performance across the haptic memory, haptic prime, and visual memory conditions. A main effect of condition was found [$\chi^2=21.92$, $df = 2$, $p < 0.001$]. Post hoc analyses were conducted using Wilcoxon Matched-Pairs tests and I found that recognition performance was significantly better in the visual memory condition than in either the haptic memory condition [$T(32) = 11$, $p < 0.001$] or the haptic prime condition [$T(32)=11.0$, $p < 0.01$]. Furthermore, performance in the haptic prime condition was significantly better than performance in the haptic memory condition [$T(32)=36.5$, $p < 0.03$]. Priming of the participant's own face facilitated the subsequent haptic recognition of his or her own face model. Indeed, using a Sign test it was found that performance in the haptic memory condition was not significantly greater than performance due to chance [$z=0.5303$, n.s.].

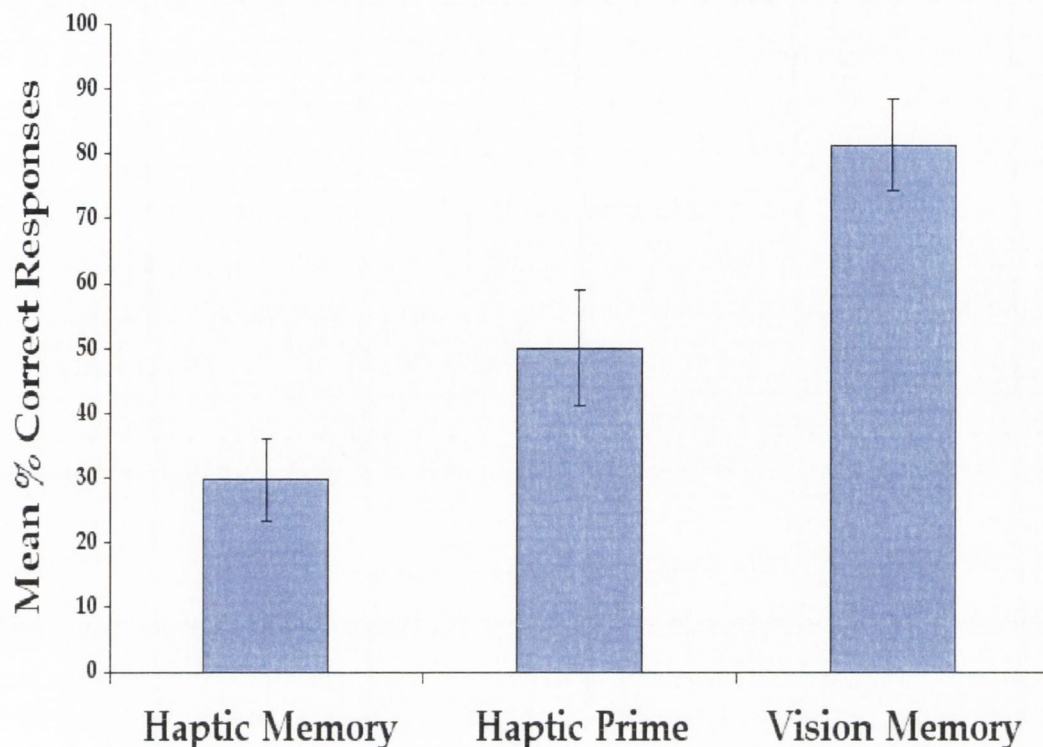


Figure 2.2. Mean percentages of correct recognition of one's own face across the haptic and visual conditions. Error bars represent ± 1 standard error of the mean.

The haptic prime trial was, however, always preceded by 2 haptic memory trials raising the concern that better performance in the haptic prime trial was simply a practice effect. To rule out this possibility, trials in the haptic memory condition were divided into their sequential trial order and analysed using a Wilcoxon Matched-pairs test. No difference in performance was found between the first and second trials in the experiment [$T(32) = 27.5$, n.s.] suggesting that the improved performance in the haptic prime condition was not simply due to practice.

2.2.3 Discussion

In this experiment it was found that participants were unable to recognise their own faces through touch alone. However, haptic recognition of their own face improved when participants were allowed to feel their own live face prior to feeling a model of their face. Nevertheless, haptic recognition was not as good as visual recognition of one's own face.

Poor haptic self-face recognition relative to visual recognition of the same face model might suggest that haptic familiarity with the face is insufficient for a tactile representation in memory. However, there may be other reasons why haptic performance was poor relative to visual performance. For example, although we have a lifetime of experience with touching our own faces, the tactile experience may be incidental to the task at hand (e.g. washing, make-up application etc.) and therefore the conditions for encoding the shape of the face in tactile memory may not be optimal. It may be necessary to attend to the shape characteristics of the face during encoding in order to form a representation of that face. For vision, attending to the shape of one's face is necessary when we view images of ourselves in photographs, film and in the mirror. However, there are few, if any, real-world tasks that require us to attend to the entire shape of our face through touch. In the Kilgour and Lederman (2002) study, good haptic recognition performance was achieved because participants were familiarised with faces through touch. Here I have shown that despite tactile familiarity, there does not appear to be a representation of the shape of one's own face in tactile memory.

Edwards et al. (2004) found that over-estimations of the grasp size necessary to grasp an object are significantly smaller for facial features such as the nose or mouth compared to similar sized neutral objects or other body parts (wrist or thumb). These authors discounted the notion that the familiarity of the action of grasping face-parts might contribute to this finding, instead proposing that familiarity of a feature underpinned a more accurate grasp size. Given the poor haptic performance in this task, however, it is possible that even a feature recognition strategy is insufficient for facilitating haptic face recognition from memory, at least without a relevant prime.

Haptic recognition performance improved after a brief priming event. This improved performance was not simply due to a practice effect with the face stimuli since no effect of practice was found across the first 2 trials (i.e. in the haptic memory condition). Thus, when participants could attend to the particular shape of their own faces through touch, they were then able to haptically identify their own face model. This result suggests that it may be possible for the haptic system to build a representation of a face shape (even if it is a short-term representation) for subsequent recognition purposes (see also Kilgour and Lederman, 2002).

However, it is important to consider an alternative explanation for the poor recognition performance in the haptic memory condition. Whilst information about

shape contributes to recognition in both vision and haptics, the haptic modality also places considerable importance on the material properties of objects (i.e. surface texture, compliance, and temperature) for accurate recognition (e.g. Klatzky, Loomis, Lederman, et al., 1993; Klatzky, Lederman, & Reed, 1987). Access to material properties of faces, (e.g. skin texture and muscle tone) has been found to optimise haptic face matching performance; haptic matching performance for live faces was significantly better than that for clay face models (Kilgour and Lederman, 2002). In this experiment participants felt rigid 3-D, uniformly textured facemasks, thus it is possible that there was a loss of likeness in the recognition process as material information that may form an integral part of their (putative) haptic face representation was missing from their target facemask (as was the material information on the distractor models which would facilitate correct rejections through comparison). Coupled with the notion that there might not be an adequate representation of one's own face shape in tactile memory, this may help to explain why recognition performance in the haptic memory condition was poor. Therefore, given the absence of material properties in the haptic stimuli one can not necessarily preclude the existence of a haptic representation of one's own face in memory based on performance in this experiment alone.

In the visual memory condition I found that participants' recognition performance was not perfect as might be expected: approximately 20% of our participants made an error in this condition. I speculated that these errors were due to recognising non-normal versions of their own face. For example, the face models lacked any pigmentation cues which might normally contribute to the recognition of faces (Yip and Sinha, 2002). Indeed, some studies have demonstrated that the addition of colour texture to facial surfaces greatly improves performance in tasks involving the recognition of faces from a novel viewpoint (Troje and Bühlhoff, 1996) and the categorisation of faces according to their sex (I. Bühlhoff and Newell, 2004).

Although visual recognition of the participant's own face model was good, the representation of their face in visual memory was clearly discrete and could not be shared across modalities: haptic self-face recognition did not seem to benefit from a visual representation of that face. This finding suggests that although familiarity may promote good visual recognition, it does not promote multisensory recognition. In the next experiment I explicitly tested whether or not familiarity promotes better recognition

across the sensory modalities by familiarising participants with previously unfamiliar tactile and visual faces.

2.3 Experiment 2

It is argued that familiar faces are processed more efficiently in the visual system than unfamiliar faces (e.g. Young, Hay, McWeeny, Flude, & Ellis, 1985) leading to more robust face perception (e.g. Schwaninger et al., 2002). It is not known, however, whether familiar faces are more easily recognised across the sensory modalities or whether familiarity benefits a sensory-modality specific representation of that face. The results of the previous experiment suggest that long-term familiarity affects modality specific representations although it may be that visual face recognition was best because of a lifetime's experience of processing information in that modality. Consequently, face representations may be 'captured' by vision when they are highly familiar in that modality because vision may be the more appropriate modality for face recognition (Welch and Warren, 1980). Thus, if familiarity promotes robust recognition across the senses then it would be expected that familiar faces would be matched more efficiently across modalities than unfamiliar faces. On the other hand, if familiarity affects the representation of a face in a modality specific manner, then there should be no benefit for matching familiar over unfamiliar faces across modalities.

2.3.1 Method

2.3.1.1 *Participants*

Thirty-nine psychology undergraduates (28 female, mean age 20.3 years, SD = 3.3) participated in this experiment to fulfil research credit requirements. Written consent was obtained from all participants prior to the experiment and none had participated in the previous experiment. All participants reported normal or corrected-to-normal vision and none reported any tactile impairment.

2.3.1.2 *Apparatus and stimuli*

Twenty-six plaster models of faces were randomly chosen from the stimulus set in Experiment 1 and used as haptic stimuli in this experiment. Digital colour images of the live faces of the individuals from whom the models were cast were used as visual stimuli. All of the individuals used as models in this experiment were unfamiliar to the participants, who had just commenced their first year of study in the School of

Psychology. This was confirmed verbally after the experiment was conducted. During the familiarisation stage of the experiment the visual images were presented as a set of laminated photographs and during the experiment they were presented as digital images using Microsoft PowerPoint on a Macintosh PowerBook G3. The images presented on the computer screen subtended a visual angle of approximately 10° vertically and 7° horizontally. Participants were required to learn a fictitious name and occupation for each face during familiarisation. Familiar, simple names (e.g. Sue, Mike, Jill etc.) and familiar occupations (e.g. builder, teacher etc.) were chosen for the participants to learn in order to reduce any memory demands on the task. The haptic stimuli were presented facing towards the participant whilst the visual stimuli were presented in a 3/4 view to maximise the amount of structural information available in the image (e.g. Baddeley and Woodhead, 1983; Blanz, Tarr, & Bühlhoff, 1999).

During the experiment, the participant sat in front of a table and the haptic stimuli were presented on the table behind an occluder (i.e. a curtain mounted on a wooden frame). Participants were required to place their hands underneath the curtain to touch each face model. A Macintosh PowerBook was positioned on the table to the right of the participant in order to display the visual images.

2.3.1.3 Design

This experiment was based on a mixed design with sensory modality matching conditions (vision-to-haptics or haptics-to-vision) as a between subjects factor and face familiarity (familiar or unfamiliar) as a within subjects factor. Participants were randomly assigned to either the visual or the tactile learning modality.

The experiment began with a familiarisation session where participants learned to identify a set of faces either through haptics or vision alone. All participants were familiarised with 13 faces, one of which was used in a subsequent practice trial, and the other 12 in the experiment. Within each visual or haptic learning group in the experiment, half of the participants were familiarised with 12 of the 24 face stimuli and half were familiarised with the other 12 stimuli. The presentation order of the face stimuli was randomised across participants.

The experiment was based on a delayed match-to-sample paradigm where participants had to match faces across sensory modalities. At encoding and test, a haptic stimulus was always presented for 60 seconds whereas a visual stimulus was presented for 1 second, in an attempt to compensate for differences in processing expertise

between vision and haptics. Stimulus presentation times vary within the face recognition literature depending on such things as the type of memory being studied (implicit or explicit), the familiarity of the face, perceptual manipulations of the image (e.g., lighting, viewpoint, image quality), and whether behavioural or neuroimaging data are collected. Although some studies have found that familiar faces can be recognised after as little as 700ms (e.g. Paller, Gonslaves, Grabowecky, et al., 2000), unfamiliar faces typically take longer to discriminate (~1 sec; e.g. Hole, 1994). Therefore, it was decided to present visual stimuli for 1 second, to ensure that unfamiliar faces were sufficiently encoded to facilitate recognition across the modalities. Whereas vision is capable of encoding the global structure of a face quickly and easily, haptic face recognition is much slower due to the processing capacities of haptic receptors, requiring the integration of serially acquired information into a coherent whole over time (Loomis and Lederman, 1986). In a previous crossmodal face matching task where no haptic exploration time limit was enforced, participants tended to respond after an average of 35 seconds (Kilgour and Lederman, 2002). However, in that study haptic stimuli comprised live faces from which both material (texture, temperature, and compliance) and geometric haptic information could be obtained to facilitate crossmodal matching performance. Indeed, within modal matching accuracy dropped significantly when facemasks, rather than live faces, were used, suggesting that available material properties optimise haptic performance. As the facemasks used in this experiment provided geometric information only, it was decided to extend the duration of haptic stimulus presentation to 60 seconds to promote, as much as possible, the configural processing of the facemasks (cf. Kilgour, 2003; Lakatos and Marks, 1999), and to facilitate optimal matching performance across the modalities.

Each face stimulus appeared twice during the matching task, once in a 'same' or matching trial and once in a 'different' or non-matching trial. Half of the total number of 24 trials included unfamiliar stimuli and half included familiar stimuli. The presentation order of the trials was randomised across participants. In order to avoid a response bias, there were equal numbers of same and different trials although the dependent variable was the mean percentage of hits only (i.e. accuracy to the same trials,

of which there were 6)^{2.2}. Familiarity was counterbalanced across the same and different trials.

2.3.1.4 Procedure

Participants were required to perform a familiarisation task prior to the experiment. Familiarisation consisted of learning the name and occupation (e.g. "Sue the accountant") of each face stimulus and subsequently identifying each face by both its name and occupation. A learning criterion of 90% accuracy was set in both the visual and haptic familiarisation tests before the participant could proceed to the experiment. The familiarisation process was repeated until this criterion was achieved. On average, the familiarisation of the haptic stimuli took about 75 minutes to complete whereas familiarisation with the visual task took about 5 minutes to complete.

During the experiment, a haptic stimulus was always presented for 60 seconds whereas a visual stimulus was presented for 1 second. The participant pressed the space bar on the computer to begin each trial. In the vision-to-haptic (i.e. V-H) matching condition, a 1 second fixation cross preceded the visual stimulus. In the haptics-to-vision (i.e. H-V) matching condition, the experimental procedure was identical to the V-H condition except that the haptic stimulus was presented before the visual stimulus. In any trial, the presentation of the second stimulus immediately followed the first stimulus. At the end of each trial, participants were required to make a same/different response by circling the appropriate answer on a response sheet.

2.3.2 Results

Figure 2.3 depicts the mean percentages of hit responses for familiar and unfamiliar faces in the crossmodal matching task.

The data were analysed using a 2 x 2 mixed design ANOVA with matching condition as a between subjects factor (V-H or H-V) and face familiarity (familiar or unfamiliar) as a within subjects factor. A main effect of learning modality was found, [$F(1,37) = 13.80, p < 0.001$], with better performance in the V-H condition than in the H-V condition. A main effect of familiarity was also found, [$F(1,37) = 4.65, p < 0.04$],

^{2.2} The second stimulus in a 'different' trial was always presented in the non-familiarised modality. For this reason, performance in the same trials only was used as the performance measure.

with better crossmodal matching performance for familiar faces than unfamiliar faces. There was no significant interaction between the factors, [$F(1, 37) = 0.016$, n.s.].

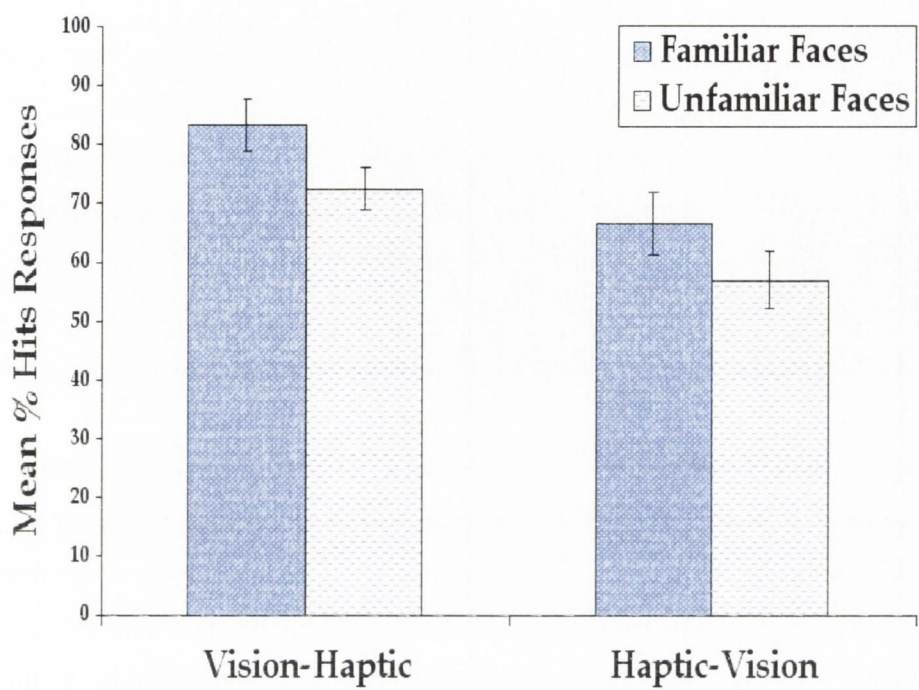


Figure 2.3. Mean percentages of correct ‘hit’ responses in the cross-modal matching task. Error bars represent ± 1 standard error of the mean.

2.3.3 Discussion

In this experiment it was found that crossmodal matching for newly familiar faces was better than for unfamiliar faces, regardless of whether face stimuli were matched from vision-to-haptics or haptics-to-vision. These results suggest that short-term familiarity promotes better recognition across sensory modalities.

Although I tried to ensure that the encoding of the face stimuli was equivalent across modalities by allowing 1 minute for haptic encoding and 1 second for visual encoding, better V-H performance than H-V performance was found. This difference in performance may have been due to the qualitative differences in information processing across the modalities which, in a matching task, may benefit V-H performance over H-V performance. For example, faces are encoded holistically in vision, however the haptic encoding of a face is not holistic but involves a feature-by-feature analysis. Previous studies have found that it is difficult to identify a feature of a face when the feature is presented in isolation or outside the context of the whole face (see Tanaka and Farah,

1993; Tanaka and Sengco, 1997). If haptics encodes faces on the basis of features then the visual recognition of a face from its feature-based haptic representation may have been less efficient than the haptic recognition of a face from its holistic, visual representation.

2.4 General Discussion

In summary, the findings support previous studies on haptic face recognition in that the results of both experiments suggest that faces can be represented and recognised through touch (Kilgour and Lederman, 2002; Kilgour et al., 2004). In this chapter I was particularly interested in the role of long-term and short-term familiarity in crossmodal face recognition. Despite a lifetime of experience with touching one's own face, in Experiment 1 there was no evidence that a representation of that face was built in tactile memory. Instead, the recognition of one's own face model through haptics seemed to rely on a brief exposure to the shape of the actual face prior to recognising the face model. Furthermore, the results of Experiment 1 suggest that the representations of faces are discrete and are not easily shared across the sensory modalities. Visual familiarity with one's own face did not appear to facilitate crossmodal recognition, since information from the visual representation could not be used by the haptic system. In Experiment 2, however, I found that faces can be matched across the modalities and despite only a relatively brief familiarisation period, this matching process is better for familiar faces than unfamiliar faces.

These results raise two related questions, first, why a representation of one's own face is not built in tactile memory and second, why a visual representation of one's own face is not shared across modalities whereas the representations of other recently familiar faces do seem to be shared across modalities. The results of Experiment 1 suggest that, although highly familiar, the representation of one's own face is modality specific and therefore cannot be shared across the modalities. Yet, the results of Experiment 2 clearly suggest that even short-term familiarity can affect a more multisensory representation of the face, at least one that can be shared across modalities. If familiarity promotes a more multisensory representation, then long-term familiarity should result in efficient sharing of the face representation across modalities. Thus, there seems to be a large discrepancy between these findings.

There may be several reasons for this discrepancy. First, it might be that the encoding of over-familiarised faces, such as our own face, is based on task-specific requirements. For example, visual encoding may result in a holistic representation of our face for the purpose of self-recognition when we view photographs, film images or mirror reflections of ourselves. Haptic encoding, on the other hand, may be less concerned with the overall shape of one's own face and tasks involving touching one's own face may rely more feature specific information (such as, judging the length of a beard before shaving or the texture of the skin before applying moisturiser). It is conceivable, therefore, that the haptic encoding of our own face is optimised for specific tasks but that the visual encoding and representation is more optimal for recognition purposes. Shape information can, however, be used by the haptic system, as shown in the haptic prime condition of Experiment 1 and the results of Experiment 2, but this information may require attention during encoding.

Another reason why the haptic recognition of one's own face model was less efficient than visual recognition may be because the face models lack information necessary for tactile recognition. For example, the plaster face models lack material properties present in a live face such as information about skin texture, tightness and muscle tone, which may have led to poor haptic recognition. Indeed, the lack of colour information, or pigmentation, may have resulted in poor visual recognition of one's own face. Therefore, it may be important that information available in a real face is made available in the face stimuli. In Experiment 2, on the other hand, the material properties of the haptic stimuli were the same at both study and test, so matching accuracy was not influenced by differences in these properties.

Finally, the representation of one's own face may be qualitatively different from the representation of the faces of other familiar persons (e.g. Keenan, Nelson, O'Connor, & Pascual-Leone, 2001; Kircher, Senior, Phillips, et al., 2001; Platek, Keenan, Gallup, & Mohamed, 2004; Turk, Heatherton, Kelly, et al., 2002). Brady, Campbell, & Flaherty (2004) found a significant preference for mirror-symmetric right-right own-face chimeras compared to a bias for composite faces of close, familiar friends comprising mirror-symmetric images of the left face. Brady et al. argued that the preference for the right side of one's own face is because of experience with mirror images of our faces. The unique way in which one's own face may be represented in visual memory may promote a more modality specific representation of one's own face, or at least one that is

not easily shared across modalities (e.g. due to mirror reversal). (It is therefore conceivable that self-face recognition is subject to a visual capture effect (Rock and Harris, 1967) which hinders successful haptic identification. This effect would not be apparent with the short-term representations of the familiar stimuli in Experiment 2, as in this instance the information provided by the visual and haptic modalities is not conflicted.)

In this study, therefore, differences in performance between the haptic recognition of one's own face relative to the recognition of other familiar faces may reflect qualitative differences between the underlying visual representations. Perhaps an examination of the crossmodal capacity for recognising the other highly familiar faces such as famous faces might produce effects reflecting those obtained with the briefly learned faces of Experiment 2.

Currently within visual face recognition research it is debated as to whether face recognition involves specific mechanisms, distinct from the mechanisms subserving the recognition of all other classes of objects. On the one hand, a variety of data provide evidence that face recognition is a unique process (see Kanwisher, 2000 for an overview). However, data also exist to show that face recognition is not 'special' but is rather a function of processing expertise, and that similar results can be obtained for non-face objects with which an individual is expert (e.g. Gauthier, Tarr, Moylan et al., 2000; Tarr and Gauthier, 2000). Whether or not the current results reflect an ability to generate haptic representations of faces that differ from the representation of other objects is beyond the scope of this thesis. It will be interesting for future research to examine whether the benefit of familiarity for the crossmodal recognition of faces is different to that for other objects. Haptic face recognition research may have much to contribute towards this specificity debate and to crossmodal research in general, in terms of the mechanisms underpinning the integration of visual and haptic information.

2.5 Conclusions

The findings reported in this chapter support recent evidence that touch, like vision, can encode and represent faces for the purpose of recognition. Moreover, the results of the experiments suggest that short-term familiarity results in a representation of a face that is more easily shared across the sensory modalities. Familiarity of a face is, however, insufficient for recognition unless those face characteristics specific for

recognition are encoded and represented by haptics. The representation of one's own face in memory may be qualitatively different in comparison to other long-term and short-term familiar faces, in that it is modality-specific and not easily shared across modalities.

Chapter 3

Are representations of faces independent of encoding modality?

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 Experiment 3 I found better within modal than crossmodal face recognition performance suggesting that face representations are specific to each modality. Moreover, in Experiment 4, this cost in cross-modal performance was found to be independent of differences in exploratory procedures across the modalities during encoding. In Experiment 5 I 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. These 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 cross-modal recognition.

3.1 Introduction

Although there has been much research into how visual and haptic information is integrated into multisensory object representations in memory (e.g. Ernst and Bühlhoff, 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, 1967; 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 and Carey, 1986; Gauthier and 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. It might therefore be expected that the contribution of information from another modality, such as touch, 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 contribute to the representation of an object in memory (Easton, Greene, & Srinivas, 1997; Reales & Ballesteros, 1999; Newell, Ernst, Tjan, & Bühlhoff, 2001). Such multisensory object representations allow for more robust recognition performance (Ernst and Bühlhoff, 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 generally occur in close temporal proximity and are thus likely to be integrated to form a coherent 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 sharing (e.g. Newell et al., 2001). Therefore it could be expected 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. For instance, it has been found that unfamiliar faces can be successfully matched across modalities (Chapter 1; Kilgour and Lederman, 2002). Perceptual matching, however, may rely on information processing that is not specific to face perception per se but involve more generic, image-based processing. The question arises as to whether or not face representations in memory are modality-independent or are more reliant on modality-specific information.

Functional models of face recognition (e.g. Bruce and 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 and Klatzky, 1990) then, in principle, haptic information could also contribute to the recognition of a face. 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 and Hole, 2000; Schwaninger, Lobmaier, & Collishaw, 2002) and configural information (e.g. Tanaka and Farah, 1993; Tanaka and 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 and Lederman, 1986). However, Kilgour and her colleagues (Kilgour, 2003; 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 and Searcy, 1993; Lewis and Glenister, 2003; Freire, Lee, & Symmons, 2000). As noted by Maurer, Le Grand, and Mondloch (2002), inversion in itself does not constitute conclusive evidence for face-specific configural processing as inversion has been found to affect the recognition of other mono-oriented objects such as cars (e.g. Gauthier, Skuldarski, Gore, & Anderson, 2000b), thus the use of alternative paradigms are necessary to substantiate claims of configural face processing in the haptic modality, and to investigate the sharing of configural face information across the modalities.

Face recognition may rely more on subtle visual differences between faces. For example, visual features such as pigmentation (Yip and 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 chapter, I investigated the extent to which face representations in memory are shared across modalities. In Experiment 3 recognition memory performance was

investigated for a set of unfamiliar faces within and across modalities. It was hypothesised that if memory representations contain information that is multisensory then performance within and across modalities should be equivalent. However, if face representations are modality-specific, then a cost in crossmodal relative to within-modal recognition was predicted, due to the lack of transfer of sensory specific face information. In Experiment 4 differences in exploratory procedures between the modalities were controlled for during the encoding of face information. Experiment 5 investigated whether featural or configural information was shared across modalities by constraining the type of information available for recognition.

3.2 Experiment 3

In this experiment crossmodal face recognition was examined and it was predicted that 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, an attempt was made 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.

3.2.1 Method

3.2.1.1 Participants

Twenty-four undergraduate students from Trinity College Dublin (mean age 20.7 years, SD= 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.

3.2.1.2 Apparatus and Stimuli

Twenty-four plaster face models were used as haptic stimuli in this experiment (see Chapter 1 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 this 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 (e.g. Baddeley and Woodhead, 1983; Blanz et al., 1999).

3.2.1.3 Design

This experiment was based on a mixed design with learning modality (vision or touch) 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: 8 target and 16 non-targets in the haptic block; 12 targets and 12 non-targets in the visual block. Although the number of targets in the haptic block was reduced to aid memory, the total number of face stimuli was maintained 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. Haptic stimuli were presented for 1 minute whilst visual stimuli were presented for 1 second (for rationale behind presentation times see Section 2.3.1.3, pg. xx). Performance was measured in terms of accuracy rates (hits plus correct rejections) based on 24 trials in each modality.

3.2.1.4 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 minutes in total. Each haptic target face was presented for 1 minute until all faces were perceived. This sequence was repeated four times. This procedure ensured that the same amount of time was spent encoding each haptic stimulus. A

learning criterion of 7/8 correct responses was set for haptics and 11/12 correct responses was set for vision. The learning block was repeated until this response criterion was achieved. On average the learning of the haptic targets took about 50 minutes to complete whereas learning of the visual targets took approximately 7 minutes.

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 modal 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 second followed by a visual face images 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 minute. 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.

3.2.2 Results

Figure 3.1 depicts the mean percentage accuracy responses for the within and across modalities face recognition task. Given the different probability levels for the visual and haptic familiarisation conditions, a 2×2 χ^2 test was used to examine possible differences across the learning conditions and found no difference between visual or haptic learning [$\chi^2=0.39$, $df=1$, n.s.]. A Wilcoxon matched-pairs tests revealed a cost in crossmodal recognition relative to within modal recognition regardless of whether faces were learned in vision [$T(12)=0.001$, $p<0.002$] or haptics [$T(12)=9.5$, $p<0.02$]. Although performance in the cross modal condition was relatively poor, it was found to be significantly better than chance for both visual [$\chi^2=21.87$, $df=11$, $p<0.03$] and haptic [$\chi^2=218.16$, $df=11$, $p<0.01$] learning.

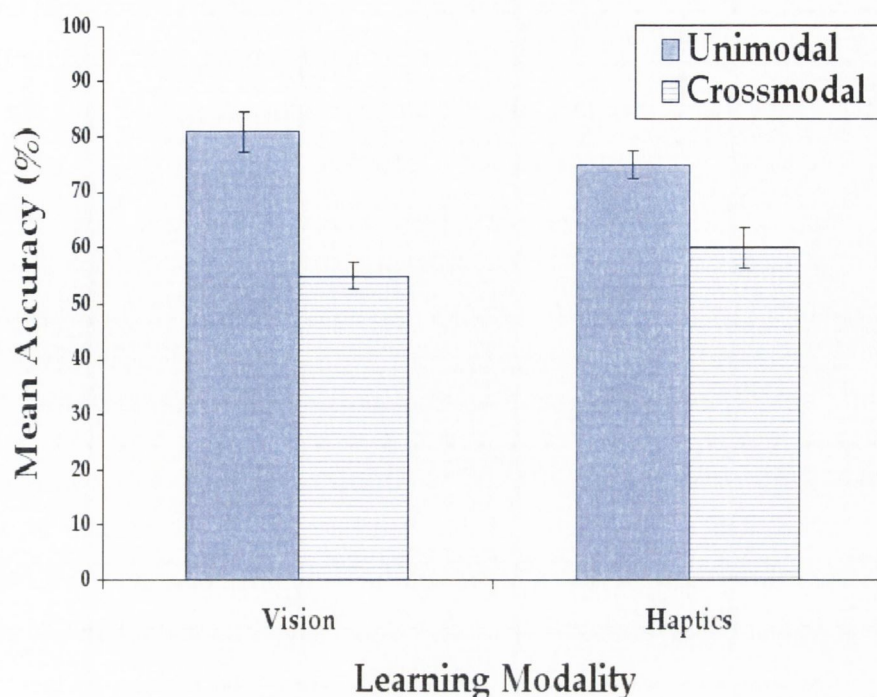


Figure 3.1. Plot showing the mean percentage of correct responses for face recognition both within and across modalities. Error bars represent ± 1 standard error of the mean.

3.2.3 Discussion

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

The 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 more so on subtle differences in modality-specific face information.

However, the manner in which face information was encoded was also 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 and 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 and Lederman, 1986; Loomis, Klatzky, & Lederman, 1991). These differences in encoding procedures may have affected representations in memory. In the next experiment I attempted to eliminate encoding differences during familiarisation, in order to investigate whether or not encoding impacted on recognition accuracy in the crossmodal conditions.

3.3 Experiment 4

In this experiment it was 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 then no cost in crossmodal relative to within modal recognition performance would be expected. Here visual face images were presented in parts to promote more serial encoding as in haptics.

3.3.1 Method

3.3.1.1 Participants

Twenty-four undergraduate students (mean age 22.25 years, SD= 3.05, 13 female) participated in this experiment for pay. Participants had normal or corrected-to-normal vision and none reported any tactile impairment. None of these participants took part in the previous experiment.

3.3.1.2 Apparatus and Stimuli

Sixteen of the plaster face models and corresponding full-face colour images used in Experiment 3 were randomly selected and used as stimuli in this experiment. The individuals used as models were unfamiliar to the participants. Using Adobe Photoshop 5.5 for the Macintosh, each visual face image was divided 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 are illustrated in Figure 3.2. They 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 as visual stimuli in the learning

session of the experiment. The whole face image was presented during the recognition task. Haptic stimuli were presented behind an occluder to the left of the monitor.

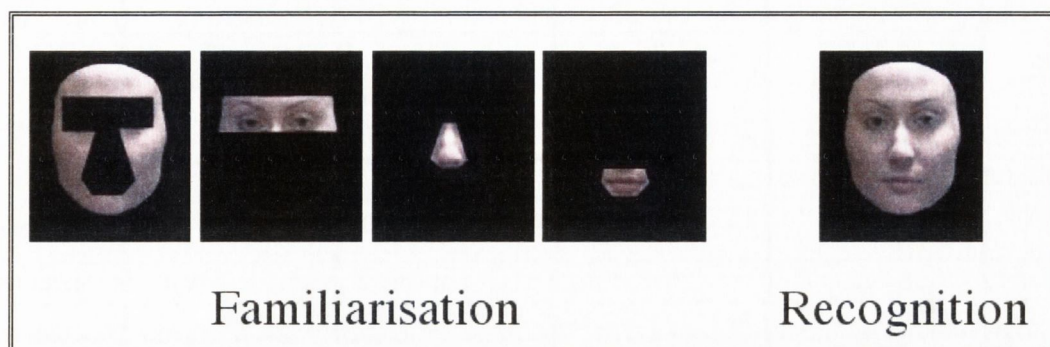


Figure 3.2. An illustration of the type of visual stimuli used in the learning and test phases of Experiment 4.

3.3.1.3 Design

This experiment was based on a 2 x 2 mixed design with learning modality (vision, haptics) as a between subjects factor and recognition modality (within or across) as a within subjects factor. A recognition block followed a learning session. During the recognition session, the within modal and crossmodal trials were blocked and each comprised 16 trials: 8 target trials and 8 non-target trials per block. These blocks were counterbalanced across participants and trials were randomly presented within each block. Haptic stimuli were presented for 1 minute whilst visual stimuli were presented for 1 second (for rationale behind presentation times see Section 2.3.1.3, pg. xx). Performance was measured in terms of accuracy rates (hits plus correct rejections) based on 16 trials in each modality.

3.3.1.4 Procedure

As in Experiment 3, participants were required to complete a learning task prior to the experiment. Each participant learned a set of 8 faces either in the visual or haptic modality. In the visual learning modality a target face was presented as four face parts and each part was randomly presented for 5 seconds within a single trial. This timing was arbitrarily chosen to facilitate learning; ultimately faces had to be learned to a set criterion level before participants could proceed to the recognition phase. Target faces were repeated four times during visual learning to approximate equivalent learning conditions between the visual and haptic conditions. As in Experiment 3, haptic target

faces were each presented for 1 minute during learning and repeated 4 times. A learning criterion of 7/8 correct responses was set in both learning blocks before the participant could proceed to the experiment. On average, learning of the haptic stimuli took about 50 minutes whereas learning of the visual stimuli took approximately 30 minutes. 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.

During the recognition phase, the task for the participants 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. The test procedure was the same as that in Experiment 3.

3.3.2 Results

The mean percentage accuracy scores (hits and correct rejections) for the within and crossmodal recognition conditions are illustrated in Figure 3.3. Although crossmodal performance was poor, performance was nonetheless greater than chance [$\chi^2 = 72.75$, $df=23$, $p<0.001$]. Using a 2 x 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 no effect of learning modality was found [$F(1,22) = 2.75$, n.s.]. However, there was a main effect of recognition task [$F(1,22) = 27.32$, $p < 0.01$] with better performance in the within-modal than the crossmodal recognition task. There was no interaction between the factors, [$F(1,22) = 1.56$, n.s.].

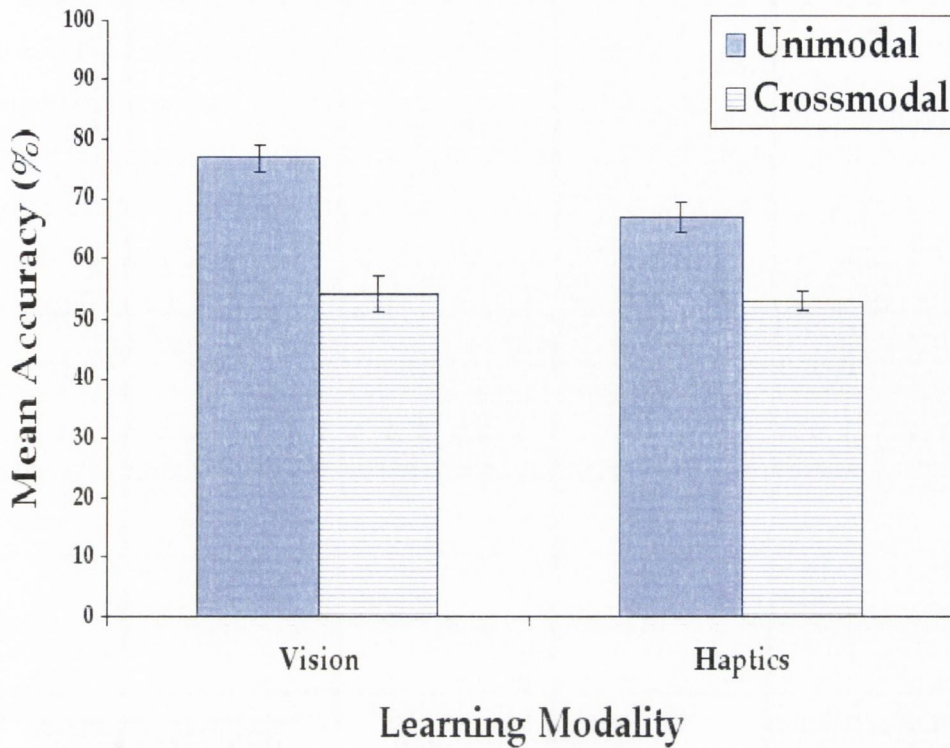


Figure 3.3. Plot showing the mean percentage correct responses for the within modal and crossmodal recognition of faces. Error bars represent ± 1 standard error of the mean.

3.3.3 Discussion

The results suggest that encoding differences across the modalities did not account for the cost in crossmodal recognition relative to within modality face recognition. In this experiment, 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 cross-modal recognition. Nevertheless, faces were recognised across modalities therefore the next question we addressed was what is the nature of the face information that is shared across modalities.

3.4 Experiment 5

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, using an alternative paradigm to inversion, whether configural face information, or featural information alone, is shared across vision and haptics.

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 Figure 3.4).

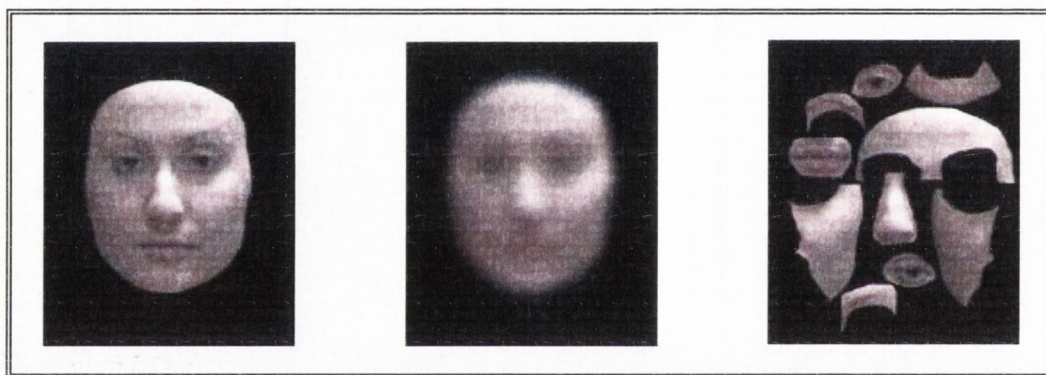


Figure 3.4. An illustration of the type of visual face stimuli used in Experiment 5. An example of an intact, blurred, and scrambled image are presented from left to right.

Collishaw and Hole (2000) found that visual 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 it appears that scrambled images are no more difficult to recognise visually than inverted images. In both instances recognition is disrupted due to an inability to process the appropriate configural face information. In this experiment, scrambled face images were substituted for inverted face images, so that featural information in all image types was

consistent (i.e. upright). By blurring a face image detailed featural information was removed. The role of featural and configural face information could therefore be studied independently (Collishaw and Hole, 2000; Schwaninger et al., 2002). It was 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.

3.4.1 Method

3.4.1.1 Participants

Twenty-four undergraduate students (mean age 24.29 years, SD= 9.08, 18 female) participated in this experiment for research credits. All were naïve to the task and none had participated in the previous experiments. All participants reported normal or corrected-to-normal vision and none reported any tactile impairment.

3.4.1.2 Apparatus and stimuli

The models in this experiment were unfamiliar to the participants. Sixteen plaster face models were used as haptic stimuli and corresponding colour images of the individuals from whom the face models were cast were used as visual stimuli. These intact images were copied and, using Adobe Photoshop 5.5, scrambled and blurred versions of each face were created. Scrambled stimuli were created by isolating the features and rearranging these into random configurations (see Schwaninger et al., 2002 for a similar procedure). Blurred images were created by applying a Gaussian blur of radius of 8 pixels to each face image. This blur factor was determined in a previous pilot experiment where several blur factors were compared.

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 the participant.

3.4.1.3 Design

This experiment was based on a one-way, repeated measures design with visual face match (intact, scrambled, 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 8 same and 8 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. Haptic stimuli were presented for 1 minute (for rationale behind presentation times see Section 2.3.1.3, pg. xx). Unlike in the previous experiments, responses here were always to visual stimuli thus allowing performance to be measured in terms of both matching accuracy (mean percentage of hits plus correct rejections on 16 trials) and reaction times.

3.4.1.4 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 minute later. The participant was then required to look at the computer monitor. A fixation cross appeared in the centre of the screen for 1 second. 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.

3.4.2 Results

Matching accuracy was measured in terms of hits plus correct rejections on 16 trials. The mean percentage accuracy scores ($\pm$ 1 standard error of the mean in parentheses) for the intact, blurred and scrambled face types respectively were 56.14% (2.53); 56.80% (1.77); and 57.32% (2.57). The accuracy data were analysed using a one-way repeated measures Friedman's ANOVA which revealed no effect of image type [$\chi^2 = 1.17$, $df = 2$, n.s.]. Performance to the matching trials was greater than chance [$\chi^2 = 243.75$, $df = 71$, $p < 0.001$].

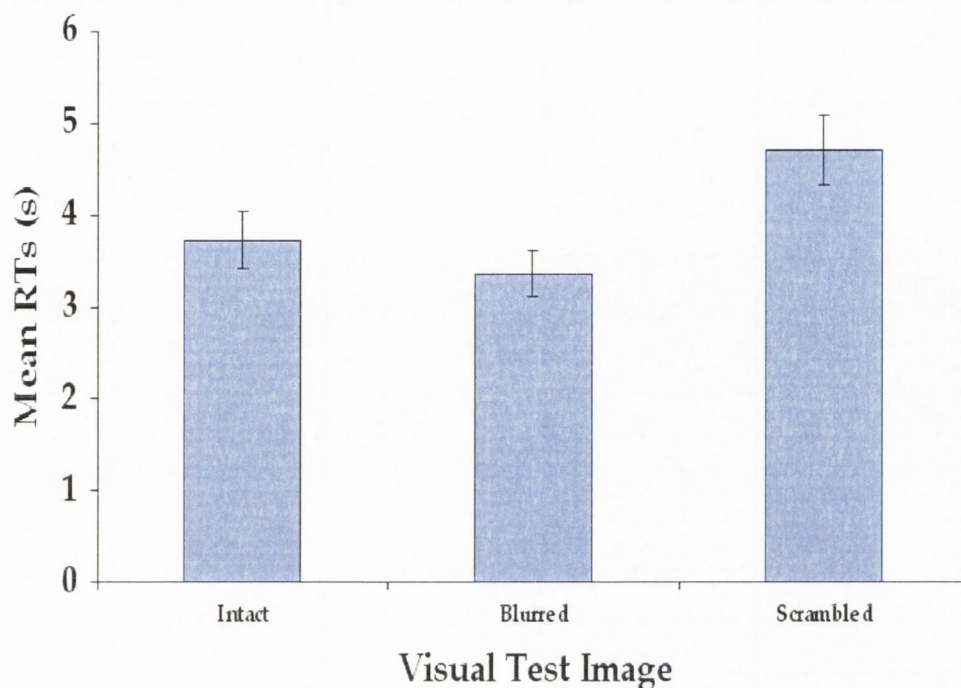


Figure 3.5. Plot showing the mean reaction times for the correct responses to each of the visual face types. Error bars represent ± 1 standard error of the mean.

Prior to analysis, response times ± 2.5 standard deviations 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 Figure 3.5. 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 scrambled images (both $p < 0.01$). There was no difference in response times between the intact and blurred images.

3.4.3 Discussion

Since matching haptic face stimuli to either intact or blurred visual face stimuli was faster than matching to scrambled face images, the 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. The results of this experiment suggest that despite differences in the ability to encode spatial information across vision and touch, configural face information facilitates crossmodal face matching. Using an alternative paradigm to inversion (cf. Maurer et al. 2002), these results provide further evidence of the importance of configural processing for face recognition in both the visual and haptic modalities, and contribute to the notion of a domain-specific processor for faces.

3.5 General Discussion

The aim of this chapter was to investigate whether visual and haptic face information encodes similar structural information to allow for efficient recognition of faces across these modalities. In summary, I found that face information represented in memory is largely modality-specific although some information can transfer across modalities (Experiment 3). I also found that these modality specific face representations were maintained even when differences in exploratory procedures during encoding were controlled (Experiment 4). The type of structural descriptions that can be shared across the modalities was investigated in Experiment 5. The 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 and Klatzky, 1990) it might be assumed that a common, crossmodal representation of face structure might facilitate crossmodal face recognition. However, the results imply that structural information alone is not sufficient for crossmodal face recognition and instead suggest an important role for modality-specific information during face recognition. Such 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 and 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 and Langton, 1994, Kemp, et al., 1996; Yip and Sinha, 2002; Vuong, Peissig, Harrison, & Tarr, 2005). In my 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. It is possible that if the visual face images were uniformly coloured (i.e. images of the facemasks) this may have resulted in better crossmodal performance. However, I decided against using such images in order 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 and Lederman, 2002). In the above experiments, such material properties were non-diagnostic of face identification since all of the haptic face stimuli were made of the same material. Instead, other haptic-specific cues may have been encoded, (possibly relating to the 3-dimensional structure of each face stimulus or poor resolution of facial features as suggested in Experiment 5), thus rendering within modal haptic recognition more efficient than cross modal recognition.

The behavioural findings of haptic face recognition studies suggest that face recognition may be a multisensory phenomenon, such that visual and haptic face recognition are possibly mediated by shared neural substrates (Kilgour et al., 2004; Kilgour and Lederman, 2002). In contrast, the results of Experiments 3 and 4 suggest that face recognition is not completely multisensory. It appears that some face information may be modality-specific and necessary for optimal within modal recognition. These 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, Furey, Riccardi, 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; Haxby, Horowitz, Ungerleider, et al., 1994; Hoffman and 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)^{3.1}. Given the findings of Experiments 3 and 4, 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 shape 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 an earlier region of visual cortex. I will return to this point in the General Discussion.

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 arrangement of facial features is more important for faces than other objects. On the other hand, research findings from Diamond and Carey (1986), among others (e.g. Gauthier and 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 this chapter, the experimental results may nevertheless lend indirect evidence to the idea that information processing is face-specific. In particular, the results of Experiment 5, obtained using an alternative method to inversion, 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

^{3.1} A separate study by Kilgour, Kitada, Servos, et al. (2005) found that haptic identification of facemasks evoked greater activation in the left fusiform gyrus relative to the identification of non-face control stimuli. Although visual face recognition imaging studies typically report greater activation in the right fusiform gyrus some do report smaller activations in the left fusiform gyrus. However, as Kilgour et al. did not include a visual face recognition task, it is not clear whether areas of haptic and visual activations in the left fusiform overlap.

independently of expertise and as such this result reflects face-specific rather than general object information processing per se.

3.6 Conclusions

It appears that the visual and haptic modalities process faces in a similar manner. However, qualitative differences in the nature of the information encoded promotes more efficient within modal relative to crossmodal face recognition.

Chapter 4

Face priming across the visual and haptic modalities

Abstract

Face recognition, as with general object recognition, appears to be subserved by two dissociable forms of memory. Explicit memory requires the conscious recollection of both perceptual information and modality-specific information that may not be optimally represented across modalities. Implicit memory, on the other hand, is thought to rely on perceptual information, namely the structural descriptions of objects or faces in memory. As the visual and haptic systems can both encode information pertaining to facial structure, it is possible that vision and haptics share perceptual descriptions of faces. In this chapter I report an implicit priming task where participants were primed either within modalities (i.e. visual face prime) or across modalities (i.e. haptic face prime). However, there was no evidence of a benefit of primed versus unprimed facemasks. Furthermore, there was no evidence of within modal visual priming, in that performance on the primed and unprimed facemasks was equivalent. Therefore, it is not possible to draw any definitive conclusions regarding putative memory representations of faces that are common to both the visual and haptic modalities.

4.1 Introduction

Researchers have long been aware that two distinct forms of memory underpin the visual recognition of words and objects, those being explicit and implicit memory (e.g. see Roediger and McDermott, 1993; Schacter and Tulving, 1994; Willingham, 1997). It is argued that explicit memory for general objects combines perceptual information with other object-related information such as semantic or contextual information, whereas implicit memory is thought to rely on perceptual object representations (Schacter, 1994; Tulving and Schacter, 1990). Explicit recollection of an object requires conscious effort, but implicit memory, on the other hand, occurs in the absence of conscious recollection and is typically evidenced through repetition

priming^{4.1}. A similar dissociation of memory systems appears to exist for visual face recognition (e.g. Bauer, 1984; de Haan, Young, & Newcombe 1987; Young and de Haan, 1992), and a substantial number of studies have demonstrated repetition priming, and thus, implicit memory for faces (e.g. Bruce, Burton, Carson, et al., 1994; Bruce, Carson, Burton, & Kelly, 1998; Bruce and Valentine, 1985; A. Ellis, Flude, Young, & Burton, 1996; A. Ellis, Young, Flude, & Hay, 1987).

It is thought that priming largely reflects the operations of a perceptual representation system (PRS) (Schacter, 1990; 1992; Tulving and Schacter, 1990). The PRS comprises domain-specific subsystems that process and represent information about the form and structure of words and objects (Schacter, 1994). For example, visual repetition priming has been found to occur in response to perceptually matched person information, that is, face images prime other images but written names do not prime face images and vice versa (Bruce and Valentine, 1985; A. Ellis, et al., 1987; A. Ellis, et al., 1996). It has been argued that structural descriptions of objects form a sub-system of the PRS and a number of research findings indicate that the structural-description system mediates visual object priming. For instance, visual investigations of implicit and explicit memory have found that changes between study and test affecting contextual or surface properties of an object (e.g. changes in size or colour) without disrupting structural properties, affect explicit recognition of that object but not implicit recognition as assessed through priming (e.g. Cave, Bost, & Cobb, 1996; Cooper, Schacter, Ballesteros, & Moore, 1992).

The visual and haptic systems can both acquire shape information for recognition (e.g. Lederman and Klatzky, 1990). Moreover, the visual and haptic modalities both appear to process faces configurally (Experiment 5; Kilgour, de Gelder, & Lederman, 2004), with configural information mediating more efficient crossmodal face matching (Experiment 5). This evidence suggests that both modalities may encode structural descriptions of faces, and it is conceivable that a single structural description might be common to both modalities. Recent studies using explicit measures of crossmodal object recognition have found within modal recognition to be significantly better than crossmodal recognition. However, crossmodal priming studies with 2-D patterns and 3-

^{4.1} Repetition priming is defined operationally as the facilitation of a response to a stimulus to which an individual has recently been exposed (e.g. Tulving and Schacter, 1990).

D objects have found that the perceptual representations mediating priming are modality independent (Easton, Greene, & Srinivas, 1997; Reales and Ballesteros, 1999). Crossmodal priming effects for common objects were found to be similar in magnitude to within modal priming effects, indicating that visual and haptic object recognition is underpinned by shared structural object representations. Therefore, it is possible that vision and haptics may share perceptual descriptions of faces, but that explicit crossmodal recognition of faces is poorer relative to within modal recognition due to context-dependent, modality-specific properties that are not optimally represented across both modalities (cf. Reales and Ballesteros, 1999; Chapter 3).

The aim of this chapter was to discover whether or not perceptual descriptions of faces are indeed common to vision and haptics, using a crossmodal repetition priming paradigm. In the visual face recognition literature, the most commonly used task for assessing implicit memory for faces is the familiarity decision task (Bruce and Valentine, 1985). In such a task, participants are first exposed to a set of familiar faces in an unrelated task, for example in a sex discrimination task (e.g. A. Ellis et al., 1987; H. Ellis, Ellis, & Hoise, 1993). After a short interval, they are then asked to make speeded familiarity decisions to a larger set of faces, comprising the exposed (i.e. primed) familiar faces, unprimed familiar faces, and a set of unfamiliar faces. 'Familiar' responses to primed familiar faces are generally faster and more accurate than responses to unprimed familiar faces. In Experiment 6 I adapted this paradigm so that participants were familiarised with some faces visually and some faces haptically prior to test. If the visual and haptic modalities share perceptual representations of familiar faces then it was hypothesised that pre-exposure to a haptic face stimulus would prime the identity of its visual counterpart and vice versa. Unlike in explicit tasks, repetition priming is thought to be impervious to modality-specific face information (e.g. Reales and Ballesteros, 1999), so the magnitude of facilitation should be equivalent across conditions^{4.2}.

4.2 Experiment 6

^{4.2} As priming effects are often in the order of milliseconds it was decided to investigate haptics-to-visual crossmodal priming only, since priming effects may be obscured in a vision-to-haptics priming task due to the time taken to disengage from a haptic face stimulus in order to respond.

4.2.1 Method

4.2.1.1 *Participants*

Sixteen undergraduate students from Trinity College Dublin (mean age 24.81 years, SD= 5.78, 10 female) participated in this experiment for a nominal payment. In all cases, participants reported normal or corrected-to-normal vision and none reported any tactile impairment. The testing procedure of this experiment was approved by the School of Psychology Ethics Committee, Trinity College Dublin. Written consent was obtained from all participants prior to conducting the experiment.

4.2.1.2 *Apparatus and Stimuli*

Sixteen plaster face models were used as target haptic stimuli in this experiment (for details of the procedure used in the casting of these faces please refer to Chapter 2). Digital colour images of the live faces of the 16 individuals from whom the face models were cast were used as target visual stimuli. A different set of 8 facemasks and another different set of 20 visual face images (4 of which were used in the priming task and not repeated in the familiarity decision task) were used as non-target stimuli. All of the individuals used as models in this experiment were previously unfamiliar to the participants. Visual stimuli were presented using SuperLab Pro on a Macintosh G4 computer. The images presented on the computer screen subtended a visual angle of approximately 10° vertically and 7° horizontally. Haptic stimuli were presented behind an occluder to the left of the computer. Correct responses and reaction time data were recorded using a button box with colour-coded keys placed directly in front of the participant, the familiar key being the button on the right.

4.2.1.3 *Design*

This experiment was based on a 2 x 2 repeated measures design with learning modality (vision or haptics) and prime condition (primed, unprimed) as factors. Unfamiliar faces were included merely to create the task demands of the familiarity decision but they were not directly related to the experimental hypothesis about the priming of familiar faces across modalities. Therefore, although accuracy and reaction time data for the unfamiliar faces are reported, they were not included in the data analysis (see Bruce, Carson, Burton, & Kelly, 1998 for a similar method).

The experiment was structured around two sessions; a learning session in which participants were familiarised with target stimuli, and a priming/test session. For the learning session, the 16 face stimuli were randomly divided into two learning groups of

8 faces, 4 male and 4 female faces in each. Participants were assigned one group as visual target stimuli, and the other group as haptic targets. The learning groups were counterbalanced across participants, as was the order of visual and haptic learning.

The priming session comprised 3 separate parts, namely a visual and haptic priming task, a 3 minute distractor task, and a familiarity decision task. In the priming task participants were asked to make male/female judgements for a set of 8 visual and 8 haptic face stimuli. The visual and haptic stimuli were blocked, and each block contained 4 target faces (primes) from the previous learning session and 4 unfamiliar faces that were used as fillers in this part of the experiment only. The blocks were counterbalanced and the presentation of face stimuli was randomised across participants. The number of correct responses and reaction times were recorded. After the priming task participants took part in a 3 minute distractor task in which they were required to write down as many world cities as possible in the time available. The distractor task was included to prevent the generation of representations of the prime stimuli in explicit memory.

Directly after the distractor task, participants were required to perform a task where they had to make speeded familiarity decisions to a series of 32 visual face stimuli which comprised the 2 groups of 8 visual and 8 haptic target faces learned in the learning session (4 primed and 4 unprimed face stimuli in each group), and a new set of 16 unfamiliar faces. This task was performed in vision only because of the technical difficulties in recording reliable timing responses to haptic stimuli. In implicit tasks, test stimuli are typically presented in a perceptually degraded form to circumvent possible contamination by explicit learning (McDermott, 2001) therefore, in this task visual stimuli were presented for only 750 ms. The order of the face trials was pseudo-randomised across participants. Performance was measured in terms of the mean percentage of correct responses (based on 4 trials per primed and unprimed familiar crossmodal and within modal conditions, and 16 trials in the unfamiliar vision condition), and reaction times.

4.2.1.4 Procedure

Participants began this experiment by completing a familiarisation or learning session. Familiarisation consisted of learning the name and occupation of each face stimulus and subsequently identifying each face by both its name and occupation. In order to reduce the memory demands in the task we chose familiar, simple names (e.g.

Jill, Frank, etc.) and simple occupations (e.g. teacher, dentist, etc.). The visual target images were presented as laminated cards and participants were required to study these images for 5 minutes in total. Each haptic target face was presented for 1 minute until all 8 faces were perceived and this sequence was repeated four times initially. A learning criterion of 100 % accuracy was set in both the visual and haptic learning tests before participants could proceed to the next session. The familiarisation process was repeated until this criterion was achieved. On average the familiarisation of the haptic stimuli took about 55 minutes to complete whereas familiarisation of the visual stimuli took approximately 7 minutes.

The day after the learning session, participants were recalled for the priming task and test session. In the priming task participants had to indicate whether a face was female or male by pressing the appropriate key on a button box placed directly in front of the participant. The priming task was blocked according to modality, with 8 face stimuli presented pseudo-randomly in each block (i.e. 8 faces presented visually and 8 faces presented haptically). The order of the priming blocks was counterbalanced across participants. In each block, 4 of the faces were from the original learning list (i.e. were familiar) and 4 were unfamiliar. Different unfamiliar face stimuli were used in the visual and haptic blocks. For each visual trial, a 1 second, central fixation cross appeared on the screen prior to the face image. A face stimulus was then presented and participants were required to respond as fast and as accurately as possible. In the haptic block, facemasks were presented behind an occluder. In a haptic trial, a 500 ms tone from the computer signalled the participant to begin exploring the facemask. Participants felt the face stimulus for 1 minute. A second 500 ms tone indicated to the participants to cease haptic exploration and to respond as fast and as accurately as possible using the button box.

After the priming task was completed, participants performed a distractor task. They were asked to write down as many world cities as they could within a 3 minute time period. Typically, participants listed 37 cities indicating that they were conducting this task appropriately.

Participants were then required to perform the test session which was the actual experiment itself. They performed a speeded familiarity judgement on a set of visual face stimuli comprising images of 16 familiar faces and a new set of 16 unfamiliar faces. The familiar face trials comprised the 8 visual target faces and the corresponding images

of the 8 haptic target faces, and in each target group 4 face trials had been primed in the sex-discrimination task, and 4 were unprimed. Face images were presented pseudo-randomly. In any one trial, a face image was preceded by a 1 second central fixation cross. A face image was then presented for 750 ms, and participants were prompted to respond 'familiar' or 'unfamiliar' by pressing the appropriate key on a button box. Participants were urged to complete this task as quickly and as accurately as possible.

4.2.2 Results

Figure 4.1 depicts the mean percentage of correct responses for the within and crossmodal familiarity decision task. Descriptive data for the unfamiliar trials are presented for illustrative purposes but are not included in the analyses for the reasons outlined above. Data were analysed using a 2 x 2 repeated measures ANOVA with priming modality (across or within) and prime (primed or unprimed) as the factors. There was a main effect of priming modality, with significantly more correct responses for within modalities relative to crossmodal priming [$F(1,15)= 77.11$, $p< 0.0001$]. However, there was neither a main effect of prime [$F(1,15)= 0.15$, n.s.], nor an interaction between the factors [$F(1,15)= 2.14$, n.s.]. Overall performance for within modal familiar, and unfamiliar faces was found to be significantly better than chance (50%) [$\chi^2= 506.25$, $df=15$, $p<0.0001$; $\chi^2= 337.50$, $df= 15$, $p<0.0001$, respectively]. However, overall performance in the crossmodal condition was significantly worse than that expected by chance [$\chi^2= 375.00$, $df= 15$, $p<0.0001$].

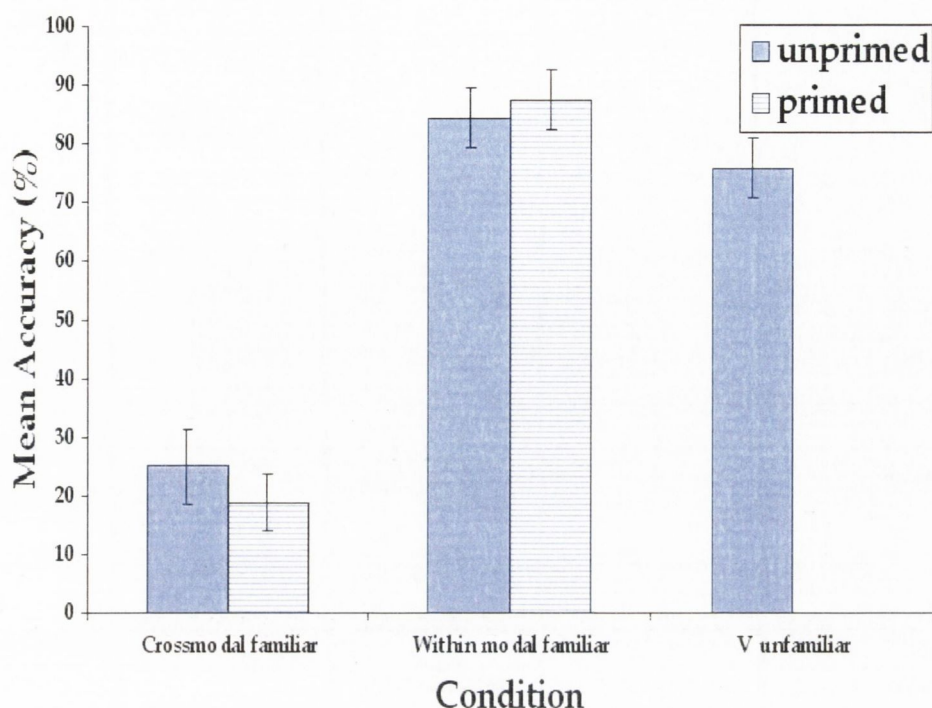


Figure 4.1. Plot showing the mean percentage of correct 'familiar' responses for the crossmodal and within modal conditions and the correct 'unfamiliar' responses to new visual face stimuli. In the familiarity decision task familiar faces were either primed or unprimed, although no effect of priming was found. Error bars represent ± 1 standard error of the mean.

Prior to analysis, response times ± 2.5 standard deviations from the mean for each participant were removed. In total, 2.3% of responses were excluded. The mean reaction times for the remaining correct responses are illustrated in Figure 4.2. A number of participants failed to respond correctly, resulting in a number of missing data (11 points). Therefore, in order to analyse the data using a 2 x 2 ANOVA, the mean reaction times for each condition were substituted into the blank cells. In this way it was possible to retain and analyse the data for all participants without affecting the overall means for each condition. There was a main effect of modality [$F(1,15) = 20.03$, $p < 0.001$] with significantly faster latencies in the within modal than crossmodal conditions. However, there was no difference in reaction times for primed and unprimed stimuli [$F(1,15) = 0.09$, n.s.]. There was no interaction between the factors [$F(1,15) = 1.44$, n.s.].

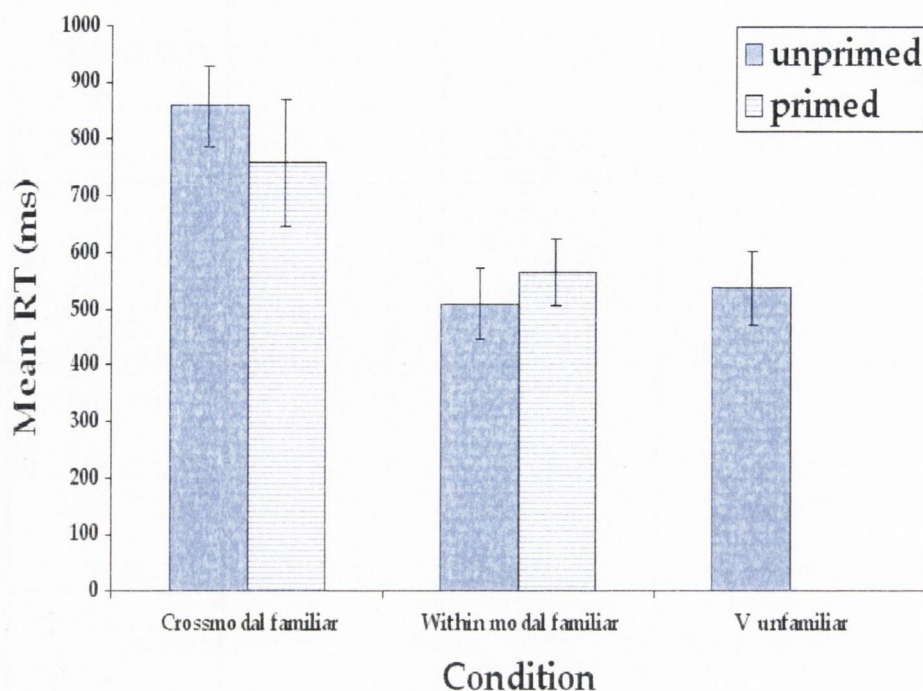


Figure 4.2. Plot showing the mean reaction times for 'familiar' responses for the crossmodal and within modal conditions and the correct 'unfamiliar' responses to new visual face stimuli. In the familiarity decision task familiar faces were either primed or unprimed, although no effect of priming was found. Error bars represent ± 1 standard error of the mean.

4.3 General Discussion

Previous studies investigating crossmodal priming of 3-D objects have found equivalent facilitation for objects primed within and across modalities relative to unprimed objects in an object naming task (Easton et al., 1997; Reales and Ballesteros, 1999). This suggests that the visual and haptic systems share perceptual descriptions of objects. This experiment was designed to investigate whether similar shared perceptual descriptions of faces might exist for the visual and haptic modalities. The underlying hypothesis was that if the haptic and visual modalities share structural descriptions of faces, then pre-exposure to a haptic face stimulus would prime a familiarity decision for its visual face counterpart to the same extent as a visual prime. However, there was no evidence of equivalent crossmodal as within modal priming in a familiarity decision task. Overall, performance for the crossmodal condition (i.e. haptically learned faces) was poor, with no difference in either response accuracy or reaction times for primed relative to unprimed faces. Furthermore, contrary to findings in the visual literature,

there was no evidence of within modal priming in the familiarity decision task for primed relative to unprimed visual face stimuli. Since there was a failure to demonstrate even within modal priming, it is not possible, on the basis of these data, to draw any definitive conclusions regarding putative perceptual descriptions of faces common to both the visual and haptic modalities.

The results of this experiment raise a number of questions for discussion. The first obvious question is why was there a failure to demonstrate within modal visual face priming? The second question is why was performance in the crossmodal condition so poor overall? Regardless of the performance for crossmodal priming, it was expected that within modal priming would be observed. However, although overall visual performance was good ($> 85\%$), this experiment failed to demonstrate within modal face priming. A large number of studies have successfully demonstrated response facilitation for visually primed faces using a variety of priming and test methods (e.g. Bruce, Burton, Carson, et al., 1994; Bruce, Carson, Burton, & Kelly, 1998; Bruce and Valentine, 1985; A. Ellis, Flude, Young, & Burton, 1996; A. Ellis, Young, Flude, & Hay, 1987). Here I adopted reported methodology which has successfully resulted in priming effects, therefore it is unclear why response facilitation in the within modal condition was not found in this experiment. However, a number of practical limitations may have influenced performance in the familiarity decision task.

Firstly, the design of this experiment was constrained by a number of factors, namely the size of the available stimulus set, and stimulus counterbalancing. Usually visual priming experiments use a large number of visual stimuli, with at least 8 face stimuli in each of the primed and unprimed conditions (e.g. Bruce et al., 1998). In this experiment however, the number of facemasks and corresponding visual face images available for use in the learning session was limited by the overall number of haptic face stimuli available. Thus, to ensure that no unfamiliar stimulus was encountered more than once in priming and test tasks and to allow counterbalancing of experimental learning lists across participants, the number of faces in each learning condition was reduced to only 8 stimuli, 4 primed, and 4 unprimed. Therefore, given the small number of data points per participant for each condition, the analysis may not have been powerful enough for significant facilitation effects to emerge.

On the other hand, it may be that the degree of familiarity with the face stimuli was not adequate enough to elicit repetition priming effects in the familiarity decision

task, despite a criterion level of 100% accuracy. A number of studies suggest that repetition priming is due to changes within the long-term representation system that responds to the familiarity of a face stimulus (Burton et al., 1990; A. Ellis, Young, & Flude, 1990; see also Jacoby and Hayman, 1987). When a familiar face is encountered, its corresponding FRU and personal identity node (PIN, the unit wherein person-related information is stored) become active and the link between the FRU and the PIN becomes strengthened. When the face is subsequently encountered it will take a shorter time for the PIN to become active, thus facilitating quicker reaction times in a familiarity decision task (Burton et al., 1999). Visual investigations into repetition priming typically use famous faces as familiar stimuli. Famous faces are generally recognisable across multiple viewpoints (A. Ellis et al, 1987), and under varying lighting conditions, and are frequently encountered as dynamic face stimuli (e.g. on television or in films). Therefore, it is reasonable to assume that structural representations of famous faces (i.e. the FRUs) are constructed from a large number and variety of exposures and consequently are very robust. Thus, pre-exposure to a famous face may activate a robust memory representation, facilitating responses in later tasks. Indeed, it has been found that repetition priming for famous faces has a low decay rate (e.g. Bruce et al., 1994), with no reduction in the amount of priming between delays of 10 minutes and 1 week, even when image formats differed between priming and test. The priming of unfamiliar faces, however, has only been found where face stimuli are presented sequentially, that is, with no intervening face stimuli between priming and test (Bentin and Moscovitch, 1988). As unfamiliar faces have no pre-existing representation in memory it was postulated that the representation generated during the first encounter was not sufficient for the unfamiliar stimuli to sustain repetition effects after long intervals.

As my stimulus set comprised faces previously unfamiliar to participants, in this experiment the familiar stimuli were newly learned, to ensure that participants were familiar with the visual and haptic face stimuli to the same extent. Although face stimuli were learned to a specified criterion level of 100% accuracy, the structural description for each face was generated from a relatively small number of exposures and only one viewpoint. Consequently, although repetition priming performance might be expected to be better for familiar than for completely unfamiliar faces, it is possible that it might be poorer than repetition priming of famous faces, as the underlying structural descriptions are not as robust. That is, the memory representations activated in the

priming task may not have been robust enough to withstand the intermittent duration between prime and test (c. 3 minutes) in this experiment. Perhaps the lack of within modal repetition priming in this experiment reflects a qualitative difference between representations of newly-learned and highly familiar faces (cf. Clutterbuck and Johnston, 2002). Future research might consider using a variety of viewpoints during the familiarisation (learning) phase, in an attempt to generate more robust structural descriptions for newly-learned faces. Alternatively, future studies might systematically investigate the time course of visual repetition priming for newly-learned, versus recently familiarised, versus well known faces (cf. Benton and Moscovitch, 1988).

The reasons outlined above might also account for the lack of response facilitation for primed relative to unprimed crossmodal familiar face stimuli. However, it is unclear why crossmodal performance in general was so poor relative to within modal performance, and chance performance. A subsequent analysis of the responses from the priming task indicated that participants were significantly better at haptically judging the sex of the familiar relative to the unfamiliar face stimuli, suggesting that participants did have stored representations of the learned faces in memory^{4.3}. Even if the visual and haptic systems don't share perceptual descriptions of faces, performance for familiar faces would still be expected to be at least at chance level. Contrary to this expectation, familiarity decisions for crossmodal faces were significantly below chance, indicative of a possible response bias. It might be possible that within the confines of the familiarity decision task, participants did not consider crossmodal faces as relevant 'familiar' stimuli. In order to prevent explicit memory contamination in the task, participants were instructed to consider as familiar any recently 'experienced' face stimulus. However, as the familiarity decision task was conducted in vision only, participants may only have considered recently encountered visual stimuli as familiar (this interpretation stems from verbal reports from participants after the task was

^{4.3} As an aside, data (correct responses) collected during the sex judgement priming task were analysed. Using a 2(modality) x 2(familiarity) x 2(sex) repeated-measures ANOVA, a main effect of both modality [$F(1,15)= 17.75, p< 0.001$] and familiarity [$F(1,15)= 11.14, p< 0.01$] was found. Visual performance was significantly better than haptic performance (close to ceiling) and the number of correct judgements was significantly greater for familiar than unfamiliar faces. There was no effect of sex, nor any significant interactions between any of the factors. Overall, visual and haptic performance was well above chance levels [$\chi^2= 565.63, df=15, p<0.0001$; $\chi^2= 237.50, df= 15, p<0.0001$, respectively].

completed). In other words, participants may have expressly refrained from responding familiar to haptically learned and primed stimuli.

4.4 Conclusions

Unfortunately, the results of this experiment do not easily allow for any conclusions to be drawn about the existence of structural descriptions of faces common to both the visual and haptic modalities at the implicit level. Hopefully future research can address some of the methodological limitations of this study to provide a more definitive insight into the kinds of memory representations that underpin crossmodal face recognition.

Chapter 5

General Discussion

The study of face recognition has long been conducted within the realm of vision; however, researchers have recently become interested in our ability to recognise faces in the haptic domain. To date only one study has provided evidence that information about human faces can be shared between the visual and haptic systems (Kilgour and Lederman, 2002). This thesis has sought to explore further the extent to which face information is shared across the visual and haptic modalities. The primary aims of this thesis were to gain a better understanding of the processing similarities that can facilitate crossmodal face recognition, as well as to gain insight into the nature of face information that can be represented in memory across vision and touch.

I will begin with a summary of the main findings from each chapter, followed by a discussion of the cumulative results wherein the data are interpreted in terms of the principal aims of this thesis. Specifically, I will discuss the implications of these results with reference to a number of important issues, such as face specificity, the neural substrates of crossmodal face recognition, and the inclusion of haptic information in extant models of face recognition. Finally, potential avenues for future research into crossmodal face recognition will be considered.

5.1 Summary of findings

In Chapter 2 I investigated the role of familiarity as a factor affecting efficient crossmodal face recognition. The results of Chapter 2 suggest that familiarity per se does not necessarily facilitate efficient crossmodal face recognition. For instance, in Experiment 1, participants were unable to recognise models of their own faces through touch from a line-up of face stimuli without a haptic prime, despite being extremely familiar with their faces through vision. On the other hand, Experiment 2 demonstrated that the crossmodal matching of faces with which participants had been familiarised was significantly better than the matching of unfamiliar faces. In trying to reconcile these disparate results it became apparent that a number of factors may be influential in the sharing of face information across modalities. Haptic familiarity with one's own face may reflect task-dependent encoding and comprise material information absent from the

test stimuli. On the other hand, the majority of visual experience with one's own face stems from mirror reflections. Thus visual (mirror-reversed) and haptic information may be discrepant resulting in a cost in haptic and crossmodal self recognition. These factors were not an issue in Experiment 2. Therefore together these results imply that familiarity alone is not sufficient for recognition, but rather it is necessary that the information specific for recognition is encoded and represented by both the visual and haptic systems.

It appears then that provided the appropriate information is encoded, familiarity promotes better recognition across the visual and haptic modalities. Chapter 3 investigated the representations and processes that mediate successful crossmodal face recognition. Experiment 3 demonstrated that explicit within modal recognition of faces was significantly better than crossmodal face recognition. This finding suggested that face memory is not multisensory in nature. However, crossmodal recognition was better than chance indicating that some face information, presumably information related to facial structure that can be encoded by both modalities, can facilitate recognition of faces across vision and haptics. In Experiment 4, serial encoding of the visual stimuli during learning ruled out encoding differences between vision and haptics as a possible reason for poorer performance in the crossmodal recognition condition. In tandem, these experiments suggested that the relative cost in performance in the crossmodal conditions might be indicative of the loss of modality-specific information that contributes to robust within modal face recognition.

Despite encoding faces in a serial fashion, the haptic system has been shown to process faces configurally (e.g. Kilgour, de Gelder, & Lederman, 2004). With this in mind Experiment 5 was designed to investigate whether configural information could be shared across the visual and haptic modalities, and whether configural information facilitated more efficient matching than facial features. Both configural and featural information were found to facilitate crossmodal matching, however, matching was more efficient when stimuli in both modalities contained configural information. It can be concluded from Chapter 3 that face information is processed in a similar manner across the visual and haptic modalities, but that qualitative differences in the nature of the information encoded facilitates more robust within modal relative to crossmodal face recognition.

The ability to successfully match and recognise faces across modalities (Experiments 2, 3, and 4), and to share information regarding the global face shape between vision and haptics (Experiment 5), suggests indirectly that these two sensory systems might both access shared representations of facial structure in memory. The possibility that a single structural description of a face is shared by vision and haptics was expressly examined in Chapter 4, and Experiment 6 was designed to this end. Although previous behavioural studies have successfully demonstrated crossmodal priming of 3-D objects thus indicating shared representations (Easton, Greene, & Srinivas, 1997; Reales and Ballesteros, 1999), Experiment 6 failed to demonstrate equivalent facilitation for haptically and visually primed faces in a familiarity decision task. However, as there was no evidence of within modal visual priming, it is not possible to draw any definitive conclusions about the existence of shared structural representations that might underpin crossmodal face recognition at the implicit level.

5.2 Interpretation of the cumulative results

There is a long history of face recognition research in the visual domain, and more recently, researchers have looked to the haptic system as an independent channel for face identification. Studies of within modal haptic face recognition have demonstrated a number of similarities to vision in terms of how faces are recognised (Kilgour, 2003; Kilgour, de Gelder, & Lederman, 2004; Kilgour, Kitada, Servos et al., 2005). It appears that haptic face recognition too is primarily configural, is susceptible to the face inversion effect, manifests face selective deficits after brain injury, and is subserved by cortical structures in the ventral temporal cortex. Given these processing similarities, it has been hypothesised that face recognition might be a multisensory phenomenon. Indeed it has been shown that faces known only in vision can be matched with faces encoded haptically and vice versa (Kilgour and Lederman, 2002), lending further credence to this hypothesis.

In this body of work I sought to further extend the findings of these initial studies by examining the extent to which face recognition is multisensory and how face information might be represented in memory so as to facilitate both within modal and crossmodal face recognition. The findings of this thesis demonstrate that face information encoded by one modality can indeed be readily shared with the other for the purposes of face recognition (Experiments 2, 3, and 5), and further, that this sharing is

affected by face familiarity (Experiment 2), and that efficient sharing is mediated by the configural processing of face information (Experiment 5). The experiments also provide behavioural evidence to suggest that crossmodal face recognition is not efficient, as encoded modality-specific information appears to play an important role in within modal face recognition (Experiments 3 and 4). However, the results clearly indicate that sufficient amounts of face information is multisensory in order to allow for successful crossmodal identification. Presumably this information based on facial structure, as 3-dimensional shape can readily be represented by both modalities (e.g. Lederman and Klatzky, 1990). Although direct evidence of shared structural descriptions of faces in memory was not obtained (Experiment 6) the experiments herein do provide some indirect evidence to suggest that this is indeed the case.

5.2.1 Processing similarities facilitating crossmodal face recognition

Shape is a diagnostic property for general object recognition for both the visual and haptic modalities (Lederman and Klatzky, 1990). Likewise, geometrical structure plays an important role in both visual and haptic face recognition; within modal recognition is significantly better than chance in both modalities even when only shape information is available (e.g. Bruce, Healey, Burton et al., 1991; Hill, Schyns, & Akamatsu, 1997; Kilgour and Lederman, 2002). The crossmodal face recognition experiments reported in Chapter 3 demonstrated that face recognition is not wholly multisensory. This may be because some face information, necessary for face recognition, is modality-specific (such as pigmentation in vision and texture in haptics) and cannot be shared across the modalities, resulting in poorer crossmodal relative to within modal recognition performance (Experiments 3 and 4). However, shape information can be represented by both modalities. As crossmodal recognition in these experiments was better than chance it might be assumed that facial structure, at least, is common to both modalities.

The importance of face shape for crossmodal recognition is also evident from the findings of Chapter 2. In Experiment 1 it was shown that participants could not haptically recognise models of their own faces, despite being highly familiar with their faces in vision. However, much of the visual experience of one's own face is gained through viewing the face in the mirror, and research suggests that the resultant memory representation may be different than that for other highly familiar faces (i.e. it is mirror

reversed) (Brady, Campbell, & Flaherty, 2004). This may therefore produce a more modality-specific representation that is not easily shared with haptics, due to a conflict between the veridical and mirror-reversed representations of face shape. Given that vision is a more accurate spatial processor than haptics, the discrepancy may result in the visual capture of face information (e.g. Welch and Warren, 1980). On the other hand, where the encoded structural information are equivalent in both modalities, faces were easily matched across vision and haptics, with better performance for familiar than unfamiliar faces (Experiment 2). Interestingly, haptic self-face recognition accuracy increased significantly when a short haptic prime was permitted before feeling each face model. Given that familiar material face properties were not available in the haptic facemasks, it can be assumed that the improved performance stems from a representation of face shape generated in haptic short-term memory during the priming phase.

On the basis of this behavioural research it can be confidently assumed that facial structure mediates crossmodal face recognition, despite a failure in Experiment 6 to demonstrate a shared structural representation of faces in memory using implicit measures. The question arises as to the nature of such a structural representation: is it primarily a parts-based (featural) or configural representation? In their recent study on reach and grasp movements, Edwards and colleagues found that the estimations of grasp size necessary to grasp an object are significantly smaller for participant's own facial features such as their nose or mouth, compared to other similar sized objects or other body parts (wrist or thumb) (Edwards, Wing, Stevens, & Humphreys, 2004). This finding can be interpreted as indirect evidence that a haptic representation of one's own face is available in memory at the featural level. If a parts-based representation of facial structure is sufficient for haptic face recognition, then it would be expected that performance in the haptic only condition of Experiment 1 should have been as good as the haptic prime condition. On the contrary, haptic face recognition was at chance, suggesting that some additional information is necessary for successful haptic face recognition, presumably configural information.

The nature of the structural representation facilitating crossmodal face recognition was more directly investigated in Experiment 5. In this crossmodal matching task the information content of the visual stimuli was manipulated so that the importance of featural and configural structural information could be assessed

independently (cf. Collishaw and Hole, 2000; Schwaninger, Lobmaier, & Collishaw, 2002). Congruent with previous visual studies (e.g. Collishaw and Hole, 2000), configural and featural information were both found to be important in terms of recognition accuracy. This finding implies that haptic representations are not solely part-based as suggested by the constraints in haptic encoding: matching to blurred images in which only configural information was available was as good as matching to intact or scrambled images. In fact, an advantage was found for configural information relative to featural information in relation to matching efficiency. Therefore it appears that configural information is not only important for within modal face recognition, but also mediates efficient crossmodal face recognition.

In summary, this body of work extends the preliminary investigations of Kilgour and Lederman (2002) by demonstrating that a number of processing similarities facilitate the sharing of face information between vision and haptics. Two commonalities in particular appear to mediate successful crossmodal face recognition. Firstly, both modalities encode 3-dimensional structural information, and this information is diagnostic for face recognition. This thesis has provided behavioural evidence indicating that structural information may be shared by both the visual and haptic modalities. Furthermore, both modalities appear to preferentially process configural face information.

5.2.2 Crossmodal face representations

Thus far the discussion of the experimental findings described in Chapters 2 and 3 has emphasised the importance of face shape for crossmodal face recognition. In addition, Experiments 3 and 4 clearly suggest that modality-specific information may be important for within modal relative to crossmodal face recognition. Collectively, these findings allow me to speculate on how face information is represented in memory to facilitate both within modal and crossmodal face recognition.

Recall that in Chapter 1 I outlined a number of candidate models describing the representation of multisensory information in memory. The two more simplistic models respectively stated that different sensory information is either a) all represented independently, or b) all multisensory. The experimental findings of this thesis support the rejection of both of these models. It is clear that face information can be recognised and matched across modalities (Experiments 2, 3, 4, and 5), but that certain face

information may be modality-specific thus promoting better within modal relative to crossmodal face recognition (Experiments 3 and 4).

A third model proposes that haptic information is mediated by the visual system, that is, haptic recognition is a product of visual imagery. The role of visual imagery was not directly studied in this collection of experiments as previous researchers have provided evidence, using a variety of methods, that visual imagery is not an influential component of crossmodal object and face recognition. For instance, brain imaging studies have found that the lateral occipital oomplex (LOC), located in the occipito-temporal pathway, responds preferentially to visual object shape compared to scrambled images or textures (Malach, Reppas, Benson, et al., 1995) and subsequently it has been discovered that this area also shows activation for haptically recognised objects (e.g. Amedi, Jacobson, Hendler, et al., 2002; Amedi, Malach, Hendler, et al., 2001; James, Humphrey, Gati, et al., 2002). It has been proposed that haptic activation of this visual cortex is representative of visual imagery and is necessary for haptic object recognition (e.g. Zhang, Weisser, Stilla, et al., 2004). This visual imagery explanation has been countered by evidence showing that visually imagining objects produces much less activation in LOC than does haptic exploration of the same objects (Amedi et al., 2001). Moreover, characteristic sounds expected to elicit visual imagery of various objects (e.g. manmade tools and devices) did not activate LOC, indicating that visual imagery is not a major factor in LOC activation, rather activation is due to either the visual or haptic encoding or representation of object shape (Amedi et al., 2002). These authors argued instead that visual and haptic activation of LOC implies a common haptic and visual representation of object shape.

In terms of haptic face recognition, Kilgour and Lederman (2002) hypothesised that if haptic face recognition was mediated by the visual system, then scores on the Vividness of Visual Imagery Questionnaire (VVIQ; a low score indicates a high ability to use mental imagery) would be negatively correlated with recognition speed and accuracy in a recognition task. However this correlation was not obtained, suggesting that a high ability to use mental imagery was not necessary for haptic and crossmodal face recognition. Further to this, Pietrini, Furey, Riccardi and colleagues (2004) found that haptic exploration of facemasks activated the same areas of ventral temporal cortex in both sighted and blind participants, indicating that visual experience is not necessary to elicit haptic face-related activations in ventral extra-striate cortex. In this thesis,

participants frequently reported that it was actually very difficult to visualise the haptic face stimuli; a similar informal finding was reported by Pietrini et al.

Although it is still possible that visual imagery may have a small contributory role, the results of the above studies suggest that it can be largely ruled out as a necessary mediator of haptic and crossmodal face recognition. Therefore one model remains as the best candidate for crossmodal face recognition; namely Millar's (2000) model of sensory convergence. According to this model, information from different sensory sources converges to share information that is represented in the same manner in both modalities, whilst modality-specific information is represented but clearly not shared across modalities. It is clear from the results presented in this thesis that crossmodal face recognition fulfils the criteria of this model almost perfectly. Although the crossmodal priming experiment failed to provide direct evidence of a shared structural representation common to both vision and haptics, the strong evidence discussed above suggests that information about face shape is encoded by both modalities. It is also apparent that modality-specific information is important for optimal within modal face recognition, since recognition performance has been found to decline in within modal haptic and visual tasks where recognition is based on solely on geometrical information (e.g. Bruce et al., 1991; Hill et al, 1997; Kilgour and Lederman, 2002). It is possible that the decline in performance for crossmodal relative to within modal face recognition arises due to the loss of modality-specific face information. Therefore, face information may be represented in memory according to the model illustrated in Figure 5.1.

This adaptation of Millar's model into one of crossmodal face representation has a number of advantages. Firstly, it can easily account for the behavioural data documented in this thesis as well as the findings of Kilgour and Lederman (2002). The ability to match and recognise faces across the visual and haptic modalities may be mediated by structural face representations common to both modalities. Secondly, it can account for the fMRI data of Pietrini et al. (2004). For example, although early processing of all stimuli activated a region in early visual cortex, later patterns of response activations for visually and haptically explored faces were located in adjoining regions of ventral extra-striate cortex, but were not found to overlap. It is possible that earlier activations represented the processing of shape information whilst later

activations reflected the higher-order processing of modality-specific properties of the face stimuli.

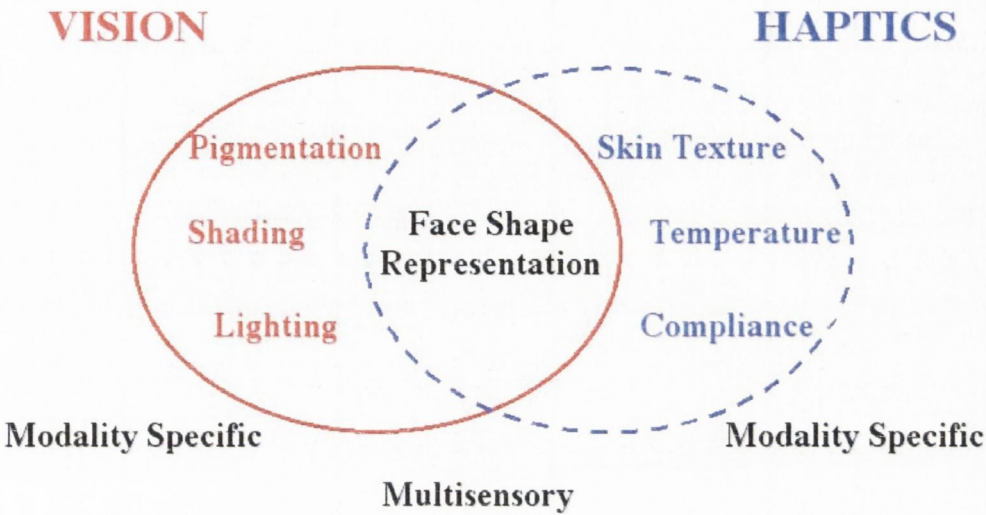


Figure 5.1. Millar's (2000) model of convergence as a model of crossmodal face representation. In the centre of the diagram redundant information about face shape is encoded by both modalities. Modality-specific information is represented on either side of the commonly represented information. When a face is seen, information within the solid red ellipse is retrieved. Alternatively if a face is explored using active touch, then face information in the dashed blue ellipse is retrieved. Crossmodal recognition is mediated by the shared structural information.

Kilgour, Kitada, Servos et al. (2005) focused on loci of response activations during haptic face identification. As they did not report areas of cortex that may have been active during both object and face identification, it is unknown whether they observed any brain activity in early visual cortex that might represent the neural analogue of shape processing. Patterns of activation in later stages of processing that differed from those typically found in visual studies of face identification (i.e. left fusiform activation rather than activation of the more typical visual right 'fusiform face area') may again reflect modality-specific aspects of face processing. I will return again to this point later in the discussion. Finally, this model of face representation can also provide an explanation of the findings with prosopagnosic LH in that damage to the shared structural representation of face shape would preclude face recognition in either modality.

5.3 Implications of results

Interpreted in unison, the experimental investigations of this thesis provide behavioural evidence to suggest that face recognition is not completely multisensory. I suggest that some information, perhaps more modality-specific information which is important for optimal within modal recognition, is not easily shared across modalities. Rather, representations of facial structure are most likely to be central to successful crossmodal face recognition. Moreover, face recognition across the modalities is affected by face familiarity and efficient recognition is mediated by configural face information. In this section I have identified 3 main areas for which these findings have important implications. Firstly, these results can contribute in a small way to the current debate of whether face recognition is special. This research also has implications for studies of brain function, in the context of both the face processing and multisensory cognition literatures. Finally, these studies can inform current models of face recognition, and suggestions can be made as to how haptic face information can be incorporated into extant models.

5.3.1 *Domain-specificity versus expertise*

A hotly debated issue in the visual face recognition literature concerns whether faces constitute a special class of objects such that they are processed by mechanisms different from those that underpin the processing of all other non-face objects. On the one hand, a number of studies have provided evidence that face recognition is a unique process (see Kanwisher, 2000 for a review). On the other hand, data also exist to demonstrate that face processing is a function of expertise, and that the same behavioural effects (e.g. configural processing, or the inversion effect) can be observed for any object class with which an individual is expert (e.g. Gauthier, Tarr, Moylan, et al., 2000; Tarr and Gauthier, 2000). This debate was not the focus of the current body of research, and consequently suitable non-face object controls were not employed. Nevertheless, the results do appear to favour the argument for domain-specificity.

The haptic modality is non-expert in its capacity for processing faces, yet even unfamiliar faces can be matched across modalities (Experiment 2 here; Kilgour and Lederman, 2002), demonstrating that face recognition can be performed by a non-expert sensory system. More particularly, processing strategies, such as configural processing, which have been argued to be a function of expertise have also been shown to extend

into the non-expert haptic modality. Configural processing, discerned largely using the inversion paradigm, has long been posited as a marker of face specificity (e.g. Bartlett and Searcy, 1993; Moscovitch et al., 1997; Yin, 1970). The inversion paradigm has recently been used to demonstrate configural processing in haptic face recognition tasks but not for haptic object recognition, suggesting face-specificity in the haptic modality also (Kilgour, de Gelder, & Lederman, 2004). However, as a visual inversion effect has been found for some mono-oriented non-face objects, even in non-experts (e.g. Gauthier et al., 2000b), the inversion effect by itself cannot be taken as confirmation of the uniqueness of face processing in either modality (cf. Maurer et al., 2002). This thesis has extended the finding of the haptic face inversion effect by using an alternative behavioural paradigm (crossmodal face matching using intact, blurred, and scrambled images) to show that efficient crossmodal face recognition appears to be mediated by configural face information (Experiment 5) adding credence to the notion that configural processing is symptomatic of a domain-specific face processing mechanism. Of course, further research is required to establish the extent to which crossmodal face processing is domain-specific.

5.3.2 Neural substrates for crossmodal face recognition

Investigations of haptic face recognition have found a number of behavioural similarities with visual face processing and that face information can be successfully shared between the two modalities (Kilgour and Lederman, 2002; Kilgour, de Gelder, & Lederman, 2004), suggesting that face recognition may be a multisensory phenomenon. In contrast to this expectation, the results of Experiments 3 and 4 in this thesis suggest that face recognition is not wholly multisensory. This finding has a number of implications in terms of identifying the neural substrate underpinning crossmodal face recognition.

Initial investigations of response activations to haptically explored face stimuli questioned whether the previously identified fusiform face area (FFA) in the right fusiform gyrus (e.g. Haxby et al., 1994; Kanwisher et al., 1997; Sergent et al., 1992) would similarly respond to haptic face stimuli, indicating a possible locus of multisensory face processing in the brain. To the contrary, two fMRI studies addressing haptic categorisation, and haptic identification of face stimuli found that haptically elicited activation did not overlap with the previously defined FFA (Kilgour, Kitada,

Servos, et al., 2005; Pietrini, Furey, Riccardi, et al., 2004). In the study conducted by Kilgour et al. haptic face identification activated the left rather than the right fusiform gyrus, as well as a number of other cortical areas previously implicated in visual face memory. Pietrini and colleagues examined category-related responses for visual and haptic face and non-face stimuli. Although they did not find a common area of activation evoked by haptic and visual face stimuli in ventral temporal cortex, haptic response activations were located close to the visual sites. However, earlier visual areas were activated during haptic recognition by all stimuli (facemasks and non-face objects alike). Perhaps therefore, early processing of faces and objects might be multisensory, with only higher-order face processing occurring in different ventral temporal areas^{5.1}.

The behavioural data reported in this thesis may help to shed some light on the disparate findings of visual and haptic face neuroimaging studies. Although direct evidence of common shape representations was not obtained, the results clearly show that encoded face shape information mediates successful crossmodal face recognition. Moreover, some face information may be modality-specific and necessary for optimal within modal face recognition. These findings fit nicely with the imaging data. Differential activation in ventral temporal cortex may well reflect the different computational requirements involved in processing face information specific to the visual and haptic modalities. However, face shape can readily be 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.

In terms of the distributed neural system of face perception outlined by Haxby et al. (2000), representations of facial identity appear to depend more on activity in the inferior occipital gyri (IOG) and the lateral fusiform gyri (LFG), the IOG being involved in the early visual processing of face information (Hoffman and Haxby, 2000). In their analysis of neural responses in visual extra-striate cortex that are activated during tactile recognition, Pietrini et al. (2004) found that both face and non-face stimulus categories evoked a response in ventral occipital cortex during haptic recognition. Interestingly, the IOG are among the brain regions included in the ventral occipital area, suggesting

^{5.1} As Kilgour et al. focused on haptic face identification specifically, only activations for haptic face stimuli that were significantly greater than haptic control objects were reported (the objects > facemask condition yielded no significant results). Therefore it is unknown whether haptic identification of facemasks and objects might have activated earlier visual areas.

that these regions, involved in early visual face processing, may also mediate early haptic face processing. Furthermore, the IOG, which have been found to respond preferentially to intact versus scrambled face images (Hoffman and Haxby, 2000), also correspond remarkably with the lateral occipital complex (LOC) which has been shown to respond preferentially to visual object shape relative to scrambled images or textures (Malach et al., 1995). Previous studies of crossmodal object recognition have found that a region in LOC represents both visual and haptic object shape information (Amedi et al., 2001; 2002). This somatosensory object-related lateral occipital area (SOLO or LOtv) appears to overlap the area of inferior temporal cortex found to respond maximally to the haptic recognition of objects (Pietrini et al., 2004), confirming the status of this area as a mediator of visual and haptic object recognition. Given that brain regions activated during early visual and haptic face processing appear to overlap (see Table 5.1 for the coordinates of the brain areas described), it seems plausible that another region of the LOC may serve as the locus of multisensory representations of face shape. For instance, it has been suggested that the posterior face regions that fall within the region of the LOC may serve to encode facial structure, whereas more anterior parts of the FFA subserve face recognition (e.g. Courtney, Ungerleider, Keil, & Haxby, 1997; Sergent et al., 1992). These posterior face regions may well mediate multisensory representations of face shape.

Naturally, confirmation of this hypothesis requires further investigation. Functional neuroimaging of crossmodal face recognition and face matching tasks may help to elucidate the role of LOC in visual and haptic face processing. Putative common representations of face shape in LOC could be assessed with a crossmodal face priming study. However, the results of Chapter 4 indicate that the crossmodal priming of faces may not be as straightforward as that of objects therefore refinements of the current methodology are required.

Table 5.1. Talairach coordinates (Talarach and Tournoux, 1988) as reported in each study described in the text (mean $\pm$ standard deviation). As these brain regions are proximal and overlapping (e.g. LOC, IOG, VO) it is possible that these areas may contain a region of cortex that mediates both the visual and haptic processing of face shape.

Study	Region	n	Talairach Coordinates (mm)			Laterality
			X	Y	Z	
Malach et al. 1995	LOC	6	42.8 (± 2.7)	-72.7 (± 8)	-18.2 (± 9.8)	right*
Hoffman & Haxby 2000	IOG	7	-31 (± 1)	-82 (± 1)	-15 (± 4)	left
		7	41 (± 2)	-79 (± 3)	-14 (± 5)	right
Amedi et al. 2001	LO	9	-45 (± 5)	-62 (± 6)	-9 (± 3)	left
Amedi et al. 2002	SOLO	8	-47	-62	-10	left
		8	45	-54	-14	right
§ Amedi et al. 2001	SOLO	9	-45	-59	-11	left
		9	47	-55	-9	right
Pietini et al. 2004	VO	5	-31	-80	-12	left
		5	26	-77	-12	right
	IT	5	-46	-58	-11	left
		5	48	-55	-10	right

LOC = lateral occipital complex; IOG = inferior occipital gyri; LO = lateral occipital area; SOLO = somatosensory object-related lateral occipital area; VO = ventrolateral occipital cortex; IT = inferior temporal cortex. * Small bilateral activation was found in 5/6 participants. § Data from Amedi et al. (2001) were re-analysed using a General Linear Model (Friston, Holmes, Worsley, et al., 1994) analysis to confirm the location of the area SOLO (Amedi et al., 2002).

In strict terms, a multisensory face representation would be coded by neurons that respond to both visual and haptic input allowing the integration of simultaneous input across the modalities (cf. Meredith, 2002). To my knowledge, no studies have provided any direct evidence of multisensory neurons in human LOC as of yet. However, a rough estimate of the sensory response properties of neurons in LO have been obtained with an fMRI design allowing the study of crossmodal repetition-related adaptation (Amedi et al., 2001). In this study, the signal produced for objects that were haptically explored after being seen was reduced compared to that for haptically palpated objects that were unseen. This has been interpreted as evidence of crossmodal adaptation, which in turn implies that neurons in at least this region of the LOC are multisensory. Of course, further research is necessary to elucidate the sensory properties of these neurons more precisely, especially with respect to face stimuli. However for the

moment, this research suggests that the processing of face shape may be underpinned by multisensory substrates.

5.3.3 Models of face recognition

In Chapter 1 I speculated about how haptic face information might be incorporated into extant models of face recognition. In terms of the cognitive models of face recognition (e.g. Bruce and Young, 1986; Burton, Bruce, & Hancock, 1999), I proposed that the face recognition units (FRUs) may comprise overlapping representations of visual and haptic face information, akin to the model of sensory convergence proposed by Millar (2000). Information pertaining to face shape encoded by both modalities may be common to the two, whilst information promoting within modal face recognition that is not easily shared across modalities is also represented. The results of this thesis do suggest, albeit indirectly, that Millar's model of convergence provides the best account of the mechanisms underpinning crossmodal face recognition. In turn, this supports the proposal that haptic face information can be incorporated into cognitive models of face recognition at the level of the FRUs. Unfortunately, Experiment 6 failed to demonstrate either crossmodal or within modal priming. Future research might examine methodological issues that may have limited this experiment (a number of suggestions were made in section 4.3). A demonstration of crossmodal priming would serve to confirm the hypothesis that FRUs comprise multisensory representations of face shape.

In section 5.3.2 I have discussed how multisensory representations of face shape mediating crossmodal face recognition might be represented in Haxby et al.'s (2000) distributed neural model of face recognition. A multisensory region in the inferior occipital gyri may mediate both the crossmodal and within modal processing of face information. In Chapter 1 I also proposed that a separate higher-order area of cortex mediating haptic face recognition is necessary to account for different regions of activation in fMRI studies for within modal haptic and visual face recognition. Modality-specific information may be relayed directly to the haptic and visual face processing areas from primary sensory cortices. This information is then bound with shape information processed in IOG, promoting optimal within modal face recognition. Again the behavioural findings in this body of work support this proposition. Better within modal relative to crossmodal face recognition may be due to the addition of

modality-specific face information, whilst information about face shape facilitates crossmodal matching and recognition. Functional neuroimaging has thus far proven a useful tool in examining the cortical areas involved in visual and haptic face recognition, and has provided a number of clues as to the locus of putative shared representations of face shape. One interesting future investigation might concern the cortical pathways linking the patterns of activation incorporated into Haxby et al.'s cortical model of face recognition. For instance, techniques such as diffusion tensor MRI (e.g. Conturo, Lori, Cull, et al., 1999) might provide insight as to how these different brain areas are connected, and whether modality-specific information is relayed directly from primary sensory cortices to separate haptic and visual face processing areas.

5.4 Limitations and directions for future research

Haptic and crossmodal face recognition is subject to a number of methodological challenges primarily stemming from the difficulty in generating a sufficiently large, and perceptually realistic, set of haptic stimuli for experimental purposes. Firstly, whilst the casting process itself was reasonably straightforward, it was a time consuming process. Thus the stimulus set was quite small (32 models in total). This had implications for data collection; the examination of multi-level factors meant that in a number of experiments (Experiments 2 and 6) performance was measured over relatively few trials (6 and 4 trials per condition, respectively). This may be an important factor when considering the failure to demonstrate within and crossmodal priming in Experiment 6; with only 4 primed and 4 unprimed stimuli in each condition the analysis may not have been powerful enough for significant facilitation effects to emerge. On a related point, the very nature of the research question in Experiment 1, concerned with the haptic recognition of one's own face, also restricted the number of trials on which performance was measured in each condition (i.e. 1 trial per condition).

Secondly, the degree of perceptual realism of both the visual and haptic stimulus sets holds a number of implications in terms of assessing crossmodal performance in all of the experiments. An unavoidable limitation experienced in generating the haptic stimulus set was the inability to simulate the material properties, specifically the texture and compliance, of each individual's face, resulting in a set of rigid, uniformly textured, haptic face stimuli. As the processing of material properties has been shown to be important for optimal haptic object and face recognition (e.g. Kilgour and Lederman,

2002; Klatzky et al., 1985; 1993; Klatzky, Lederman, & Reed, 1987), it could be argued that haptic face processing in this body of work was disadvantaged relative to vision (at least for Experiments 2-6), as not all of the information requisite for optimal recognition performance was available. This point has particular relevance for Experiment 1 wherein participants were required to recognise their own facemasks using touch. If the haptic representation of one's own face in memory comprises both material and structural information, then the absence of material information in the haptic stimuli could help to explain participants' poor performance in this task.

In terms of visual recognition, the highly precise colour images used in Experiments 2-6 provide all of the visual cues encoded from live faces, such as structure, pigmentation, 3-dimensional shape-from-shading, and texture information, facilitating optimal visual face recognition. Indeed, loss of modality appropriate information may also explain poorer than expected visual performance in the visual condition of the Experiment 1. Unlike in subsequent experiments where visual encoding and recognition was based on high-quality digital colour images of the live faces of the face models, participants had to visually recognise their own facemasks which lacked appropriate colour information, resulting in sub-optimal visual recognition performance (cf. Yip and Sinha, 2002).

Differences in the potential information available to participants from visual and haptic stimuli may also hold interesting implications for crossmodal matching performance obtained in Experiments 2 and 5. A visual encoding advantage has been observed in crossmodal matching tasks (Easton et al., 1997; Jones, 1981), and this advantage is thought to be a function of the relative efficiency by which vision, in comparison to haptics, can extract information from the standard object, which then guides exploration of the comparison object (Jones, 1981). This visual encoding advantage may explain the results of Experiment 2, where significantly better vision-to-haptics than haptics-to-vision crossmodal face matching was found. In Experiment 2, the standard stimuli in the visual learning condition contained rich, modality appropriate information. However, the haptic standards, providing only structural information, did not provide information necessary for generating optimal haptic face representations. Moreover, vision is capable of encoding the global structure of the face, quickly and easily, whereas haptic face processing requires the serial exploration of the face stimuli and the integration of this information into a coherent representation over time, which

can be relatively error prone (Loomis and Lederman, 1986). This may go some way to explain why vision-to-haptics matching performance is significantly better in Experiment 2. This same argument may explain the overall poor crossmodal matching performance observed in Experiment 5 where crossmodal matching was conducted in haptics-to-vision only. Already at a loss as regards material information, if participants failed to encode all of the structural information available from the haptic standards, then that information was non-diagnostic, and the likelihood of an accurate match was reduced further.

A final point to note in relation to the visual and haptic stimulus sets, and one that can easily be explored in future research, concerns differences in the information about the eye area that was available from the stimuli in each modality. The visual stimulus set comprised images of live faces in which the eyes were open, whereas the eyes were closed as a consequence of the casting processes for haptic stimuli. This difference has implications in terms of processing shape information around the eye area. The surface of the open eye is moist, which may result in a distorted reflection of light from the surface, influencing its perceived shape. Furthermore, the difference in luminosity between the open eye surface and surrounding eye region may also serve to give an altered perception of eye shape that may not be perceived when the eye is closed. This may lead to perceived differences in shape between vision and haptics thus affecting crossmodal recognition performance. The face is an important social stimulus, and in terms of social interaction, eye gaze and contact serve a number of functions, such as indicating social attention and expressing intimacy (Kleinke, 1986). Researchers have recently found that direct eye gaze confers processing advantages during face encoding and increases the accuracy of subsequent face recognition (e.g. Hood, Macrae, Cole-Davies, & Dias, 2003; Vuilleumier, George, Lister, & Armony, 2005), whilst diverted gaze confers greater advantages relative to eyes closed at encoding (Hood et al., 2003). Therefore in this body of work, particularly in Experiments 4, and 6 where visual stimuli comprised images of faces looking directly at the camera, vision may have been at an advantage during encoding and retrieval. Having their eyes closed in the visual stimuli may serve to elucidate the importance of eye information in vision and its influence on crossmodal face recognition performance.

Undoubtedly, a limitation of this research was the failure of Experiment 6 to provide evidence of crossmodal face priming. However, as the experiment also failed to

yield evidence of within modal priming, the hypothesis that a common representation of face shape underlies crossmodal face recognition cannot be rejected. Indeed, the data from Chapters 2 and 3 provide indirect evidence in support of this hypothesis, therefore the implications of the cumulative results still have merit, and open a number of potential avenues for future research. The first step may be to obtain concrete behavioural evidence of crossmodal face priming. However, beyond this, there are a number of exciting challenges for researchers of face recognition and multisensory perception alike. Crossmodal face recognition is an emerging area of interest with huge potential and scope for future work. I have identified a number of possible studies of interest in the discussion of the implications of my research findings, for instance, the identification of crossmodal face activation in LOC.

The mechanism underpinning the integration of visual and haptic face information is also of interest, especially in terms of facial identity. Previous research investigating the integration of visual and haptic object information suggests that sensory integration is information driven, such that integration occurs in a statistically optimal fashion, with greater weight attributed to the sensory information that is most reliable in terms of making a perceptual judgement about an object (Ernst and Banks, 2002). The visual perception of facial identity has been shown to be categorical (e.g. Rotshtein, Henson, Treves, et al., 2005); therefore, if visual identity information captures haptic information, then in a bimodal task with differing sensory information, haptic recognition performance would approximate results obtained in a visual categorisation task. However, this remains to be seen.

An interesting application of multisensory integration research might include addressing the processing deficits associated with prosopagnosia. The integration of sensory information is subject to a number of guiding principles. For example, as outlined in the introduction, a coherent multisensory percept is characterised by the close temporal proximity and spatial co-location of sensory stimuli (illusory percepts such as the ventriloquism effect notwithstanding). Another principle that typically governs multisensory integration is the 'inverse effectiveness principle' (Wallace and Stein, 1994). This principle stipulates that when sensory stimuli are integrated the facilitation of the neuronal response is maximal when the responses to the individual inputs is weakest. This principle may be applied to face recognition. For instance, prosopagnosic LH shows face recognition deficits in both vision and haptics respectively. In other

words, the sensory input in the respective modalities is very weak, resulting in significantly poorer face recognition performance relative to control participants in within modal tasks. However, the inverse effectiveness principle suggests that the integration of two sources of impoverished face information might facilitate the generation of more robust face representations. Therefore it might be expected that LH's recognition performance would improve slightly or even significantly in a multisensory face recognition task. A significant improvement in recognition performance could signal a potential avenue of exploration in terms of rehabilitation, not only for prosopagnosia, but for general object agnosia as well.

5.5 Concluding remarks

In this thesis I have used a number of simple variations of the basic crossmodal matching and recognition memory paradigms to investigate crossmodal visual and haptic face recognition. I have confirmed the findings of previous research in demonstrating that the haptic system is indeed capable of representing and recognising faces, and that face information can be shared across the visual and haptic modalities. Moreover, this body of work extends these findings by showing that crossmodal face recognition can be facilitated by familiarity, and that efficient crossmodal recognition appears to be mediated by configural face information. These experiments also provided behavioural evidence to suggest that face recognition is not multisensory per se. It appears that modality-specific information may promote optimal within modal face recognition performance. However, the results clearly indicate that enough information is multisensory to allow successful crossmodal recognition, and furthermore, imply that the visual and haptic crossmodal face recognition may be underpinned by common representations of face shape.

References

A

- Amedi, A., Jacobson, G., Hendler, T., Malach, R., Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital cortex. *Cerebral Cortex*, 12 (11), 1202-12.
- Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, Z. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience*, 4, 324-330.
- Amedi, A., von Kriegstein, K., van Atteveldt, N.M. Beauchamp, M.S., & Naumer, M.J. (2005). Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research*, 116(3-4), 559-571.

B

- Baddeley, A. & Woodhead, M. (1983). Improving face recognition ability. In S. Lloyd-Bostock & B. Clifford (Ed.s) *Evaluating Witness Evidence*, (pp 125-136). Chichester: Wiley.
- Bartlett, J.C. & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology*, 25, 281-316.
- Barton, J.J., Press, D.Z., Keenan, J.P., O'Connor, M. (2002). Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology*, 58(1), 71-78.

- Bauer, R.M. (1984). Autonomic recognition of names and faces in prosopagnosia: A neuropsychological application of the guilty knowledge test. *Neuropsychologia*, 22, 457-469.
- Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403, 309-312.
- Bentin, S., & Deouell, L.Y. (2000). Structural encoding and identification in face processing: ERP evidence for separate mechanisms. In N. Kanwisher & M. Moscovitch (Ed.s) *The Cognitive Neuroscience of Face Processing*. UK: Psychology Press Ltd.
- Bentin, S., & Moscovitch, M. (1988). The time course of repetition effects for words and unfamiliar faces. *Journal of Experimental Psychology: General*, 117(2), 148-160.
- Bentin, S., Allison, T., Perez, E., Puce, A., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, 8, 551-565.
- Bertelson, P (1999). Ventriloquism: a case of crossmodal perceptual grouping. In G. Ascherlseven, T. Bachmann, J. Musseler (Eds.), *Cognitive Contributions to the Perception of Spatial and Temporal Events*. (pp. 347-362). Amsterdam: Elsevier Press.
- Bertelson, P. & Radeau, M.(1981). Crossmodal bias and pereptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, 29, 578-574.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115-147.

- Biederman, I., & Gerhardstein, P.C. (1993). Recognising Depth-rotated objects: Evidence for 3D viewpoints invariance. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 1162-1182.
- Blanz, V., Tarr, M.J., Bülthoff, H.H. (1999). What object attributes determine canonical views? *Perception*, 28, 575-599.
- Bodamer, J. (1947). Die Prosop-Agnosie. *Archiv für Psychiatrie und Nervenkrankheiten*, 179, 6-53.
- Brady, N., Campbell, M., & Flaherty, M. (2004). My left brain and me: a dissociation in the perception of the self and others. *Neuropsychologia*, 42(9), 1156-1161.
- Bresciani, J.P., Ernst, M.O., Drewing, K., Bouyer, G., Maury, V. & Kheddar, A. (2005). Feeling what you hear: auditory signals can modulate tactile tap perception. *Experimental Brain Research*, 162(2), 172-180.
- Bruce, V. (1983). Recognising faces. *Philosophical Transactions of the Royal Society London, B*, 302, 423-436.
- Bruce, V. (1986). Influences of familiarity on the processing of faces. *Perception*, 15, 387-397.
- Bruce, V, Burton, A.M., Carson, D., Hanna, E., & Mason, O. (1994). In C. Umiltà & M. Moscovitch (Eds.), *Attention & Performance XV* (p.p 179-201). Cambridge, MA:MIT Press.
- Bruce, V., Carson, D., Burton, A.M., & Kelly, S. (1998). Prime time advertisements: Repetition priming from faces seen on subject recruitment posters. *Memory & Cognition*, 26(3), 502-515.
- Bruce, V, Healey, P., Burton, A.M., Doyle, T. (1991). Recognising facial surfaces. *Perception*, 20, 755-769.

- Bruce, V., Henderson, Z., Hancock, P., Burton, A.M., & Miller, P. (1999). Verification of face identities from face images captured on video. *Journal of Experimental Psychology: Applied*, 5(4), 339-360.
- Bruce, V., Henderson, Z., Newman, C., & Burton, A.M. (2001). Verification of face identities from faces captured on video. *Journal of Experimental Psychology: Applied* 7(3), 207-218.
- 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., & Valentine, T. (1985). Identity priming in the recognition of familiar faces. *British Journal of Psychology*, 76, 373-383.
- Bruce, V. & Valentine, T. (1986). Semantic priming of familiar faces. *Quarterly Journal of Experimental Psychology*, 38A, 125-150.
- Bruce, V., Valentine, T., & Baddeley, A. (1987). The basis of the 3/4 view advantage in face recognition. *Applied Cognitive Psychology*, 1, 109-120.
- Bruce, V. & Young, A.W. (1986). Understanding Face Recognition. *British Journal of Psychology*, 77, 305-327.
- Bruck, M., Cavanagh, P., & Ceci, S.J. (1991). Fortysomething: recognising faces at one's 25th reunion. *Memory & Cognition*, 19, 221-228.
- Bruyer, R., Laterre, C., Seron, X., Feyereisen, P., Strypstein, E., Peirard, E., & Rectem, D. (1983). A case of prosopagnosia with some observed covert remembrance of familiar faces. *Brain and Cognition*, 2, 257-284.
- Bülthoff I, Newell FN (2004) Categorical perception of sex occurs in familiar but not unfamiliar faces. *Visual Cognition*, in press.
- Burton, A.M. & Bruce, V. (1993). Naming faces and naming names: Exploring an interactive-activation model of person recognition. *Memory*, 1, 457-480.

- Burton, A.M., Bruce, V., and Hancock, J.B. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science*, 23(1), 1-31.
- Burton, A.M., Bruce, V., & Johnston, R.A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology*, 81, 361-380.
- Burton AM, Wilson S, Cowan M, Bruce V (1999) Face recognition in poor quality video: evidence from security surveillance. *Psychological Science*, 10, 243-248.

C

- Calder, A.J., & Young, A.W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, 6, 641-651.
- Calvert, G.A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, 11, 1110-1123.
- Calvert, G.A., Bramner, M.J., & Iversen, S. (1998). Crossmodal identification. *Trends in Cognitive Sciences*, 2, 247-253.
- Calvert, G.A., & Campbell, R. (2003). Reading speech from still and moving faces: the neural substrates of visible speech. *Journal of Cognitive Neuroscience*, 15(1), 57-70.
- Calvert, G.A., Campbell, R., & Bramner, M.J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in human heteromodal cortex. *Current Biology*, 10, 649-657.
- Campbell, R., Brooks, B., de Haan, E., & Roberts, T. (1996). Dissociating face processing skills: decisions about lip-read speech, expression, and identity. *Quarterly Journal of Experimental Psychology: A*, 49(2), 295-314.

- 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.
- Cave, C.B., Bost, P.R., & Cobb, R.E. (1996). Effects of colour and pattern on implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22(3), 639-653.
- Clavagnier, S., Falchier, A., & Kennedy, H. (2004). Long-distance feedback projections to area V1: implications for multisensory integration, spatial awareness, and visual consciousness. *Cognitive, Affective, and Behavioural Neuroscience*, 4(2), 117-126.
- Clutterbuck, R., & Johnston, R.A. (2002). Exploring levels of face familiarity by using an indirect face matching task. *Perception*, 31(8), 985-944.
- Collishaw, S.M., & Hole, G.J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception*, 29, 893-909.
- Conturo, T.E., Lori, N.F., Cull, T.S., Akbudak, E., Snyder, A.Z., Shimony, J.S., McKinstry, R.C., Burton, H., & Raichle, M.E. (1999). Tracking neuronal fibre pathways in the living human brain. *Proceedings of the National Academy of Sciences, of the U.S.A.*, 96, 10422-10427.
- Cooper, L.A., Schacter, D.L., Ballesteros, S., & Moore, C. (1992). Priming and recognition of transformed three-dimensional objects: Effects of size and reflection. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18, 43-57.
- Courtney, S.M., Ungerleider, L.G., Keil, K., & Haxby, J.V. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature*, 386, 608-611.

D

- de Haan, E.H.F., Young, A.W., & Newcombe, F. (1987). Face recognition without awareness. *Cognitive Neuropsychology*, 4, 385-415.
- de Gelder, B. & Rouw, R. (2000). Structural encoding precludes recognition of face parts in prosopagnosia. In N. Kanwisher & M. Moscovich (Ed.s) *The Cognitive Neuroscience of Face Processing* (pp. 89-102). UK: Psychology Press Ltd.
- de Gelder, B., Bachoud-Levi, A.C., & Degos, J.D. (1998). Inversion superiority in visual agnosia may be common to a variety of orientation polarised objects besides faces. *Vision Research*, 38, 2855-2861.
- Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, 115, 107-117.
- Driver, J. & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology*, 10, R731-R735.

E

- Easton, R.D., Greene, A.J., & Srinivas, K. (1997). Transfer between vision and haptics: Memory for 2-D patterns and 3-D objects. *Psychonomic Bulletin & Review*, 4, 403-410.
- Edelman, S. & Bulthoff, H.H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, 32, 2385-2400.
- Edwards, M.G., Wing, A.M., Stevens, J., & Humphreys, G.W. (2005). Knowing your nose better than your thumb: measures of over-grasp reveal that face-parts are special for grasping. *Experimental Brain Research*, 161(1), 72-80.

- Ellis, A.W., Flude, B.M., Young, A.W., & Burton, A.M. (1996). Two loci of repetition priming in the recognition of familiar faces. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 22, 295-308.
- Ellis, A.W., Young, A.W., & Flude, B.M. (1990). Repetition priming and face processing: priming occurs within the system that responds to the identity of a face. *Quarterly Journal of Experimental Psychology: A*, 42(3), 495-512..
- Ellis, A.W., Young, A.W., Flude, B.M., & Hay, D.C. (1987). Repetition priming of face recognition. *Quarterly Journal of Experimental Psychology*, 39A, 193-210.
- Ellis, H.D. (1987). Processes underlying face recognition. In R. Bruyer (Ed.), *The Neuropsychology of Face Perception and Facial Expression*. Hillsdale, NJ: Lawrence Erlbaum.
- Ellis, H.D., Ellis, D.M., & Hoise, J.A. (1993). Priming effects in children's face recognition. *British Journal of Psychology*, 84, 101-110.
- Ellis, H.D., Jones, D.M., & Mosdell, N. (1997). Intra- and inter-modal repetition priming of familiar faces and voices. *British Journal of Psychology*, 88, 143-156.
- Ernst, M.O. & Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429-433.
- Ernst, M.O. & Bühlhoff, H.H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8(4), 162-169.
- Etcoff, N. (1984). Selective attention to facial identity and facial emotion. *Neuropsychologia*, 22, 281-295.
- Etcoff, N., Freeman, R., & Cave, K.R. (1991). Can we lose memory for faces? Content specificity and awareness in prosopagnosia. *Journal of Cognitive Neuroscience*, 3, 25-41.

F

- Farah, M.J., Drain, H.M. & Tanaka, J.W. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception & Performance*, 21(3), 628-634.
- Farah, M.J., Levinson, K.L., & Klein, K.L. (1995). Face perception and within-category discrimination in prosopagnosia. *Neuropsychologia*, 33, 661-674.
- Farah, M.J., Rabinowitz, C., Quinn, G.E., & Liu, G.T. (2000). Early commitment of neural substrates for face recognition. In N. Kanwisher & M. Moscovitch (Ed.s) *The Cognitive Neuroscience of Face Processing* (pp. 117-123). UK: Psychology Press Ltd.
- Farah, M.J., Wilson, K., Drain, M., and Tanaka, J.N. (1995). The inverted face inversion effect in prosopagnosia: Evidence from mandatory face-specific perceptual mechanism. *Vision Research*, 35 (14), 2089-2093.
- Foxe, J.J. & Schroeder, C.E. (2005). The case for feedforward multisensory onvergence during early cortical processing. *Neuroreport*, 16(5), 419-423.
- 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.
- Friston, A.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., & Frackowiak, R.S.J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2 (4), 189-210.

G

- Gauthier, I., Skudlarski, P., Gore, J.C., & Anderson, A.W. (2000b). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3(2), 191-197.
- Gauthier, I. & Tarr, M.J. (1997). Becoming a "Greeble" expert: Exploring the face recognition mechanism. *Vision Research*, 37, 1673-1682.
- Gauthier, I., Tarr, M.J., Moylan, J., Anderson, A.W., Skudlarski, P., & Gore, J.C. (2000a). Does visual subordinate-level categorisation engage the functionally defined fusiform face area? In N. Kanwisher & M. Moscovich (Ed.s) *The Cognitive Neuroscience of Face Processing* (pp. 143-163). UK: Psychology Press Ltd.
- Gauthier, I., Williams, P., & Tarr, M.J. (1998). Training "Greeble" experts: A framework for studying expert object recognition processes. *Vision Research*, 38, 2401-2428.
- Gerlach, C., Aaside, C.T., Humphreys, G.W., Gade, A., Paulson, O.B., & Law, I. (2002). Brain activity related to integrative processes in visual object recognition: bottom-up integration and the modulatory influence of stored knowledge. *Neuropsychologia*, 40, 1254-1267.
- Gibson, J.J. (1966). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.
- Grill-Spector, K., Kourtzi, & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, 41, 1409-1422.
- Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzhak, Y., & Malach, R. (1998). A sequence of object processing stages revealed by fMRI in the human occipital lobe. *Human Brain Mapping*, 6, 316-328.

H

- Hancock, J.B., Bruce, V., & Burton, A.M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, 4(9), 330-337.
- Harman, K.L., & Humphrey, G.K. (1999). Encoding 'regular' and 'random' sequences of views of novel three-dimensional objects. *Perception*, 28, 601-615.
- Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425-2430.
- 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.
- Hay, D.C., & Young, A.W. (1982). The human face. In A.W. Ellis (Ed.), *Normality and Pathology in Cognitive Functions*. London: Academic Press.
- Hay, J.C., Pick, H.L., & Ikeda, K. (1965). Visual capture produced by prism spectacles. *Psychonomic Science*, 2, 215-216.
- Hill, H, Schyns, PG, & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition*, 62, 201-222.
- 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.
- Hole, G. (1994). Configural factors in the perception of unfamiliar faces. *Perception*, 28, 341-359.

- Hood, B.M., Macrae, C.N., Cole-Davies, V., & Dias, M. (2003). Eye remember you: the effects of gaze direction on face recognition in children and adults. *Developmental Science*, 6(1), 67-71.
- Humphreys, G.W. (1999a). Integrative agnosia. In G.W. Humphreys (Ed.), *Case Studies in the Neuropsychology of Vision*. (pp. 41-58). Hove: Psychology Press.
- Humphreys, G.W. (1999b). Neural representations of objects in space: a dual coding account. In G.W. Humphreys, J. Duncan, A. Triesman (Eds.), *Attention, Space, and Action*. (pp. 165-82). Oxford: Oxford University Press.

J

- Jacoby, L.L., & Hayman, G.A. (1987). Specific visual transfer in word identification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 13, 456-463.
- James, T.W., Humphrey, G.K., Gati, J.S., Servos, P., Menon, R.S., & Goodale, M.A. (2002). Haptic Study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia*, 40, 1706-1714.
- 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 and S. Ballesteros (Eds.), *Touch and blindness: psychology and neuroscience*. Mahwah, NJ: Lawrence Erlbaum.
- Johnston A, Hill H, Carman N (1992) Recognising faces: effects of lighting direction, inversion, and brightness reversal. *Perception*, 21, 365-375.
- Jones, B. (1981). The developmental significance of crossmodal matching. In R. Walk & H. Pick (Eds.) *Intersensory Perception and Sensory Integration* (pp. 109-130). New York: Plenum Press.

K

- Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, 3(8), 759-763.
- Kanwisher, N., & Moscovitch, M. (2000). *The cognitive neuroscience of face processing*. Psychology Press Ltd, UK
- Kanwisher, N., Chun, M.M., McDermott, J., & Ledden, P. (1996). Functional imaging of human visual cognition. *Cognitive Brain Research*, 5, 55-67.
- 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.
- Kappers, A.M.L. & Koenderink, J.J. (1999). Haptic perception of spatial relations. *Perception*, 28, 781-795.
- Katanoda,,K., Yoshikawa, K., & Sugishita, M., (2000). Neural substrates for the recognition of newly learned faces: a functional MRI study. *Neuropsychologia*, 38, 1616-1625.
- Keenan JP, Wheeler MA, Gallup GG Jr, & Pascual-Leone, A. (2000). Self-recognition and the right pre-frontal cortex. *Trends in Cognitive Sciences*, 4(9), 338-344.
- 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., & Lederman, S.J. (2002). Face recognition by hand. *Perception & Psychophysics*, 64, 339-352.
- 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 & Cognition*, 59(3), 246-257.
- Kim, J.J., Andreasen, N.C., O'Leary, D.S., Weiser, A.K., Ponto, L.L., Watkins, G. & Hitchwa, R.D. (1999). Direct comparison of the neural substrates of recognition memory for words and faces. *Brain*, 122, 1069-1083.
- Kircher TT, Senior C, Phillips ML, Benson PJ, Bullmore ET, Brammer M, Simmons A, Williams SC, Bartels M, David AS (2000) Towards a functional neuroanatomy of self processing: effects of faces and words. *Cognitive Brain Research*, 10 (1-2): 133-144.
- Klatzky, R.L. and Lederman, S.J. (1999). The haptic glance: A route to rapid object identification and manipulation. In D. Gopher & A. Koriati (Eds.) *Attention and Performance XVII: Cognitive Regulation of Performance: Interaction of theory and application* (pp. 165-196). Mahwah, NJ: Erlbaum.
- Klatzky, R.L., Lederman, S.J., & Metzger, V. (1985). Identifying objects by touch: An "expert system". *Perception and Psychophysics*, 37, 299-302.
- Klatzky, R.L., Lederman, S.J., & Reed, C.L. (1987). There's more to touch than meets the eye: The salience of object attributes for haptics, with and without vision. *Journal of Experimental Psychology: General*, 116, 356-369.

Klatzky, R.L., Loomis, J.M., Lederman, S.J., Wake, H., & Fujita, N. (1993). Haptic identification of objects and their depictions. *Perception & Psychophysics*, 54(2), 170-178.

Kleinke, C.L. (1986). Gaze and eye contact: a research review. *Psychological Bulletin*, 100, 78-100.

L

Lakatos, S. & Marks, L.E. (1999). Haptic form perception: Relative salience of local and global features. *Perception & Psychophysics*, 61, 895-908.

Leder, H. & Bruce, V. (2000). When inverted faces are recognised: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 53A, 513-536.

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

Lederman, S.J., Klatzky, R.L., & Reed, C.L. (1993). Constraints on haptic integration of spatially shared object dimensions. *Perception*, 22, 723-743.

Lederman, S.J., Klatzky, R.L., Chataway, C., & Summers, C.D. (1990). Visual mediation and the haptic recognition of two-dimensional pictures of common objects. *Perception and Psychophysics*, 47, 54-64.

Lerner, Y., Hendler, T., Harel, M., & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cerebral Cortex*, 11, 287-297.

Leube, D.T., Erb, M., Grodd, W., Bartels, M., & Kircher, T.J. (2003). Successful episodic memory retrieval of newly learned faces activates a left frontoparietal network. *Cognitive Brain Research*, 18, 97-101.

- Leveroni, C.L., Seidenberg, M., Mayer, A., Mead, L.A., Binder, J.R., & Rao, S.M. (2000). Neural systems underlying the recognition of familiar and newly learned faces. *Journal of Neuroscience*, 20, 878-886.
- Lewis, M.B. & Glenister, T.E. (2003). A sideways look at configural encoding: Two different effects of face rotation. *Perception*, 32, 7-14.
- Lewis, M.B.(2001). The lady's not for turning: rotation of the Thatcher illusion. *Perception*, 30, 769-774.
- 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. (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 & Performance*, 16, 106-120.
- 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.

M

- Malach, R., Reppas, J.B., Benson, R.R., Kwong, K.K., Jiang, H., Kennedy, W.A., Ledden, P.J., Brady, T.J., Rosen, B.R., & Tootell, R.B.H. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Neurobiology*, 92, 8135-8139.
- Malone, D.R., Morris, H.H., Kay, M.C., & Levin, H.S. (1982). Prosopagnosia: A double dissociation between the recognition of familiar and unfamiliar faces. *Journal of Neurology, Neurosurgery, & Psychiatry*, 45, 820-822.

- Marr, D. (1982). *Vision*. San Francisco: Freeman Publishers.
- Marotta JJ, Genovese CR, Behrmann M (2001) A functional MRI study of face recognition of patients with prosopagnosia. *Neuroreport*, 12(8): 1581-1587.
- Massaro, D.W. (1999). Speechreading: illusion or window into pattern recognition. *Trends in Cognitive Sciences*, 3(8), 310-317.
- Maurer, D., Le Grand, R., & Mondloch, C.J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, 6(6), 255-260.
- 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.
- McDermott, K.B. (2001). Memory, explicit and implicit. In V.S. Ramachandran (Ed.) *Encyclopaedia of the Human Brain*. Vol. 2 (pp. 773-781). London: Academic Press.
- McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.
- McNeil, J.E. & Warrington, E.K. (1993). Prosopagnosia: a face-specific disorder. *Quarterly Journal of Psychology: Human Experimental Psychology*, 46A, 1-10.
- Meredith, M.A. (2002). On the neuronal basis for multisensory convergence: a brief overview. *Cognitive Brain Research*, 14, 31-40.
- Meredith, M.A. (2004). Corticocortical connectivity of cross-modal circuits. In G. Calvert, C. Spence, & B.E. Stein (Eds.) *The Handbook of Multisensory Processes*. London: MIT Press.
- Millar, S. (2000). Modality and mind: convergent active processing in interrelated networks as a model of development and perception by touch. In M.A. Heller (Ed.), *Touch, Representation, and Blindness*. (pp. 99-142). Oxford: Oxford University Press.

Morein-Zamir, S., Soto-Faraco, S., Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Research*, 17, 154-163.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience*, 9, 555-604.

Munhall, K.G., Gribble, P., Sacco, M., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics*, 58, 351-362.

Murray, J.E. (2004). The ups and downs of face perception: Evidence for holistic encoding of upright and inverted faces. *Perception*, 33, 387-398.

N

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 FN, Ernst MO, Tjan BS, Bülthoff HH (2001) Viewpoint dependence in visual and haptic object recognition. *Psychological Science*, 12, 37-42.

Newell, F.N., & Findlay, J.M. (1997). Effects of depth rotation on object identification. *Perception*, 26, 1231-1257.

Newell, F.N., Chiroro, P., & Valentine, T., (1999). Recognising Unfamiliar Faces: The Effects of Distinctiveness and View. *Quarterly Journal of Experimental Psychology*, 52A, (2), 509-534.

O

O'Toole, A.J., Edelman, S., & Bülthoff, H.H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. *Vision Research*, 38: 2351-2363.

P

- Paller, K.A., Gonslaves, B., Grabowecky, M., Bozic, V.S., & Yamada, S. (2000). Electrophysiological correlates of recollecting faces of known and unknown individuals. *NeuroImage*, 11, 98-110.
- Pallis, C.A. (1955). Impaired identification of faces and places with agnosia for colours. *Journal of Neurology, Neurosurgery, and Psychiatry*, 18, 218-224.
- perception. *Perception*, 16, 747-759.
- Perrett, D.I., Rolls, E.T., & Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, 47, 329-342.
- Perrett DI, & Mistlin AJ (1990) Perception of facial characteristics by monkeys. In W.C. Stebbins, & M.A. Berkley (Eds), *Comparative Perception, Complex signals* Vol 2, pp 187-215. New York: Wiley.
- Peterson, M.A., & Rhodes, G. (2003). *Perception of faces, objects, and scenes: Analytic and holistic processes*. Oxford University Press, Oxford UK.
- Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M.I., Wu, W.H., Cohen, L., Guazzelli, M., & Haxby, J.V. (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.
- Platek, S.M., Keenan, J.P., Gallup, G.G. Jr., & Mohamed, F.B. (2004). Where am I? The neurological correlates of self and other. *Cognitive Brain Research*, 19(2), 114-22.
- Prather, S.C., Votaw, J.R., Sathian, K. (2004). Task-specific recruitment of dorsal and ventral visual areas during tactile perception. *Neuropsychologia*, 42, 1079-1084.

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.

Puce, A., Allison, T., Gore, J.C., & McCarthy, G. (1995). Face-sensitive regions in the human extrastriate cortex studied by fMRI. *Journal of Physiology*, 74, 1192-1199.

R

Rakoner, S.S., & Cahlon, B. (2003). *Face Recognition: Cognitive and computational processes*. John Benjamin's Publishing Co, Philadelphia.

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 & Cognition*, 25, 644-663.

Rock, I., & Harris, C.S. (1967). Vision and touch. *Scientific American*, 216, 96-104.

Roediger, H.L. III, & McDermott, K.B. (1993). Implicit memory in normal human subjects. In H. Spinnler & F. Boller (Eds.) *Handbook of Neuropsychology* (p.p. 63-131). Amsterdam: Elsevier Press.

Rotshtein, P., Henson, R.N.A., Treves, A., Driver, J., & Dolan, R.J. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience*, 8(1), 107-113.

S

Sathian, K., Zangaladze, A. Hoffman, J.M., Grafton, S.T. (1997). Feeling with the mind's eye. *Neuroreport*, 8, 3877-3881.

- Schacter, D.L. (1994). Priming and multiple memory systems: Perceptual mechanisms of implicit memory. In D.L. Schacter & E. Tulving (Eds.) *Memory Processes 1994* (pp. 233-268). Cambridge, MA: MIT Press.
- Schacter, D.L. (1990). Perceptual representation systems and implicit memory: Toward a resolution of the multiple memory systems debate. In A. Diamond (Ed.) *The development and neural bases of higher cognitive functions*. Annals of New York Academy of Sciences, no. 608 (pp. 543-571).
- Schacter, D.L. (1992). Implicit knowledge: New perspectives on unconscious processes. *Proceedings of the National Academy of Science, U.S.A.*, 89, 11113-11117.
- Schacter, D.L., & Tulving, E. (1994). *Memory Systems 1994*. Cambridge, MA: MIT Press.
- 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.
- Schweinberger, S.R., Herholz, A., Sommer, W. (1997). Recognising famous voices: Influence of stimulus duration and different types of retrieval cues. *Journal of Speech, Language and Hearing Research*, 40, 453-463.
- Schweinberger, S.R., Pfütze, E-M., Sommer, W. (1995). Repetition priming and associative priming of face recognition: evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 722-736.
- Sergent, J. & Signoret, J.L. (1992). Varieties of functional deficits in prosopagnosia. *Cerebral Cortex*, 2, 275-388.
- Sergent, J., Ohta, S., & McDonald, B. (1992). Functional neuroanatomy of face and object processing: A positron emission tomography study. *Brain*, 115, 15-36.

- Shams, L., Kamitani, S., Shimojo, S. (2002). What you see is what you hear. *Nature*, 408, 788.
- Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology*, 11, 505-509.
- Slutsky, D.A., & Recanzone, G.H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, 12, 7-10.
- Stein, B.E. & Meredith M.A. (1993). *The Merging of the Senses*. Cambridge, MA: MIT Press.
- Sutherland, N.S. (1968). Outline of a theory of pattern recognition in animal and man. *Proceedings of the Royal Society of London, B171*, (297-317).

T

- Talairach, J., & Tournoux, P. (1988). *A Co-planar, Stereotaxic Atlas of a Human Brain*. New York: Thieme Medical Publishers.
- Tanaka JW, Farah MJ (2003) The holistic representation of faces. In Peterson MA, Rhodes G (eds) *Perception of Faces, Objects and Scenes: Analytic and holistic processes*. Oxford: Oxford University Press, pp. 53-74.
- 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 & 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.
- Tarr, M.J. & Bülthoff, H.H. (1999). *Object Recognition in Man, Monkey & Machine*. (Cognition Special Issues). Cambridge, MA: MIT Press.

- Tarr, M.J., & Cheng, Y.D. (2003). Learning to see faces and objects. *Trends in Cognitive Sciences*, 7(1), 23-30.
- Tarr, MJ, & Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, 3(8), 764-769.
- Thompson, P. (1980). Margaret Thatcher - a new illusion. *Perception*, 9, 483-484.
- Troje NF, Bühlhoff HH (1996) Face recognition under varying pose: the role of texture and shape. *Vision Research*, 36,1761-1771.
- Tulving, E., & Schacter, D.L (1990). Priming and human memory systems. *Science*, 247, 301-306.
- Turk, D.J., Heatherton, T.F., Kelly, W.M., Funnell, M.G., Gazzaniga, M.S., & Macrae, C.N. (2002). Mike or me? Self recognition in a split-brain patient. *Nature Neuroscience*, 5(9), 841-842.

V

- Valentine, T. (1988). Upside-down faces: a review of the effects of inversion upon face recognition. *British Journal of Psychology*, 79, 471-419.
- VanLancker, D.R., Kreiman, J., & Cummings, J. (1989). Voice perception deficits: Neuroanatomical correlates of phonagnosia. *Journal of Clinical and Experimental Neuropsychology*, 11, 665-674.
- Vuilleumier, P., George, N., Lister, V., Armony, J., & Driver, J. (2005). Effects of perceived mutual gaze and gender on face processing and recognition memory. *Visual Cognition*, 12(1), 85-101.

- Von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A-L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience*, 17(3), 367-376.
- 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.

W

- Wallace, M.T. & Stein, B.E. (1994). Crossmodal synthesis in the mid-brain depends input from cortex. *Journal of Neurophysiology*, 71(4), 429-432.
- Welch, R.B., & Warren, D.H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88(3), 638-667.
- Welch, R.B. and Warren, D.H. (1986). Intersensory interactions. In K.R. Buff, L. Kaufman, & J.P. Thomas (Eds.) *Handbook of Perception and Human Performance: Vol. II. Sensory processes and perception*. (pp. 1-36). New York: Wiley.
- Willingham, D.B. (1997). Systems of memory in the human brain. *Neuros*, 18, 5-8.
- Winston, P.H. (1975). Learning structural descriptions from examples. In P.H. Winston (Ed.) *The Psychology of Computer Vision* (pp. 157-209). New York: McGraw-Hill.

Y

- Yarbus, A.L. (1967). *Eye Movements and Vision*. New York: Plenum Press.
- Yin R.K. (1970). Face Recognition. A dissociable ability? *Neuropsychologia*, 8, 395-402.

- Yin, R.K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81, 141-145.
- Yip AW, Sinha P (2002) Contribution of colour to face recognition. *Perception*, 31(8), 995-1003.
- Young, A.W., Hay, D.C., & Ellis, A.W. (1985). The faces that launched a thousand slips: Everyday difficulties and errors in recognising people. *British Journal of Psychology*, 76, 495-523.
- Young, A.W. (1998) *Face and Mind*. UK: Oxford University Press.
- Young, A.W., & de Haan, E.H.F. (1992). Face recognition and awareness after brain injury. In A.D. Milner & M.D. Rugg (Eds.) *The Neuropsychology of Consciousness* (p.p. 69-90). San Diego, CA: Academic Press.
- Young, A.W., Hay, D.C., McWeeny, K.H., Flude, B.M., & Ellis, A.W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, 14(6), 737-746.
- Young, A.W., Hellowell, D., & Hay, D.C. (1987). Configural information in face perception. *Perception*, 16, 747-759.
- Young, A.W., McWeeny, K.H., Hay, D.C., & Ellis, A.W. (1986b). Matching familiar and unfamiliar faces on identity and expression. *Psychological Research*, 48, 63-68.

Z

- Zhang, M., Weisser, V.D., Stilla, R., Prather, S.C., & Sathian, K. (2004). Multisensory cortical processing of object shape and its relation to mental imagery. *Cognitive, Affective, and Behavioural Neuroscience*, 4, 251-259.



School of Psychology

University of Dublin, Trinity College

Dublin 2, Ireland

Tel: +353 1 608 1886

Fax: +353 1 671 2006

Department of Psychology Ethics Committee

5th January 2003

Dear Sarah,

The Department of Psychology Ethics Committee met recently to consider your application entitled 'Cross Modal Face Recognition' for approval. I am pleased to tell you that the Committee have decided to approve your application.

Yours sincerely,

Chairperson,
Department of Psychology
Ethics Committee