Touching up on faces
-A multisensory perspective on acquiring expertise with faces-

Dissertation

zur Erlangung des Grades eines Doktors
der Naturwissenschaften

der Fakultät für Biologie
und
der Medizinischen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt
von

Lisa Whittingstall,
geb. Dopjans

aus Herdecke, Deutschland

Februar 2011

Tag der mündlichen Prüfung:	21. April 2011
Dekan der Fakultät für Biologie:	Prof. Dr. H. A. Mallot
Dekan der Medizinischen Fakultät:	Prof. Dr. I. B. Autenrieth
1. Berichterstatter:	Prof. Dr. Christian Wallraven
2. Berichterstatter:	Prof. Dr. Heinrich H. Bülthoff
Prüfungskommission:	Prof. Dr. Uwe Ilg Prof. Dr. Christian Wallraven Prof. Dr. Heinrich H. Bülthoff Prof. Dr. Marc Ernst Prof. Dr. Fiona Newell

Declaration

I hereby declare that I have produced the work entitled 'Touching up on faces - A multisensory perspective on acquiring expertise with faces', submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.

Lisa Whittingstall
Tübingen, February 10, 2011

This thesis is dedicated to my parents,

for their wholehearted and ever-present love and support,

for hour-long phone calls and care packages,

and for their persistence in trying to understand what this thesis is all about.

Acknowledgments

I would like to thank a number of people who have supported me throughout the various stages of my thesis (in one way or another) and made it an unforgetable experience:

I am very grateful

to Christian Wallraven for his constant support and encouragement.

to Heinrich Bülthoff for creating an outstanding research environment at the Max Planck Institute for Biological Cybernetics.

to Martin Breidt and Mirko Thiesen for providing invaluable technical assistance with producing stimuli.

to the 'Nikolauspflege- Stiftung für blinde und sehbehinderte Menschen' for their great co-operation. In particular, I want to thank all the blind students for their participation in our experiments and for sharing their stories with me.

to Amir Amedi, Haim, Zohar, Ran, Ella and Uri for introducing me to the Israeli way of life- and doing research.

to Andreja for our Saturday morning trips to downtown Jerusalem.

to Philipp just for being my big brother.

to Kathrin and Jozien for always coming up with a reason to have a glass- or two- of sparkling wine.

to Verena for late night ice cream in Times Square.

to John for being my swimming buddy and to Edel for the loveliest Irish accent.

to Jenny and Brian for peanut chicken.

to Kamil for the 'return of the Mack'.

to Connie, Pat and Bruce for being aaaaaawesome.

Last, but certainly not least, to my husband Kevin for his unconditional love and support and for always making me laugh!

Abstract

Faces are often considered to be different from other object categories both because of the sociobiological necessity for humans to differentiate members of their own group, and because of the differences in cognitive demands for face perception relative to general object recognition. A tremendous amount of research has shown that human observers are experts at *visual* face recognition due to specialized *visual* mechanisms for face processing that evolve with perceptual expertise. In this Ph.D. thesis, we introduce *haptic* face processing as a tool for studying the mechanisms underlying face processing expertise from new angles.

In the first study, we tested face recognition within and across the haptic and visual modalities. We showed that both the haptic and visual systems have the capacity to process faces, and that face-relevant information can be shared across sensory modalities. Interestingly, we found this information transfer across modalities to be asymmetric and limited by haptic face processing. In the second study, we investigated whether and how the way in which haptic information is gathered affects the encoding, processing and storage of that information and consequently how it affects face recognition performance. This was achieved by using a gaze-restricted display to constrain the visual system to sequential, self-directed exploration, promoting serial encoding in vision in much the same way that the haptic system encodes

objects. We found that face recognition was equally disrupted using gaze-restricted vision and haptics as compared to unrestricted vision with a clear switch from configural to featural processing. In the third study, we focused on the role of perceptual expertise on information processing of serially encoded faces using a gaze-restricted display, i.e. whether participants can learn to efficiently recognize faces that are serially encoded. We showed that practice with a previously novel way of perceiving faces, i.e. through serial encoding, can lead to some of the recognition effects typically associated with unrestricted visual face recognition, indicating that - at least for vision - serial encoding of information might allow for expert face processing. In the fourth study, we used a different approach to the question how perceptual expertise shapes efficient face processing strategies by studying haptic face recognition in the sighted, congenitally blind, and acquired blind. Our results demonstrated the crucial role of visual input for the development of efficient face processing capabilities inasmuch as a lack of relevant visual experience cannot be compensated for or improved by purely perceptual haptic expertise.

Overall the thesis highlights how modality-specific differences in information acquisition affect processing strategies in high-level tasks, such as face recognition, and under which conditions those differences can be compensated for by perceptual expertise. Studying haptic face processing (and its contribution in a cross-modal context) is, therefore, a valuable new tool for studying face processing itself, the mental representation and the role of perceptual expertise, but also the differences and commonalities of information processing in the visual and haptic modalities.

Zusammenfassung

Gesichter werden oftmals als verschieden von anderen Objektkategorien betrachtet, sowohl aufgrund der soziobiologischen Notwendigkeit für Menschen Mitglieder ihrer eigenen Spezies zu unterscheiden als auch aufgrund der unterschiedlichen kognitiven Anforderungen für Gesichtswahrnehmung im Vergleich zur generellen Objekterkennung. Umfangreiche Studien haben gezeigt, dass menschliche Betrachter Experten für *visuelle* Gesichtserkennung sind, basierend auf spezialisierten *visuellen* Mechanismen für die Verarbeitung von Gesichtern, die sich mit perzeptueller Expertise entwickeln. In dieser Dissertation stellen wir *haptische* Gesichtserkennung als eine neue Methodologie zur Untersuchung der Mechanismen vor, die der Expertise für Gesichtsverarbeitung zu Grunde liegen.

In der ersten Studie testeten wir Gesichtserkennung innerhalb und über die visuelle und haptische Modalität hinweg. Wir konnten zeigen, dass sowohl das haptische als auch das visuelle System zur Verarbeitung von Gesichtern fähig sind und dass gesichtsbezogene Informationen zwischen den sensorischen Modalitäten geteilt werden kann. Interessanterweise war der Transfer von Informationen asymmetrisch und beschränkt durch haptische Informationsverarbeitung. In einer zweiten Studie untersuchten wir ob und wie die Art und Weise in der haptische Information gesammelt wird Enkodierung, Verarbeitung und

Speicherung dieser Information und, folglich, die Gesichtserkennungsleistung, beeinflusst. Dies wurde möglich durch die Verwendung eines Displays, welches das Gesichtsfeld einschränkte, um das visuelle System auf serielle, selbstgesteuerte Exploration zu begrenzen. Hierdurch wurde die serielle Enkodierung im Visuellen gefördert, ähnlich der Art und Weise in der das haptische System Objekte enkodiert. Wir fanden dass haptische und visuell beschränkte Gesichtserkennung im Vergleich zu normaler visueller Gesichtserkennung gleichermaSSen beeinträchtigt waren, einhergehend mit einem deutlichen Wechsel von globalen (sog. konfiguralen) hin zu lokalen (sog. merkmalsbasierten) Strategien der Informationsverarbeitung. In einer dritten Studie untersuchten wir die Rolle der perzeptuellen Expertise im Hinblick auf Informationsverarbeitung von seriell enkodierten Gesichtern unter Verwendung des visuell beschränkten Displays, also, ob Versuchsteilnehmer erlernen können seriell enkodierte Gesichter effizient zu erkennen. Wir zeigten, dass Übung mit einer im Vorfeld neuartigen Art und Weise Gesichter wahrzunehmen (d.h. durch serielle Enkodierung), zu einigen, für unbeschränkte visuelle Gesichtserkennung typischen Erkennungseffekten führen kann. Dies deutet daraufhin, dass - zumindest für das Visuelle - serielle Enkodierung von Informationen Gesichtsverarbeitung auf Expertenniveau ermöglicht. In der vierten Studie verfolgten wir einen anderen Ansatz, um die Frage zu beantworten wie perzeptuelle Expertise effiziente Gesichtsverarbeitungsstrategien formt, indem wir Gesichtserkennung in Sehenden, Geburtsblinden und Erblindeten untersuchten. Unsere Ergebnisse unterstrichen die Wichtigkeit des visuellen Inputs für die Entwicklung von effizienten Gesichtsverarbeitungsfähigkeiten insofern als ein Fehlen von relevanter visueller Erfahrung nicht durch rein haptische perzeptuelle Expertise kompensiert oder verbessert werden kann.

Zusammenfassend zeigt die Dissertation wie modalitäten-spezifische Unterschiede in der Informationserfassung Verarbeitungsstrategien bei der Gesichtserkennung beeinflussen und unter welchen Voraussetzungen solche Unterschiede durch perzeptuelle Expertise kompensiert werden können. Die Untersuchung von haptischer Gesichtserkennung (und ihr Beitrag in einem kross-modalen Zusammenhang) ist daher eine wertvolle neue Methodologie um Gesichtsverarbeitung selbst, deren mentale Repräsentation und die Rolle von perzeptueller Expertise, aber auch die Unterschiede und Gemeinsamkeiten der Informationsverarbeitung in der visuellen und haptischen Modalität, zu untersuchen.

Statement of Personal Contribution

All studies are published or submitted under the name Lisa Dopjans.

Study 1: Dopjans, L., Wallraven, C., and Bülthoff, H.H. (2009). Cross-modal transfer in visual and haptic face recognition. *IEEE Transaction on Haptics, 2*, 236-240

I designed the experiments together with C. Wallraven, and H.H. Bülthoff. I created all the stimuli, programmed and carried out all experiments, performed all data analysis, and wrote the paper.

Study 2: L. Dopjans, H.H. Bülthoff and C. Wallraven (2010). Serial exploration of faces: Comparing vision and touch. *submitted*

I designed the experiments together with C. Wallraven, and H.H. Bülthoff. I created all the stimuli, programmed and carried out all experiments, performed all data analysis, and wrote the paper.

Study 3: L. Dopjans, H.H. Bülthoff and C. Wallraven (2011). Learning to recognize faces through serial exploration. *submitted*

I designed the experiments together with C. Wallraven, and H.H. Bülthoff. I created all the stimuli, programmed and carried out all experiments, performed all data analysis, and wrote the paper.

Study 4: L. Dopjans, C. Wallraven and H.H. Bülthoff. Visual experience is necessary for efficient haptic face recognition.

I designed the experiments together with C. Wallraven, and H.H. Bülthoff. I created all the stimuli, programmed and carried out all experiments, performed all data analysis, and wrote the paper.

Table of Contents

Chapter 1

Introduction

Visual face recognition is of strong evolutionary significance across many biological species, including humans, as the face carries different categories of information that are all critical to survival. For example, our remarkable face-recognition skills are used to choose a mate, to communicate with others, and to differentiate between friend or foe. In fact, extensive research has established that face processing is an essential function of visual perception, to such an extent that faces are visually processed with an expertise that surpasses general object recognition. Amongst the hallmarks that have been proposed to distinguish expert face recognition mechanisms from general object recognition are:

- Faces more than other objects require the ability to make fine-level discriminations (Damasio (1990); Damasio et al. (1982)). This inequality is the result of differences in both task demands and stimulus characteristics. Specifically, face recognition is more demanding of discrimination processes due to face identification at an individual level (as opposed to a basic level for other object classes) and the homogeneity of faces as a stimulus class (Rosch et al. (1976)).

- Faces are processed holistically (each component part is processed in interaction with multiple other parts; Rhodes (1988); Maurer et al. (2002); Davidoff et al. (1986)).
- Expert face recognition relies on configural processing (the encoding of spatial relations between features). Spatial distances between features are particularly diagnostic for faces as all faces share a common global configuration of features (e.g., eyes above nose above mouth; Tanaka and Sengco (1997)).

Not surprisingly, almost all of the face research from which these general principles derive involves the visual system. In terms of general object recognition, however, previous research suggests that information from *multiple sources of sensory inputs* contributes to object representations (Easton et al. (1997); Newell et al. (2001); Reales and Ballesteros (1999)). For example, objects are not only recognized visually, but can also be recognized using the sense of touch, especially when objects are actively explored haptically (Klatzky et al. (1987)). In this context, little is known about how face information is integrated across modalities and whether information from other sensory modalities can contribute to the formation of robust face representations. Functional models of visual face recognition (e.g., Bruce and Young (1986)) suggest that structural information from a face (a feature-based description together with a representation of the spatial arrangement or configuration of those features) is encoded and represented in face memory for person identification. Since touch can also encode structural information (e.g., Lederman and Klatzky (1990)), this information could, in principle, also contribute to face memory. In fact, although both the visual and haptic systems are able to extract many properties of objects and faces, to efficiently recognize them, both systems rely heavily on shape features (Kilgour and Lederman (2002); Klatzky et al. (1987)).

On the other hand, as the receptor and peripheral nerve systems of the eyes and skin are quite different, it is no surprise that general object recognition performance using haptics and vision also comprises differences. Perhaps the most salient of these differences is the time required for haptic recognition, especially for objects where the features are spatially isolated, such as faces (Kilgour and Lederman (2002)). In this case, the visual receptor array of the eye has an advantage over the somatosensory array of the fingers, because it can sample large parts of the environment simultaneously. Haptics, in contrast, must usually sample spatially separated features sequentially, imposing higher working memory demands on haptic exploration of objects (Loomis (1981)). This difference suggests that visual object recognition may involve very fast or even simultaneous spatial feature integration, so that both local facial features and their global configuration can be rapidly processed (Tanaka and Sengco (1997)), whereas haptic object recognition may involve serial feature integration (Loomis et al. (1991); Loomis and Lederman. (1986)). If either of these factors, serial encoding and/or higher working memory demands, limits performance, as seems likely, this effect would be exaggerated for a stimulus class such as faces, which requires analysis of several spatially isolated features for successful recognition.

Given these limitations of the haptic modality, the question arises whether the latter is actually capable of not just recognizing faces but employing hallmarks of expert face recognition such as fine-level discriminations and holistic and configural processing. Inasmuch as we have little to no training in haptic face recognition throughout life, it is possible that participants might be able to develop strategies to compensate for processing differences introduced by serial encoding. After all, it is well established that visual expert face processing takes many years to develop and requires training and experience (LeGrand et al. (2003); Carey and Diamond (1977); Dahl et al. (2009); Hay and Cox (2000); Maurer et al. (2002); Mondloch et al. (2003); Pellicano and Rhodes (2003); Schwarzer (2000)).

The human brain is remarkably plastic in its learning capabilities. Not only can information from different sensory modalities be integrated efficiently (e.g., Ernst and Banks (2002)) but if this information is conflicting, for example due to sensory manipulations such as prism adaptation, compensatory strategies like remapping can be employed. For example, spatial and even temporal re-mappings are quickly adapted to in the visual (e.g., Ghahramani et al. (1996); Redding et al. (2005)) and auditory (e.g., Cunningham et al. (2001)) domains. In addition the brain is even capable of learning totally new object classes (e.g., Gauthier and Tarr (1997)) and sensory modalities (for example, a 'sense of direction') as demonstrated by sensory substitution and augmentation studies (e.g., Nagel et al. (2005)). Here we ask whether the brain can not only learn how to recognize faces by touch but also whether we can become experts in this new task. Studying haptic face recognition (and its contribution in a cross-modal context) is, therefore, a valuable new approach that will provide new insights not only into face processing itself, the mental representation and the role of expertise, but also into the differences and commonalities of information processing in the visual and haptic modalities, and consequently the capabilities and limitations of these modalities themselves. In this thesis, we further investigated the mechanisms underlying (haptic) face processing from new angles using psychophysical experiments. We began by defining a fully controlled stimulus set of 3D face masks. In a first study, we tested face recognition within and across the haptic and visual modalities. In a second study, we investigated whether and how the way in which haptic information is gathered affects the encoding, processing and storage of that information (i.e., a question that is related to the serial vs. holistic manner of perceiving haptic stimuli), and consequently how it affects performance in a high-level cognitive task such as face recognition. In a third study, we focused on the role of perceptual expertise on information processing of serially encoded faces using a gaze-restricted display, i.e. whether participants can learn to efficiently recognize faces that are serially encoded, for example, by learning to accurately integrate information gained through serial encoding into a more global representation. In a fourth study, we used a different approach to the question how perceptual expertise shapes efficient face processing strategies by studying face recognition in the blind. More specifically, we investigated haptic face recognition in the sighted, congenitally blind, and acquired blind to assess the role of visual input for the development of efficient face processing capabilities.

The thesis is structured as follows: we first provide an overview of the relevant literature to provide a theoretical framework for the work conducted in this thesis, followed by a detailed description of haptic stimulus creation, then present the four studies conducted in the scope of this thesis and close by a general discussion of these studies and future avenues of research in (haptic) face recognition.

Chapter 2

Related work

This section provides an overview of the relevant literature in the following four research areas to provide a theoretical framework for the work conducted in this thesis:

- Haptic object recognition
- Visual face recognition
- Haptic face recognition
- Perceptual expertise

2.1 Haptic object recognition

In this section we will briefly review some of the important properties of haptic object recognition. In the following, we refer to 'haptic processing' as an information-processing perceptual system that combines input from both the cutaneous (mechanoreceptors embedded in the skin) and kinesthetic (mechanoreceptors located within the body's muscles, tendons and joints) systems and that is associated in particular with active touch (Loomis and Lederman. (1986)).

Klatzky, Lederman and colleagues carried out seminal work on haptic object recognition. They found that people are extremely accurate and very fast at haptically identifying three-dimensional common objects (Klatzky et al. (1985)). Similarly, in a study with five-year olds, Bushnell and Baxt (1999) demonstrated that children were virtually error-free at haptically discriminating between previously presented and newly presented common objects. Such work serves as an existence proof that the human haptic system can perform certain aspects of object processing surprisingly well. In further experiments, which investigated the nature of the haptic representations of objects in memory and the underlying processes by which those representations are stored and manipulated, Lederman and Klatzky (1987, 1990) observed participants' hand movements as they identified common objects haptically. They found that people use a number of distinctive movements, so-called exploratory procedures (EPs), which they classified into six types (lateral motion, pressure, static contact, holding, enclosure, and contour following). They also showed that the ease of extracting object properties varies according to the EP used. For instance, texture is extracted best by lateral motion, a back-and-forth rubbing motion of the fingers over a surface, whereas this EP provides little or no information about object shape. In contrast, gripping objects in the hand allows for fast extraction of global shape and texture, whereas detailed shape information is harder to perceive by gripping. Following an object's contour, although time-consuming, allows for extraction of both texture and exact shape.

Klatzky and Lederman (1993) suggested a hierarchical organization of object properties extracted by the haptic system. At the highest level the distinction is made between geometric properties of objects and material properties. Geometric properties, which can be divided into shape and size, are specific to particular objects whereas material properties such as texture, compliance and temperature are independent of any one sampled object. A number of studies by Klatzky, Lederman and colleagues point to the importance of material properties in identification and similarity judgments (see, for example, Klatzky and Lederman (2000)). Klatzky et al. observed that participants who were freely identifying common objects often reported attending to their material properties (Klatzky et al. (1985)). In a follow-up study they found that participants who explored common objects with a single finger, while wearing a heavy glove, improved significantly when the tip of the glove was cut off to expose the object's material (Klatzky and Lederman (1993)). Finally, Klatzky and Lederman (1995) found that a 200 ms touch with the fingertip was sufficient to identify 25% of a set of objects selected to have large surfaces and to be particularly identifiable by texture (e.g., sandpaper).

The work of Klatzky & Lederman on object property extraction and exploratory procedures was extended by a series of studies by Cooke et al. (2007, 2010). In these studies the perceptual effects of changing EPs were visualized and quantified using a multidimensional scaling (MDS) framework that was developed for studies of cross-modal human perception and validation of computer vision algorithms (Cooke et al. (2005, 2007)). They used a set of parametrically-defined, novel 3D objects to investigate perceptual similarity by having participants provide similarity ratings after exploring the objects using one of four EPs (contour-following, gripping, tip-touching, or lateral motion on the objects' centers). The results showed that the specific type of hand movement used to explore the objects does indeed affect perceptual similarity, dependent upon the dimensions which can physically be extracted using a specific hand movement. Specifically, lateral motion on the objects' centers, which does not provide any information about global shape changes, yielded one-dimensional perceptual representations in which the single dimension corresponded to texture. In contrast, two perceptual dimensions, shape and texture, were needed to explain similarity data when participants gripped the objects, followed their contours, or touched their tips.

In another study, Cooke et al. (2007) investigated how perceptual similarities and categorization vary when different sensory modalities are used to explore objects. Subjects explored the objects using either vision alone, touch alone, or both vision and touch. Despite sharing common dimensions and ordinal relationships, there were two clear differences amongst modality-specific perceptual spaces. First, compared to the visual condition, larger individual differences were observed in similarity weights used in the haptic condition and even greater differences in the visuohaptic condition. Second, the relative weights of shape and texture dimensions differed: on average, shape dominated texture when objects were seen, while shape and texture were roughly evenly-weighted when objects were either touched, or both seen and touched. This finding agrees with the notion that vision is specialized for the extraction of object macro-geometry (Lederman et al. (1996)).

In addition, Lakatos and Marks (1999) investigated whether participants emphasize the local features or the global form during exploration of 3D objects. The task was to make similarity judgments of unfamiliar geometric forms (e.g. cube, column) which contained distinctive local features such as grooves and spikes. The data suggests greater salience for local features in early processing, with global features becoming more equal in salience as processing time increased. Objects with different local features, but similar in overall shape, were judged less similar when explored haptically than when vision was available. Longer exposure time (increasing from 1s to 16s), however, produced greater similarity ratings for objects that were locally different but globally similar, indicating the increasing salience for global shape over time.

Furthermore, Newell et al. (2001) have found evidence for orientation dependency in haptic object recognition. When people extract local features of 3D objects, they appear to have a bias toward encoding the back of the object - the reverse of vision. Participants viewed or haptically explored objects made of LEGO blocks and then tried to recognize the one they had been exposed to. On some trials, the objects were rotated 180° between exposure and recognition test. When exposure and test were in the same modality (vision or touch), performance suffered if the objects were rotated. For cross-modal recognition, however, performance was better when the objects were rotated. Moreover, when exposure and testing were exclusively in the haptic domain, performance was better for objects explored from the back than those explored from the front, indicating that haptic object recognition is indeed orientation specific.

2.2 Visual face recognition

Faces are often considered to be different from other object categories both because of the sociobiological necessity for humans to differentiate members of their own group, and because of the differences in cognitive demands for face perception relative to general object recognition (Sergent et al. (1992)). A tremendous amount of research demonstrates functional and cortical specialization for faces in a variety of populations, including normal adults, patients, infants and monkeys; these studies used a variety of methodologies, including behavioral, imaging, electrophysiological and single-cell recording techniques. In the following we will briefly discuss some of the important properties of visual processing of faces in relation to processing of other object categories, going from low-level to high-level characteristics. As an exhaustive review of the extensive literature on visual face recognition is outside the scope of this thesis, the following discussion is intended to provide a context for the work presented in this thesis.

Visual face recognition has been shown to be remarkably robust in that it is strikingly tolerant to resolution reduction (Harmon and Julesz (1973); Harmon (1973); Yip and Sinha (2002)) and that this ability to handle degradations increases with familiarity (Burton et al. (1999); Roark et al. (2003); Liu et al. (2003)). However, it remains unclear how increased experience with a given individual leads to an increase in the robustness of the encoding. In the same context, high-spatial frequency information (e.g., images which contain only contour information) has been shown to be an insufficient cue for human face recognition on its own (Davies et al. (1978)). In fact, it seems to be the presence of additional photometric cues, not just contour information that leads to the recognizability of human generated line-drawings (Bruce and Young (1998); Bruce et al. (1992); Pearson and Robinson (1985)).

Recent studies have investigated whether shape or pigmentation is more important for face recognition. The approach taken has been to create sets of faces that differ from one another in terms of only their shape or only their pigmentation using either laser-scanned models of faces (O'Toole et al. (1999)), artificial faces or morphing photographs of faces (Russell et al. (2004)). For each of these classes of stimuli, similar performance levels have been found experimentally providing evidence that the two kinds of cues are used equally in face recognition. In this context, it has also been found that face-shape appears to be encoded in a slightly caricatured manner. Rhodes (1996) demonstrated that 'caricatured' versions of faces support recognition performance at least equal to or better than that achieved with veridical faces, even if the exaggerated deviations were created in shape alone (Brennan (1985)) or in a combination of both shape and pigmentation cues (Benson and Perrett (1991)). These results have been taken to suggest a norm-based representational space for faces, often referred to in the literature as 'face-space' (Valentine (1999)).

As mentioned above, both pigmentation and shape are about equally important in visual face recognition. However, Yip and Sinha (2002) have shown that when shape cues are compromised (e.g., by reductions in resolution), the brain relies on color cues to determine identity. In such circumstances, recognition performance with color images is significantly better than with gray-scale images. They suggest that either color provides diagnostic information important for face identification or that color might facilitate low-level image analysis (e.g. segmentation) and thus indirectly aid face recognition.

An additional focal point in research on face processing has been whether face processing relies primarily on the facial configuration (i.e., features and their spatial interrelations) or more on the features themselves. Much of the visual face research has unequivocally emphasized the primacy of holistic or configural processes (e.g., Gelder and Rouw (2000); Freire et al. (2000)). The difference between holistic and configural is not precisely understood (and the terminology differs among researchers, e.g. Maurer et al. (2002); Gauthier and Tarr (2002)), but it is established that, in comparison to objects, processing of faces involves (i) a stronger and mandatory perceptual integration across the whole and (ii) a more precise representation of the 'second-order' deviations from the basic ('first-order') shape, including precise spatial-relational information (e.g., the distance from corner of left eye to tip of nose) and precise feature shape (Yovel and Kanwisher (2005)). To investigate this configural (Tanaka and Farah (1993)) recognition strategy, researchers have predominantly used the inversion effect, which is defined as a decrease in performance when recognizing inverted as opposed to upright faces (Valentine (1988); Yin (1969)). Yin trained adults on a series of faces which later had to be identified from pairs made up of seen and unseen faces. Performance in the test phase was high when these faces were presented upright but suffered remarkably when all faces were inverted. The difference in performance was much smaller for non-face objects like houses, suggesting that this is not a characteristic of general object recognition but may be face specific (Diamond and Carey (1986)). The predominant explanation for the decrease in face recognition performance, induced by vertical inversion, is that this transformation selectively impairs our ability to extract configural information from faces, while leaving featural processing largely intact (Leder and Bruce (2000)). This featural processing of inverted faces is illustrated by the well-known Margaret Thatcher Illusion (Thompson (1980), Figure 2.1). To date, disproportionate effects of inversion remain one of the most cited pieces of evidence for face specificity, and inversion continues to be widely used as a control condition to test for the specificity of a particular process (see Carey and Diamond (1977) for early replication of a disproportionate effect for faces compared to houses, building and dog faces; Scapinello and Yarmey (1970); Valentine and Bruce (1986); Yarmey (1971)). However, it remains unclear what accounts for the difficulty in recognizing inverted faces, as the same information is present in both, inverted and upright faces.

Figure 2.1: The Margaret Thatcher Illusion, (Thompson, 1980).The eyes and mouth of the image on the right have been vertically inverted. When the whole face is inverted as well, this manipulation is not apparent. If the reader turns this page around, however, the manipulation is grotesquely obvious.

Interestingly, the inversion effect is not present in infants. In fact, a number of studies have shown that this pattern of results (i.e. a remarkable deficit in recognition of inverted faces, but no such deficit for inverted images of non-face objects) takes many years to develop (Carey and Diamond (1977); Hay and Cox (2000); Maurer et al. (2002); Mondloch et al. (2003); Pellicano and Rhodes (2003); Schwarzer (2000)). Whereas six year old children do not show an inversion effect, eight year olds show some and ten year old children exhibit near adult-like performance. However, although six year olds are not sensitive to inversion, they are susceptible to the Thatcher Illusion (Thompson (1980); Lewis (2003)), suggesting, perhaps, that configural information is important for face processing throughout child development, but that this information has not been integrated yet into the face recognition system. This pattern of behavior suggests that a shift in strategy occurs over the course of several years. Initially, infants and toddlers seem to adopt a largely feature-based strategy for recognizing faces. Gradually, a more sophisticated strategy involving configural information processing evolves. Additionally, this finding serves as indirect evidence for the role of configural information in achieving the robust face recognition performance that adults exhibit.

Sadr et al. (2003) and others (Davies et al. (1977); Fraser et al. (1990)), however, have shown that when taken alone, features are sometimes sufficient for facial recognition in that just one feature (such as eyes, or the mouth) can be enough for recognition of many famous faces. Not all facial features, however, contribute equally to the representations underlying identity assessments. A variety of techniques have been employed to assess the relative salience of facial features. These include, for example, the study of eye movements, psychophysical experiments with spatially filtered stimuli, multidimensional scaling, and the use of subjective questionnaires. In general, the region around the eyes appears to be most important for visual face recognition (e.g., Keating and Keating (1982); Leder et al. (2001); Mangini and Biederman (2004); Schyns et al. (2002); Sekuler et al. (2004)); more precisely, people visually attend foremost to the eyebrows (Schyns et al. (2002); Sadr et al. (2003)), followed in descending order of importance by the eyes, mouth, and, finally, nose (Fraser et al. (1990); Haig (1986); Janik et al. (1978)). However, despite the important role of certain facial features, Sinha and Poggio (1996) have demonstrated the insufficiency of internal features and even their mutual configuration, while highlighting the perceptual importance of the full head configuration for face recognition.

In addition to spatial cues, the human visual system has been demonstrated to use a number of temporal cues for face recognition. Recognizing a face across variations in viewing angles is a very challenging task which the human visual system can solve with remarkable ease. It has been suggested that temporal association forms the basis for linking different images of the same object into a coherent and consistent object representation. Wallis and Bülthoff (2001) demonstrated this role of temporal association in learning of faces in a series of experiments. They briefly exposed participants to movies containing a rotating head which morphs between one individual and another as it rotates from frontal to profile views. The results showed that this impaired participants' ability to distinguish between the two faces contained in the sequence. Generally, rigid motion (e.g., from a camera moving around a motionless head) has been shown to facilitate recognition of previously viewed faces (Schiff et al. (1986); O'Toole et al. (2002)), though with only little, if any, benefit of seeing these views during the learning phase. Non-rigid facial motion (exhibited in emotive facial expressions or speech movements), however, has been demonstrated to facilitate face recognition significantly (Lander and Chuang (2005)). Using subtle morphs of form and facial motion in unfamiliar faces, Knappmeyer et al. (2003) showed that non-rigid facial motion from one face applied to the form of another face can bias an observer to misidentify the latter as the former. These effects come from a natural sequence of moving images, not merely from having more views available. It has, therefore, been suggested that dynamic cues from talking and expressive movements provide information about aspects of facial structure that go beyond the information contained in multiple viewpoints.

Further evidence for the view of faces as a 'special' visual object class was provided by the observation that face recognition emerges developmentally early. Newborns selectively gaze at 'face-like' patterns (e.g., three dots within an oval that represent the two eyes and a mouth) only hours after birth, whereas they do not pay much attention to impossible faces (e.g., vertical inversion of the triad of dots). This finding suggests some innately specified representation for faces (Johnson et al. (1991); Farah et al. (1998); Johnson and Morton (1991)). However, as a counterpoint to this idea of innate preferences for faces, Simion et al. (2001) have shown that newborns consistently prefer top-heavy patterns (e.g. a T-shape) over bottom-heavy patterns (e.g. an inverted T-shape). It is unclear whether this is the same preference exhibited in earlier work, and if it is whether it is face-specific or related to some other general-purpose preference. Levine et al. (1988) argued that, relative to other visual stimuli, humans have developed an expertise through constant exposure to faces and practice in recognizing and differentiating them.

On the basis of the results discussed above, it is suspected that unique cognitive and neural mechanisms may exist for face pro-

cessing in the human visual system and most researchers agree that a subset of visual abilities, sometimes referred to as a face module, may be dedicated to such processing (e.g., Kanwisher et al. (1997)). Indeed, there is a great deal of evidence that the primary locus for human face processing may be found in the fusiform gyrus of the extrastriate visual cortex (McCarthy et al. (1997)). This region shows an intriguing pattern of selectivity (schematic faces do not give rise to much activity) and generality (animal faces do elicit a good response) (Tong et al. (2000)), suggesting a strong domain-specific response for faces (McKone et al. (2007); Kanwisher and Yovel (2006)). In keeping behavioral results, the 'fusiform face area' (FFA) also appears to exhibit an 'inversion effect' (Kanwisher et al. (1998)) making it an ideal candidate for a dedicated face processing module. However, a second line of evidence suggests that rather than being a true 'face module' the FFA may be an 'expert module' responsible for performing either subordinate or expert-level categorization of generic objects (Diamond and Carey (1986); Gauthier and Tarr (1997); Gauthier et al. (1999b, 2000a, 2003); Xu (2005)).
In the same context, researchers have addressed the question how faces are mentally represented within the FFA. A popular model is the concept of a 'face space' (Valentine (1991); Burton and Vokey (1998)), which is considered to be a multi-dimensional space whose dimensions represent the physiognomic features that are used to encode faces - i.e., correspond to ways in which faces can vary. The properties of this structure can be used to explain various effects with faces, such as the own-race bias and the effects of typicality or distinctiveness on the recognition of unfamiliar faces (e.g., Hancock et al. (1996); Vokey and Read (1992, 1995)). These theoretical proposals, however, are made in the absence of any proposal about the nature of the underlying

dimensions but rely, instead, on the presumably general abstract properties of any multi-dimensional space. Plausible candidates, however, might include dimensions that code for hair color and length, face shape and age, or, alternatively might be based on statistical properties of face images such as those extracted by principal component analysis. Closely associated with the dimensional structure of face space is the intuition that the distribution of faces within this space is not uniform but, rather, increases in density toward the centre, reaching maximum density at the centre. Therefore, typical faces are thought to be located in areas of high density closer to the centre of the space, and distinctive faces to be located in the more rarefied periphery. Hence, the typicality of a face is an inverse function of the distance of the face from the centre of the space. Given these notions, a common conclusion about face space is that the majority of faces occupies the region of highest density, implying that the majority of faces would therefore be typical. These findings have been confirmed by fMRI studies in humans (Loffler et al. (2005)) and extended in neurophysiological studies in monkeys (Giese and Leopold (2004)). Loffler et al. (2005) provided evidence for a representation in which individual faces are encoded by their direction (facial identity) and distance (distinctiveness) from a prototypical (mean) face. They found an increase in the fMRI signal with increasing distance from the mean face, when they varied facial geometry (head shape, hair line, internal feature size and placement). In addition, adaptation of the fMRI signal showed that the same neural population responds to faces falling along single identity axes within this space. Giese and Leopold (2004), on the other hand, elucidated the neural principles of the encoding of face spaces in visual cortex. They tested two models realizing example-based and norm-referenced encoding by comparing the experimentally measured tuning properties of neurons in macaque area IT with predictions from the two models. As they found a better agreement with the norm-referenced encoding model, their results were taken to suggest that a majority of IT neurons might represent deviations from a norm-face, which is determined by an average over the distribution of typically occurring faces.

In short, hallmarks of visual processing of faces, relative to visual processing of other object categories are that it is highly practiced (e.g., Gauthier et al. (2003)), based predominantly on overall configuration (e.g., Maurer et al. (2002)), orientation specific (e.g., Farah et al. (1995a); Sergent (1984)), and identity specific (e.g., Rosch et al. (1976)) and may be subserved by specialized cognitive and neural mechanisms (e.g., Kanwisher et al. (1997)).

2.3 Haptic face recognition

Recent research reveals that humans are also capable of haptically recognizing both facial identity and facial expression of emotions in live faces, 3D face masks (rigid molds taken from live faces) and 2D raised-line depictions. In this section, we will briefly review the literature on haptic face recognition and its relation to vision to provide a context for the work presented in this thesis.

Kilgour and Lederman (2002) extended the study of face processing beyond the visual domain, investigating participants' capability to identify unfamiliar live human faces and rigid face masks using only their sense of touch. Participants haptically matched the facial identity of live actors with a success rate of

79% (chance = 33%). When rigid face masks were used, accuracy declined to 59%, implicating material-specific properties of the face as important sources of haptic information about facial identity. Since, however, performance with the 3D masks remained well above chance, the authors concluded that haptic face encoding at the subordinate level involves both geometric and material properties with greater emphasis being placed on the geometric cues (3D structural information). This study constitutes an existence proof that people are indeed able to recognize faces haptically, and in doing so, offers initial support for the proposal that face recognition could even be a bimodal phenomenon. Other researchers have since confirmed this initial result using 3D face masks (Casey and Newell (2005); Kilgour et al. (2004, 2005); Pietrini et al. (2004)). The object information that accounts for our ability to recognize faces haptically might be derived from structurally invariant 3D contours and distinctive material properties, such as skin texture and compliance (perceived in terms of skin smoothness, softness, firmness, etc.).

Given that humans have little to no experience with haptic face recognition, does exploring a face manually really constitute face recognition? Kilgour et al. (2004) provided the first piece of evidence when they reported the first known case of haptic prospagnosia (a condition in which individuals have considerable difficulty in recognizing faces). Their study required a prosopagnosic patient, with previously documented poor visual performance, and neurologically intact controls to haptically match 3D face masks. The patient's ability to identify faces haptically was as deficient as it was visually, with similar patterns of matching impairment: matching accuracy was at chance level only, response times were significantly slower than those of normal controls, and he demonstrated a paradoxical inversion effect (i.e., better performance for inverted than upright faces) visually and haptically (for possible neural explanations for the paradoxical inversion effect, see Farah et al. (1998); Gelder and Rouw (2000)). These results were taken to indicate that face-processing deficits can be found across different input modalities.

Having established that humans can haptically discriminate and identify faces at levels well above chance, two core questions about haptic face processing have been considered to date: (1) how does the haptic system process facial identity, i.e., what is the relative importance of configural (the layout of the face parts), as opposed to feature-based (isolated face parts) processes in the haptic perception of facial identity and how does it relate to vision, and (2) how does the haptic system represent facial identity and can this representation be shared across the haptic and visual modalities?

Kilgour and Lederman (2006) used a haptic version of the face inversion paradigm to assess the role of configural versus featural processing in haptic face recognition. Participants performed a temporally unconstrained matching task with upright and inverted 3D clay face masks and non-face control objects (teapots). Their results demonstrated a significant haptic inversion effect for faces in terms of accuracy, but none for teapots. When they restricted exploration time to 10s in a small follow-up experiment, forcing participants to adopt local feature-based exploration strategies (previously shown by Halberstadt et al. (2003) for vision), however, they failed to find an inversion effect for faces. Taken together, this pattern of results was taken to indicate that, to this extent, haptic and visual processing of facial identity are similarly influenced by orientation and that participants might use configural (as opposed to feature-based) processing more with upright than with inverted faces, paralleling the results for visual face recognition (Gelder and Rouw (2000); Freire et al. (2000)).

Recently, McGregor et al. (2010) studied the relative importance of configural, as opposed to feature-based, processes in the haptic perception of facial identity using 2D raised-line drawings in a face-identity learning task involving scrambled, as well as upright and inverted faces. The upright and scrambled displays produced equivalent performance. Because scrambling faces alters the global facial configuration, McGregor et al. concluded that it was not used to haptically process facial identity portrayed in 2D raised-line drawings. Scrambled faces also produced higher accuracy than inverted faces. Since face inversion alters the local configural information about the features, McGregor et al. further concluded that participants haptically processed only *local* configural information about the 2D features, the features themselves being treated as oriented objects within a body-centered frame of reference.

Given that recent haptic research provides ample evidence that faces can be discriminated and identified both visually and haptically, the question about the nature of the information underlying haptic face recognition and whether or not this information is the same as in vision arises. If this is the case, and if haptic information contributes to visual information to achieve robust face recognition performance, face information should be easily shared across modalities. Kilgour and Lederman (2002) first addressed the question of cross-modal transfer in both directions-vision as the input modality with haptic matching and haptics as the input modality with visual matching- and found that, for the most part, cross-modal transfer of face information between vision and touch occurred. In addition, and since vision was involved in both conditions, they investigated whether participants adopted a visual mediation strategy to identify the faces. As any performance similarities between the two modalities may be attributed to the transformation of haptic input into a visual image that is subsequently re-processed by visual mechanisms and/or to the modalities sharing common supra-modal processes. However, they could not find any evidence to support this hypothesis. That is, participants did not appear to include a visual-translation stage when performing the face matching task haptically but rather relied on the basic processing mechanisms associated with the sense of touch.

Casey and Newell (2005) tested whether or not familiarity promotes better recognition across the sensory modalities. They found that cross-modal matching for newly familiar faces was better than for unfamiliar faces, regardless of whether face stimuli were matched from vision to touch or touch to vision. However, although they tried to ensure that the encoding of the face stimuli was equivalent across modalities by allowing 1 min for haptic encoding and only 1 s for visual encoding, they found better performance for visual-haptic transfer than for haptic-visual transfer. These results were taken to indicate that even short-term familiarity can produce a more multisensory representation of the face, at least one that can be shared across modalities.

The work by Kilgour and Lederman (2002) and Casey and Newell (2005) on cross-modal transfer in face recognition was extended in a recent series of experiments conducted by Casey and Newell (2007). First, they examined recognition performance for a set of unfamiliar faces within and across the visual and haptic modalities, using live-size plaster face models as stimuli in the haptic modality and 2D color images in the visual modality. They found a decrease in cross-modal face recognition performance relative to unimodal recognition of faces. This cost was incurred independently of the learning modality, suggesting that face recognition is not underpinned by a single multisensory modality. While these results seemingly contradict those found by Casey and Newell (2005) (see above), this evidence of bi-directional cross-modal transfer (whether partial or complete) confirms that vision and touch processes have access to at least some common structural representations. However, to the extent that transfer is incomplete, the two modalities may well represent different aspects of the object in light of the material-geometry distinction above, that is, a relatively stronger emphasis on structure for vision and material for touch. Secondly, Casey and Newell (2007) investigated whether differences in the manner in which faces are encoded affect the representation of faces in memory. Whereas 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 is limited to a serial exploration of an object requiring integration of object information over time (Loomis and Lederman. (1986); Loomis et al. (1991)). Therefore Casey and Newell (2007) made encoding in vision and haptics more similar by presenting part-scrambled face images during visual learning, to promote serial encoding. However, as they again found a decrease in cross-modal relative to within-modal face recognition, their results suggested that it is not encoding differences that account for the observed cost in cross-modal transfer but rather the nature of information represented by each modality. Hence, they raised the question whether featural or configural information is shared best across modalities (again, featural information is defined as isolated face parts, whereas configural information is defined as the layout of the face parts). They found that configural information is processed by the haptic system and that it is this type of information that enables more efficient face matching across modalities. In summary, this pattern of results was taken to indicate 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 promotes more robust within-modal relative to cross-modal recognition.

As discussed above, another important topic in visual face processing has been the neuroanatomical location that underlies visual face processing. Much of the research (e.g., McCarthy et al. (1997)) points to the fusiform gyri (particularly in the right hemisphere) as the significant neural substrate. Likewise, attention has been drawn to the underlying neural correlates of haptic face recognition and whether haptic and visual face processing share common neural substrates.

Kilgour et al. (2005) conducted an fMRI study in which participants were intensively trained to identify specific face masks that were molded from live faces and specific control objects. When these stimuli were presented in the scanner, face masks activated left fusiform and right hippocampal/ parahippocampal areas (and other regions) more than control objects, whereas the latter produced no activity greater than the face masks. Moreover, they found a greater activation of the left fusiform gyrus for face masks than for control objects. This pattern of results was taken to indicate that these ventral occipital and temporal areas may play an important role in the haptic identification of faces at the subordinate level and that the left fusiform gyrus may be recruited more for face masks than for control objects because of the increased need for sequential processing by the haptic system.

In addition, Kilgour et al. (2004) used the case of a prosopagnosia patient to infer areas involved in haptic face recognition from his pattern of brain damage in combination with his behavioral results. They, therefore, suggested an influence of the fusiform gyrus, other extrastriatal visual regions (Deibert et al. (1999); James et al. (2002); Ostry and Romo (2001)), including areas in the parietal lobes (supramarginal and angular gyri, Deibert et al. (1999)) and the lateral occipital complex (LOC) and the claustrum, which has been shown to be involved in haptic-visual cross-modal transfer (Hadjikhani and Roland (1998)).

James et al. (2006)) specifically investigated the influence of familiarity on haptic face identification. Subjects were carefully trained to identify a subset of 3D plaster face masks ('familiar' subset) using their left hand only. In the scanner, they were then haptically presented with old and new objects to judge for familiarity. The left fusiform gyrus was activated more strongly by the haptic presentation of familiar (cf. unfamiliar) objects, suggesting that this area specifically differentiates haptically familiar and unfamiliar faces.

Two additional fMRI studies offer complementary evidence for the suggestion that there is some overlap between vision and touch in their neural representations of facial identity, but that at least some information is preserved in separate, modality-specific channels. Pietrini et al. (2004) examined whether the information about faces that is represented in the ventral visual pathway by distinct patterns of neural activity is strictly visual or a more abstract, supramodal presentation of object form. Specifically, they tested whether the patterns of response elicited by haptic recognition of faces are distinct and to what extent these differential patterns of response during tactile recognition are similar to the patterns of response elicited by visual recognition of faces.

They studied haptic recognition in both sighted and blind individuals. By studying congenitally blind individuals who have no visual memories for faces, they could rule out that the responses evoked by tactile recognition in visual areas are merely the result of visual imagery. Furthermore, they found that the activity evoked by haptic recognition of face masks in sighted participants was unrelated to the pattern of response during visual face recognition.

In a related study, Kitada et al. (2009) examined brain organization for haptic and visual identification of human body parts (faces, hands and feet) vs. non-biological category of control objects (bottles). In accord with Pietrini et al. (2004), haptic and visual object identification activated largely disjoint networks. However, it is possible that face sensitivity may be shared across sensory modalities in small regions, of which locations are spatially varied across participants. The authors examined two regions which produced the strongest activation in haptic and visual face identification. These two discrete areas, HFR ('haptic face region') and FFA ('fusiform face area') were sensitive to 3D face masks (cf. controls) whether presented haptically or visually. Nevertheless, the corresponding activation patterns across object categories (faces, feet, hands, and bottles) were different for FFA and HFR regions. Kitada et al. concluded that although both regions within the fusiform gyrus are sensitive to faces, independent of sensory modality, the sub-region that is most sensitive to haptically presented faces (HFR) is functionally distinct from that which is most sensitive to visually presented faces. Finally, they addressed the use of visual imaging vs. multisensory processing of faces and other body parts by including a third condition in which participants were required to visually image targeted exemplars of face masks (as well as other body parts).

Several measures of visual imagery were obtained, involving both behavioral (i.e., VVIQ: Marks (1973); subjective reports regarding the extent to which participants used visual imagery) and neuroimaging measures (i.e., neural activation in visual imagery vs. haptic conditions). Various correlational analyses converged in showing that at best, visual mediation could account for only a relatively minor portion of the increase in category-specific signal observed with haptically presented faces (and other body parts). The authors concluded that visual imagery is not necessary to achieve good haptic perception of facial identity (or other body parts).

In short, previous research on haptic face recognition has shown that (1) humans can haptically discriminate and identify faces at levels well above chance (Kilgour and Lederman (2002); Casey and Newell (2007)), (2) haptic and visual processing of facial identity are similarly influenced by orientation (Kilgour and Lederman (2006); Kilgour et al. (2004)), (3) information can be shared across the haptic and visual modalities bi-directionally (Kilgour and Lederman (2002); Casey and Newell (2007)) to a certain extent and (4) visual and haptic face recognition may be subserved by discrete neural mechanisms (Pietrini et al. (2004); Kitada et al. (2009)).

2.4 Perceptual expertise

There are a number of behavioral and neural characteristics that distinguish novices and experts. It goes without saying that experts know more than novices about their domain of expertise. They can verbalize more properties, describe more relationships, make more inferences, and so forth (e.g., Ericsson et al. (2006); Kim and Ahn (2002); Murphy and Wright (1984); Johnson and

Mervis (1997)). This is, after all what makes them experts. Our focus here is on behavioral and neural changes in visual cognition that underlie perceptual expertise. In this section, we will briefly review the literature on real-world and laboratory-trained perceptual expertise in terms of relevance for the work conducted in this thesis.

2.4.1 Real-world expertise

Real-world perceptual expertise represents the endpoint on the continuum of perceptual category learning and has been investigated across a variety of domains such as face recognition (e.g., Tanaka (2001); Dufour et al. (2006); O'Toole et al. (1994)), dog show judges (e.g., Diamond and Carey (1986); Robbins and McKone (2006); Tanaka and Curran (2001)), bird watchers (e.g., Gauthier et al. (2000a); Johnson and Mervis (1997); Palmeri and Blalock (2000); Xu (2005)), car experts (e.g., Gauthier et al. (2000a); Grill-Spector et al. (2004); Rossion et al. (2007)), chess players (Chase and Simon (1973)), fingerprint examiners (Busey and Vanderkolk (2005)), radiologists (e.g., Myles-Worsley et al. (1988); Nodine and Krupinski (1998); Nodine et al. (1999)), expert fisherman (Boster and Johnson (1989)), and tree experts (Lynch et al. (2000); Proffitt et al. (2000)), based on which the following core features of perceptual expertise have been proposed (Palmeri and Cottrell (2010)):

- *The use of implicit versus explicit knowledge.* While novices often rely on explicitly verbalized category knowledge in the form of rules or ideal cases that are acquired from reference manuals or explicit instruction (e.g., Allen and Brooks (1991)) or that are created through induction (e.g., Johansen and Palmeri (2002)), expert categorization often seems removed from explicit and conscious deliberation (e.g., Brooks et al. (1991); Sloman (1996)) despite the fact that experts have more verbal knowledge about a domain.

- *An advantage in terms of speed and accuracy.* Experts show a marked improvement in terms of speed and accuracy for recognition compared to novices (Newell and Rosenbloom (1981); Heathcote et al. (2000); Rickard (1997); Palmeri (1999)). One important aspect of this speed up is the so-called *entry-level shift* (Jolicoeur et al. (1984); Tanaka and Taylor (1991)). For novices, categorizations at the basic level ('dog' or 'bird') are faster than categorizations at either the superordinate ('animal' or 'plant') or the subordinate level ('robin' or 'terrier'). The fastest level of categorization is often described as the entry-level into conceptual knowledge. For experts, there is an entry-level shift whereby subordinate-level categorizations are made as quickly as basic-level categorizations. Similarly, expertise leads to an increase in the ability to attend to more fine-grained perceptual features, typically associated with subordinate-level processing (Johnson and Mervis (1997); Tanaka and Taylor (1991)).

- *Derivation of configural information.* In addition to increases in the ability to process objects of expertise at the subordinate level, there is also a trend for greater holistic and configural processing of objects of expertise. Experts process not just the individual features but also the relations among them (Gauthier and Tarr (2002); Maurer et al. (2002)). Experts but not novices show poorer accuracy and longer reaction times when stimuli are presented upside down than when they are upright. This inversion effect is well established in the domain of face recognition (Yin (1969)) and is also observed for dog (Diamond and Carey (1986)) and Greeble (Gauthier et al. (1998)) experts. Similarly, Rossion and Curran (2010) found that groups of car experts showed car inversion effects, but the magnitude of the inversion effect correlated with degree of expertise within the expert group indicating that the shift to configural processing occurs gradually during expertise acquisition. Moreover, experts process individual features of a stimulus more poorly than novices, the so-called composite effect; when two parts of two different stimuli are presented as a composite, experts are slower and less accurate in recognizing one of the parts when the composite is upright or fused, compared with when the composite is inverted or the two parts are not fused (Hole (1994); Young et al. (1987)). Both manipulations are thought to disrupt holistic or configural processing, thus affecting experts but not novices. In fact, experts are highly sensitive to changes in the configuration of features, but only when objects are presented in a familiar orientation (Maurer et al. (2002); Mondloch et al. (2002); Gauthier and Tarr (1997)). The claim that sensitivity to configural information mediates expertise is supported by studies showing that children process faces in a part-based manner, whereas adults (as well as experts in their domain of expertise) process information in a more holistic fashion (see 'visual face recognition', Carey and Diamond (1994); Diamond and Carey (1986)). An extensive amount of research, which to discuss is outside the scope of this review, has been dedicated to investigating how the highly specialized skill of face recognition emerges during infancy, continues to develop throughout childhood, and becomes adult-like in late adolescence (e.g., Carey and Diamond (1977); Hay and Cox (2000); Maurer et al. (2002); Mondloch et al. (2002, 2003); Pellicano and Rhodes (2003); Schwarzer (2000)). In contrast, Robbins and McKone (2006) found no evidence of facelike holistic processing in dog experts viewing images of dogs as measured by the inversion task, the composite paradigm and sensitivity to contrast reversal. Thus, while most of the research suggests that real-world experts process objects of expertise holistically, other studies indicate that there may be multiple pathways to perceptual expertise involving both holistic and featural analysis.

- *Lack of verbal interference.* Novices and experts show different patterns of interference. Novices are easily distracted whereas experts may be able to simultaneously engage in other tasks while making expert decisions. Part of this apparent lack of interference may be because experts no longer use explicit verbalizable routines, so concurrent verbal activity does not interfere with performance. But when experts engage in tasks that tap the same representational resources used for other domains of expertise, they suffer interference in ways unseen in novices (Gauthier and Curby (2005); Gauthier et al. (2003); Rossion et al. (2004); see also Curby and Rossion (2010)). Gauthier et al. (2003) studied interference related to holistic processing of cars and faces. Behavioral results suggested that car experts process upright cars more holistically than novices, and this interfered with holistic processing of faces. The presence of interference between faces and car expertise suggests shared neural mechanisms for processing these stimuli (Rossion et al. (2007)).

- *Generalization of knowledge.* Experts have the ability to rapidly learn and correctly recognize new exemplars more quickly and accurately than novices. That is, expertise allows generalization to previously unknown members of an expert object class, at least so long as new objects are similar to other objects in their domain of expertise (i.e., they vary systematically in the same way as other learned objects; Gauthier and Tarr (1997, 2002); Tanaka et al. (2005)). This restriction is nicely demonstrated by the other-species effect (superior recognition of members of one's own species, e.g., Dufour et al. (2006); Pascalis and Bachevalier (1998); Pascalis et al. (2002); Scott et al. (2005, 2006a)) and the other race effect (superior recognition of members of one's own race, e.g., Chance et al. (1982); Meissner and Brigham (2001); O'Toole et al. (1994)) in face recognition. Both effects reflect differential experience with own- and other race/species faces and suggest that perceptual exposure plays an important role in shaping face perception abilities.

- *Neural markers of expertise.* Experts show different patterns of brain activity than novices. For example, recent investigations into the neural correlates of perceptual expertise have shown that the fusiform face area (FFA) is not just involved in face recognition but is activated by objects of expertise in real-world experts for cars and birds (Gauthier et al. (2000a); Xu (2005); but see Grill-Spector et al. (2004)) and by objects of expertise created in the lab (Gauthier and Tarr (1997, 2002)). Similarly, event-related potential (ERP) markers for face recognition such as the N170 exist, which shows highest amplitude when observing objects of visual expertise over objects that are not (Tanaka and Curran (2001); but see Scott et al. (2006b)).

2.4.2 Laboratory-trained expertise

One way to mimic the acquisition of a natural expert system like face processing is to train participants to better discriminate different classes of objects. While it is not expected to equate laboratory-trained expertise to real-world expertise, as the latter occurs on the scale of years (e.g., Carey and Diamond (1977); Maurer et al. (2002); Mondloch et al. (2003)), whereas typical laboratory training studies require only hours of training (e.g., Gauthier and Tarr (1997, 2002); Malpass et al. (1973)), training studies allow for the manipulation of different factors that may contribute to the acquisition of expertise, providing better control over variables influencing this process, such as level of categorization, supervised versus unsupervised training, and stimulus type.

The use of training studies to examine the acquisition of perceptual expertise originated with training participants with novel objects called Greebles (Gauthier et al. (1999a); Gauthier and Tarr (1997, 2002); Gauthier et al. (1999b, 1998); Scott et al. (2006b, 2008)). The first of these investigations found that training not only led to faster and more accurate responses, but training also increased the configural (and thus face-like) processing of Greebles (Gauthier and Tarr (1997)), evidenced by increased reaction time to trained Greeble configurations (studied parts, in a studied configuration) compared to transformed Greeble configurations (studied parts, in a different configuration). The development of configural processing occurred gradually over the course of training, becoming evident first for features that were close to one another, and later for more distal features, suggesting a widening window of spatial attention. Tests of generalization of learning after Greeble training suggested that learning generalized to Greebles that were structurally similar to the training set, but did not generalize to Greebles that were less similar to the training set (Gauthier et al. (1998)).

Tanaka et al. (2005) applied the Greeble training protocol to teach bird expertise. Participants learned to classify species of wading birds and species of owls at either the subordinate or basic level of abstraction. Results indicate that subordinate, but not basic-level training (i) increased discrimination on previously trained wading birds and owls, (ii) resulted in an 'entry-level' shift, i.e., similar category verification times for basic and subordinate level judgments, and (iii) led to greater generalization to novel exemplars within the trained species and to novel (untrained) species within that family (i.e., owls or wading birds). These results suggest that learning to individuate at the subordinate but not the basic level increases discrimination in experts.

Neuroimaging studies have more closely linked behavioral changes due to training with corresponding changes in brain activity as measured by fMRI and ERP (event-related potential) methods (Gauthier et al. (1999b, 2000b); Rossion et al. (2002); Scott et al. (2006b, 2008); Tarr and Gauthier (2000)). Using fMRI, two areas in the right hemisphere, the occipital face area and fusiform face area, previously associated with face processing have also been found to increase in activation after Greeble training (Gauthier et al. (1999b, 2000b); Tarr and Gauthier (2000)). More specifically, these areas are both recruited when Greeble novices become Greeble experts. Moreover, recent results of training studies have further clarified the importance of individuation training as well as unsupervised exposure in the formation of expert perceptual abilities. For example, studies looking at electrophysiological and behavioral changes over time (Scott et al. (2006b, 2008)), have refined our understanding of the N170 component (previously found to index face processing) and suggest that increases in the N170 are due to increased exposure to object categories (such as faces or objects of expertise).

It is critical to note, however, that these findings pertain only to visual expertise, and, in fact, little is known about expertise in modalities other than vision.

James and colleagues previously trained participants for 10 to 12 hours in order to be able to haptically identify faces and Greebles with 100% accuracy (James et al. (2006)). Unfortunately, their study only assessed the influence of familiarity on brain activation during haptic exploration of 3D face masks and did not test for hallmarks of face expertise for haptic face processing.

Behrmann and Ewell (2003) trained participants- for at most 2 hours- on haptic identification of two-dimensional patterns. Since they found an inversion and a part-whole effect (better recognition of the whole than a part pattern) for tactile pattern recognition for experts but not novices, indicating sensitivity to configural processing in experts, they suggested that expertise manifests itself in a qualitatively similar fashion in vision and haptics. Crucially, however, the patterns were designed to be encoded at a 'haptic glance' (Klatzky and Lederman (1995)) and did not require serial exploration. The authors, therefore, manipulated an important characteristic of haptic information encoding, rendering it more similar to visual information encoding. Further studies are, therefore, necessary to investigate how expertise manifests itself in the haptic modality and whether it is really in a qualitatively similar fashion as vision.

Finally, a recent brain imaging study by Saito et al. (2007) has shown the activation of the primary visual cortex in a tactile discrimination task performed by sighted individuals. Critically, however, this activation of visual areas was found only for participants who had been trained in tactile discrimination (e.g., players of mahjongg who were experts in discriminating the tiles that are used to play the game). Interestingly, the activation in the primary visual cortex in the expert group was also reported (although reduced in magnitude) when Braille characters were presented, thus showing that the effect extended beyond the specific category of expertise (mahjongg cards) of the participants and generalized to an untrained category (Braille cards). These results were taken to indicate that the involvement of visual areas when participants perform tactile tasks (a result previously observed only in blind individuals; e.g., Sadato et al. (1998)), might be, at least in part, related to the strengthening of cross-modal connections/ representations as a function of intensive practice. That is, people who are experienced with a certain class of visual stimuli may learn to associate those stimuli with their tactile equivalent. This would also result in an activation of visual areas when the stimulus is presented haptically (cf. Murray et al. (2004, 2005)).

Overall, training studies provide a powerful tool for elucidating the mechanisms involved in the acquisition of expertise, and the use of these studies for further understanding what makes an expert 'expert' is invaluable.

Chapter 3

General Methods

In this section, we describe in detail the creation of the haptic stimuli used in the experiments in this thesis.

3.1 Stimulus database

We used the morphable MPI-Face-Database that contains images of 7 views of 200 (100 female and 100 male) laser-scanned (Cyberware^{TM} 3030PS) heads without hair. The 200 head models were newly synthesized by morphing real scans to avoid close resemblances to individuals who may not want to appear on a computer screen or in scientific publications. We picked faces that did not contain any artifacts from scanning and were perceptually most distinctive.

3.2 Scanning method

Scanning was done using the Cyberware Head & Face 3D Color Scanner 3030. Prior to being scanned, each participant was required to wear a swimming cap to cover the hair. The scanner created a profile of the face by shining a low-intensity laser beam in the shape of a vertical stripe onto the head. The laser beam moved around the head in 15 s while a video sensor captured the profile at a rate of about 30 times per second, sampling the shape of the head on a regular cylindrical grid of 512 x 512 points with a resolution of 0.8° horizontally and 0.615° vertically. Simultaneously, a second video sensor acquired color information in the same spatial resolution. Consequently each scan is represented by two sets of data of 512 x 512 points each. One set describes the 3D shape of the head (geometric data); the other one the RGB-color values of each pixel of the head image (textural data). Face images were subsequently cropped at the hairline such that the resulting faces were devoid of hair or scalp, but included the ears and ended at the neck. After processing, each face was represented by approximately 7 x 10^4 vertices and the same number of color values.

3.3 Printing preparation

We created one set of life-size and a second set of smaller-than-life masks of the same faces (Printing preparation is described for smaller-than-life faces. If parameters differed for the two sets additional values for life-size (LS) faces are given in [brackets]). All stimuli from the face database had to be edited using the graphics package 3D Studio Max (Autodesk). The imported face models were all arranged at (0, 0, 0) in absolute world coordinates. First, to smooth the borders of the face masks, we selected them using soft selection with a fallout of 1.0. The so selected border was then smoothed applying a mesh relaxation operator (Object-space modifier relax) with a relax value of 1.0

and five iterations. The boundary points were not kept fixed nor were the outer corners saved. Secondly, to establish all faces at the same height above the pedestal and to ensure sufficient space for chin exploration we elongated the neck of all face masks. For this, we selected only the lower border using soft selection again, with a falloff of 1.75. The border was manually extruded and cut applying a slice operator, arranging the slice plane at (0, 0, -13) [life-size faces :(0, 0, -10)] in absolute world coordinates, to remove the bottom. Thirdly, the scanned-faces were thickened to ensure stability of the printed objects, using Claytools Software. Faces (scaled in cm) were imported using an edge sharpness parameter value of 0.04 cm and a thickness of 0.6 cm [0.3 cm]. To avoid contour lines that are created by the conversion of volumetric data back to a triangle mesh, the data was reduced from around 2000000 triangles to about 300000 triangles using the 'Reduce for Export' function [Models were reduced for export to 25% of their original size]. The exported obj-files were smoothed in 3DS in the nasal area using soft selection with a falloff value of 1, and again applying the relax operator with a relax value of 1.0 and three iterations, keeping boundary points fixed. Furthermore, we deleted all noise shells that were created by Claytools during the export by selecting the main (largest parts of the) face and then deleting the inverse selection set.

Finally, we created a pedestal using a hexagon with a radius of 5, a fillet of 2.5 and a height of 3 [1]. The hexagon was scaled by factor 200 along the x- and y-dimensions and arranged with an offset of 2.5 along the y -axis and of -14 [-10] along the z- axis. [Pedestals for life-size faces were hollow with a shell thickness of 4mm and rotated by 90° along the z-axis compared to small faces.] The face mask and the pedestal were unified applying a boolean operation on the two objects. The platform was created by subtracting a hexagon (radius 5.05, fillet 2.5, height 3[1], scaled by factor 200) from a box with a length and width of 11, a height of 4 and scaled by the factor 200, applying a boolean operation on the two objects.

Furthermore, we created a second platform for presenting the faces upside-down. This was done by subtracting 4 sides of a hexagon (radius 5.1, fillet 2.5, height 3 [1], scaled by factor 200) from a box with a length of 30 and width of 20, a height of 4.75 [2.75] and scaled by the factor 200. The support plate was reshaped by subtracting two boxes with a length of 10, a width of 3.5 and a height of 2 from the base of the sunken hexagon with a distance of 0.808 to both sides of the structure. The remaining 'tongue' was reshaped by subtracting a box with a length of 2.6, a width of 6.25, and a height of 2, from which a sphere with a radius of 3.125 and 32 segments had been subtracted beforehand. All operations were performed by applying a boolean operation on the respective objects.

Finally, we used the printing software Magics for Objet to prepare the printing platform and connect to the 3D printer. All objects were rescaled by factor 5 [10], oriented upright [lying on their nose] and positioned such that six [one] face masks fit on the platform. Printing of a set of six face masks took about 72 hours.

3.4 Printing

Printing was performed by an Eden 250 printer. PolyJet inkjet technology works by jetting white, acrylic-based photopolymer materials in ultra-thin layers (16μ) onto a build tray layer by layer until the part is completed. Each photopolymer layer is cured by UV light immediately after it is jetted, producing fully cured models that can be handled and used immediately, without post-curing. The gel-like support material, which is specially designed to support complicated geometries, is easily removed by hand and water jetting.

3.5 Stimuli

Smaller-than-life face masks weighed about 138 ± 5 g each and measured 89 ± 5.5 cm wide, 120 ± 7.5 cm high and 103.5 ± 5.5 cm deep. Life-size face masks weighed about 422 ± 20 g each and measured 147 ± 13 mm wide, 202 ± 12 mm high, and 190 ± 15 mm deep.

Chapter 4

Cross-modal transfer in visual and haptic face recognition

Abstract- We report four psychophysical experiments investigating cross-modal transfer in visual and haptic face recognition. We found surprisingly good haptic performance and cross-modal transfer for both modalities. Interestingly, transfer was asymmetric depending on which modality was learned first. These findings are discussed in relation to haptic object processing and face processing.

4.1 Introduction

The visual information provided by human faces is of strong ecological significance, for example, for communication, identification and mate-selection. Strong psychophysical (e.g., Gauthier et al. (2003); Maurer et al. (2002)), neurophysiological (e.g., Tsao et al. (2006)) and neuroimaging (e.g., McKone et al. (2007)) evidence suggests that faces are visually processed with an expertise that surpasses general object recognition. Not surprisingly, almost all research that supports this perspective has focused on vision. In terms of general object recognition, however, previous research suggests that information from *multiple sources of sensory inputs* contributes to object representations (Easton et al. (1997); Newell et al. (2001); Reales and Ballesteros (1999)). Such multisensory object representations allow for more robust recognition performance (Ernst and Bülthoff (2004)), i.e. no single sensory signal can provide reliable information about the three-dimensional structure of the environment in all circumstances. This incompleteness might be resolved by combining information from different sources that complement each other, thus rendering recognition less sensitive, for example, to ambiguous cues. In this context, little is known about how face information is integrated across modalities and whether information from other sensory modalities can contribute to the formation of robust face representations. Functional models of visual face recognition (e.g., Bruce and Young (1986)) suggest that *structural* information from a face (a feature-based description together with a representation of the spatial arrangement or configuration of those features) is encoded and represented in face memory for person identification. Since touch can also encode structural information (e.g., Lederman and Klatzky (1990)), this information could, in principle, also contribute to face memory. As the structure of a face remains unchanged whether it is presented to vision or to active touch (Reales and Ballesteros (1999)), a single, abstract representation (a similar structural description) would be created after the face is perceived. Were this true, we would expect to obtain substantial and symmetric (independent of direction of transfer) cross-modal transfer, inferring that the face representations are common to vision and touch, and that these representations are primarily structural. Cross-modal transfer, however, should disappear when changes across modalities interfere with the structural descriptions of the faces. Therefore, although haptic face recognition seems like an unusual task at first inasmuch as we have little to no training in haptic face recognition throughout life, we suggest that if people are good at it (and, especially, at an unexpected crossmodal transfer task) then this provides evidence for efficient transfer of shape information across the two modalities.

Recently, a few studies have started to investigate whether faces can be recognized haptically and if so, what commonalities haptic recognition might share with visual face recognition. Kilgour and Lederman (2002) for the first time demonstrated participants' capability to identify unfamiliar live human faces and face masks using only their sense of touch, showing that face information might be shareable across the senses. Since then, other studies have confirmed this result using 3D face masks (Casey and Newell (2005, 2007); Kilgour and Lederman (2006); Kilgour et al. (2004, 2005); Pietrini et al. (2004)).

Given that recent haptic research provides ample evidence that faces can be discriminated and identified both visually and haptically, the question arises whether the nature of the information that underlies haptic face recognition is the same as that in visual face recognition. If this is the case, face information should be easily and symmetrically shared across modalities. Two recent studies (Kilgour and Lederman (2002); Casey and Newell (2005)) both showed that unfamiliar faces can be successfully matched across modalities. Perceptual matching, however, may rely on information processing that is not specific to face perception *per se.* In a follow-up study (Casey and Newell (2007)), an old/new recognition task was used in which haptic memory was aided by reducing the number of haptic learning stimuli. Here cross-modal face recognition was worse than within-modal recognition. This cost was incurred independently of the learning modality, suggesting that face recognition is *not* underpinned by a single multisensory representation. One major shortcoming of this study was, however, the use of *different* stimuli in the two modalities: haptic stimuli were life-size plaster faces while the visual stimuli were 2D, color images of the same individuals. The experiment therefore deprived haptic face recognition of important information such as texture while providing additional information (e.g., colour) for visual face recognition. As objects and pictures did not contain the exact same information, recognition of one from the other would require comparison of information at a higher, supposedly more abstract level. The involvement of abstract information transfer could reduce the importance of other, more perceptual, cues and thus prevent controlled measurement of a transfer effect. In contrast, we used the *same* 3D face masks for experiments in both modalities allowing us to investigate cross-modal transfer at a lower, more perceptual level.

Here we present for the first time a fully controlled stimulus set that enables us to investigate cross-modal information transfer in visual and haptic face recognition at a lower, more perceptual level. More specifically, we will address two important questions: can we generalize from haptically learned faces to the visual domain and vice versa? And if so: at what level is information shared, e.g. is this cross-modal transfer symmetric? If we find symmetric cross-modal transfer, this will provide evidence in favor of shared representations and processes between visual and haptic face recognition. Moreover, by using an old/new recognition paradigm with *identical design* in both modalities we directly address the important question of visual versus haptic memory effects in face recognition.

4.2 General methods

Three-dimensional (3D) models of nineteen faces were taken from the MPI-Face-Database Troje and Bülthoff (1996) and edited for printing using the graphics package 3D Studio Max (Autodesk). Two sets of 3D face masks were printed with the use of an Eden 250 printer (Objet Geometries Ltd.). The first set consisted of life-size face masks that weighed about 422 ± 20g each and measured 147 ± 13 mm wide, 202 ± 12 mm high and 190 ± 15 mm deep. The second set consisted of small face masks that weighed about 138 ± 5g each and measured 89 ± 5.5 mm wide, 120 ± 7.5 mm high and 103.5 ± 5.5 mm deep. See *Figure 4.1 (A)* for an example of the stimuli used.

The apparatus used for visual and haptic face recognition is shown in *Figure 4.1 (B)*. The experimenter placed the face on a mount placed behind an opaque curtain such that participants could not see the face masks during haptic exploration. All faces were rigidly fixed to the platform and always presented from a frontal view. Participants used a chin rest that was placed 30 cm away from the stand on which the objects were presented. The curtain could be slid back to reveal the face masks for visual face recognition. During haptic exploration of the faces, an arm-rest was provided to prevent exhaustion.

Each experiment was performed by a different set of naive participants, who were paid 8 Euros an hour. All participants reported right-handedness, normal tactile sensation and normal or corrected-to-normal vision.

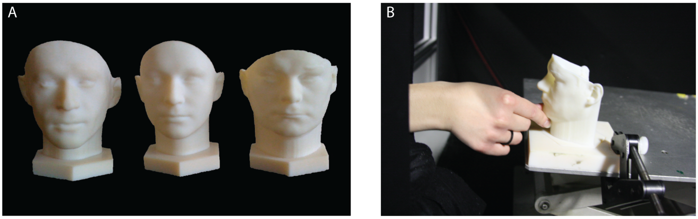

Figure 4.1: (A)Three example stimuli used for haptic face recognition. (B) Experimental setup for haptic face recognition.

4.3 Experiment 1

The aim of the first study was two-fold: (1) to show that haptic face discrimination is possible using our 3D face masks and (2) to investigate the effect of stimulus size on haptic face perception. The latter became necessary as we set out to perform our experiments on cross-modal face recognition with smaller-than-life face masks due to technical constraints. However, decreasing the size of a pattern might affect recognition performance. We therefore compared discrimination performance for two stimulus sizes: life-size and smaller-than-life faces.

4.3.1 Methods

The stimulus set included 2 sets of 12 faces each; one set with life-size face masks and the second set with small face masks. The faces differed in size only.

Twenty-four participants performed a same/different face discrimination task in one of two conditions, i.e. with either life-size or small faces. They were sequentially presented with pairs of face masks which they were asked to explore haptically, using their own exploratory procedure. Face masks were always presented frontally and shown one at a time for 7 sec with an interstimulus interval (ISI) of 5 sec in which the faces were exchanged by the experimenter. A tone signaled the beginning and end of exploration time. After the presentation of the second face of each pair, participants were asked to report whether they had been shown the same face twice or two different faces by pressing a 'same' or 'different' labeled key on a keyboard. They were instructed to respond as accurately and quickly as possible.

Five time-unlimited and five time-limited practice trials were given before the experiment, which consisted of 3 blocks of 78 randomized trials (due to time constraints each object was only compared once with itself and once with every other object resulting in 12 + (12*11)/2 = 78 trials). The order of appearance of stimuli was randomized over blocks. No feedback was provided for either practice or experimental trials.

To exclude the use of obvious strategies such as participants always answering same or different due to the asymmetric design, we calculated performance on same and different trials separately. Performance, given in proportion correct ±SEM, was analyzed using one-tailed t-tests for each condition to test whether performance was above chance (50%), and whether performance was significantly better for life-size than small faces.

4.3.2 Results

Haptic face discrimination performance was above chance in each condition (Life size faces: 74.77 ± 2.49% correct on same trials, $t_{11} = 5.83, p < 0.001$, 74.62 ± 1.39% correct on different trials, $t_{11} = 10.35, p < 0.001$, average percent correct 74.64

± 0.97%; Small faces: 80.79 ± 2.04% correct on same trials, $t_{11} = 8.86, p < 0.001$, and 68.73 ± 1.73 % correct on different trials, $t_{11} = 6.33, p < 0.001$, average percent correct 70.58 ± 1.5%). Most importantly, we found *no significant difference* in performance across conditions (same trials: $t_{22} = -1.09, p = 0.29$, different trials: $t_{22} = 1.55, p = 0.13$).

4.3.3 Discussion

Decreasing the size of a pattern might affect recognition performance when cutaneous spatial resolution limits haptic recognition. However, while we showed that participants were able to discriminate our stimuli at levels well above chance, we did not find an overall effect of size on discrimination performance. Performance tended to be slightly better for life-size faces for 3 out of the 12 faces, suggesting that it might be due to some characteristic feature in the respective faces that was enhanced in the larger faces, rather than a general advantage of life-size versus small faces.

4.4 Experiment 2

Having established that our face stimuli are suited for haptic (face) processing, the goal of Experiment 2 was to test cross-modal transfer from the haptic to the visual modality using an old/new recognition task.

4.4.1 Methods

First all 18 participants were haptically familiarized with 3 faces (out of 19 total) that were randomly chosen from 6 sets of 3 faces each. We labeled each face with a short first name. Participants were allowed to explore the faces only haptically using the right hand, with no constraint on either the exploratory procedure or the duration of exploration. They were told to explore the face masks carefully and to learn their names because they would be asked to recognize those particular faces later. No further information was given about the nature of the following experiment during the familiarization. Haptic learning of the three faces took 4 min on average. In the subsequent identification task, participants had to name each randomly presented face mask after haptic exploration. Feedback was provided in that participants were told whether the face was recognized correctly or not. Each face mask had to be identified correctly twice before the experiment continued.

The old/new recognition task immediately followed the familiarization and consisted of 4 blocks of 19 trials corresponding to 3 old (learned) and 16 new faces (each object was shown once per block). This asymmetric design was chosen because of time constraints for haptic learning. Face masks were shown one at a time in random order with an ISI of 10 sec in which the faces were exchanged. In the within-modal blocks 1 to 3, they were asked to explore each face mask *haptically* and to report whether it was one of the three faces they had learned (old) or not (new). Audio signals indicated begin and end of the exploration. As before, participants were free to use their own exploratory strategy to explore the faces. Although exploration time was unrestricted, they were instructed to respond as quickly and accurately as possible by pressing an ”old” or ”new” labeled key on a keyboard with their left hand. Participants took about 10 min to complete a haptic block. In the cross-modal block 4, participants were asked to perform the old/new recognition task *visually.* Participants had not been informed about this cross-modal recognition task before. This was to assess if participants were able to form a visual representation from haptic input. A curtain revealed the faces until the participant responded by pressing the respective key on a keyboard. No feedback was provided in any test trial in either modality.

Responses were converted to standard d' scores and analyzed using one-tailed t-tests for each block to test whether performance was above chance[1]. Paired t-tests were then used to compare performance across within-modal blocks, and to compare Block 3 to Block 4 to assess cross-modal transfer.

4.4.2 Results

Figure 4.2 (A) shows recognition performance for Experiment 2 across participants, for each of the within-modal blocks (H-H) and for the cross-modal block (H-V).

Haptic face recognition performance was significantly above chance in each block (Block 1: $t_{17} = 4.23, p < 0.001$; Block 2: $t_{17} = 4.01, p < 0.001$; Block 3: $t_{17} = 4.45, p < 0.001$), although decreasing significantly across blocks 2 and 3 ($t_{17} = 2.13, p < 0.05$). Cross-modal recognition was not significantly above chance. ($t_{17} = 1.66, p = 0.11$).

Our results demonstrate participants' ability to learn and recognize faces haptically. Nonetheless, overall performance was rather poor with mean $d' < 1$. In contrast to within-modal face recognition, we found cross-modal recognition to be at chance level only, indicating that participants were not able to recognize the three haptically learned faces visually. One interpretation might be that information learned during haptic face explo-

---

[1] Following standard practice we set negative d' numbers to 0, which due to our asymmetric design resulted in a slight increase of the chance level from 0 to roughly 0.5.

ration cannot be accessed for visual face recognition, i.e. a failure to share a representation of the learned faces across modalities. Alternatively, the inability to recognize faces across modalities may be due to quick fading of haptic memory. This interpretation is supported by the decrease in performance across blocks as well as by debriefing reports of participants wherein they mentioned difficulties in recalling the learned faces. Therefore, in the next experiment we investigated whether refreshing haptic memory, prior to each block, would improve recognition performance, both within and across modalities.

4.5 Experiment 3

Experiment 3 tested the hypothesis that haptic memory was the limiting factor for the lack in cross-modal transfer found in Experiment 2.

4.5.1 Methods

18 naive participants took part in Experiment 3. The design was the same as in Experiment 2 except that haptic memory was refreshed by repeated exposure to the three learned faces. That is, the identification task was conducted *before each test block.*

In addition to the statistical comparisons outlined in Experiment 2, we used two-tailed t-tests to compare the data from Experiments 2 and 3 to assess whether refreshing haptic memory improved within-modal and cross-modal recognition.

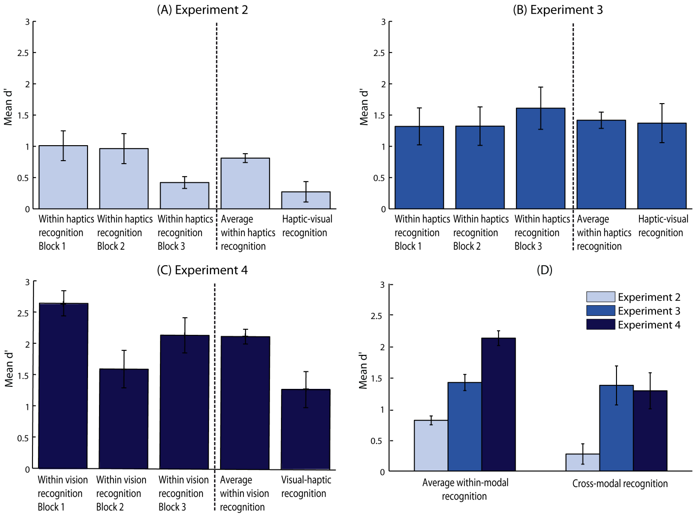

Figure 4.2: Plots showing face recognition performance, measured as the mean d' scores for (A) Experiment 2, (B) Experiment 3, and (C) Experiment 4 for within and cross-modal recognition. Within-modal recognition was in the haptic modality (H-H) in Experiments 2 and 3. Cross-modal transfer was tested in the visual modality (H-V). In Experiment 4, within-modal recognition was tested in the visual modality (V-V) and cross-modal transfer in the haptic domain (V-H). Panel (D) compares recognition performance across experiments. Note, that due to our asymmetric design chance level is actually not 0 but 0.5. Error bars represent ± 1 SEM.

4.5.2 Results

As *Figure 4.2 (B)* shows, haptic (within-modal) face recognition was above chance in each block (Block 1: $t_{17} = 4.46, p < 0.001$; Block 2: $t_{17} = 4.29, p < 0.001$; Block 3: $t_{17} = 4.77, p < 0.001$) with no significant difference in performance across blocks (Block 1 vs 2: $t_{17} = -0.01, p = 0.99$; Block 2 vs. 3: $t_{17} = -1.12, p = 0.28$; Block 1 vs. 3: $t_{17} = -0.70, p = 0.50$). In contrast to Experiment 2, visual (cross-modal) recognition was significantly *above* chance ($t_{17} = 4.39, p < 0.001$). Within-modal recognition was not significantly better than visual recognition ($t_{17} = -0.56, p = 0.59$), so we found no cost for cross-modal transfer. Performance in Experiment 3 was significantly better than in Experiment 2 for both overall within-modal recognition ($t_{34} = -5.53, p < 0.001$) and cross-modal recognition ($t_{34} = -3.05, p < 0.01$) (*Figure 4.2 (D)*).

4.5.3 Discussion

We found that refreshing haptic memory improved both within-modal and cross-modal recognition, with cross-modal recognition now significantly above chance and no worse than within-modal recognition. This pattern of results suggests that the lack of a cross-modal transfer effect in Experiment 2 was not due to a failure to share information of the learned faces across modalities, but rather due to quick fading of haptic memory. If face recognition memory is underpinned by a single multisensory representation, we would expect to find efficient cross-modal transfer for visual as well as haptic learning.

4.6 Experiment 4

This experiment tested whether cross-modal transfer would also happen from vision to haptics.

4.6.1 Methods

The same design as Experiment 2 was used except that we interchanged learning and recognition modalities. Faces were learned visually by 18 naive participants. Within-modality recognition was then tested with visually presented faces whereas cross-modal transfer was tested with haptically presented faces. Visual learning of the three faces took about 2 min, completion of a visual block about 2 min.

In addition to the analyses conducted for Experiment 2, we used two-tailed t-tests to compare Experiments 3 and 4 to assess whether learning modality influenced within-modal and cross-modal recognition performance.

4.6.2 Results

Visual (within-modal) face recognition was significantly above chance in each block (Block 1: $t_{17} = 13.12, p < 0.001$; Block 2: $t_{17} = 5.32, p < 0.001$; Block 3: $t_{17} = 7.60, p < 0.001$) as was haptic (cross-modal) recognition ($t_{17} = 4.45, p < 0.001$). No consistent pattern was found across blocks. Performance decreased significantly across blocks 1 and 2 ($t_{17} = 3.46, p < 0.01$), but not across blocks 1 and 3 ($t_{17} = 1.87, p = 0.08$) (*Figure 4.2 C)*). Unlike Experiment 3, within-modal recognition was significantly better than cross-modal recognition ($t_{17} = -3.13, p < 0.01$). Moreover, within-modal recognition was significantly better in Experiment 4 than Experiment 3 ($t_{34} = -2.34, p < 0.05$), but, interestingly, no significant difference was found for cross-modal recognition ($t_{34} = 0.2, p = 0.84$) (*Figure 4.2 (D)*).

4.6.3 Discussion

Compared to Experiment 3, our results demonstrate a clear advantage for within-modal recognition using vision as the learning modality. Interestingly, we found a cost in cross-modal (V-H), relative to within-modal (V-V) face recognition. This is in contrast to Experiment 3 where cross-modal (H-V) recognition was as good as within-modal (H-H) recognition. Cross-modal recognition was equally accurate from vision to haptics (V-H) and from haptics to vision (H-V). It is unlikely that the cost in cross-modal (V-H), relative to within-modal (V-V) face recognition is due to fading visual memory since recognition performance did not decline across within-modal blocks. We propose that haptic processing performance might constitute a limiting factor for recognition performance within and across modalities. Information from visually and haptically learned faces thus might be equally accessible to the haptic modality. Haptic performance, therefore, does not seem to benefit from the information from visually learned faces that is advantageuos for within-modal visual face recognition.

4.7 General Discussion

The aim of our study was to investigate whether visual and haptic modalities encode similar information about faces to allow for efficient cross-modal transfer. In summary, we first showed that stimulus size does not significantly affect haptic face discrimination performance (Experiment 1). We then replicated previous results (e.g., Kilgour and Lederman (2002)) showing that our 3D face stimuli can be learned and recognized using touch alone (H-H, Experiment 2). Above-chance cross-modal recognition was only possible when haptic memory was refreshed, suggesting a memory effect in cross-modal face recognition (Experiment 3). Here, cross-modal H-V recognition was as accurate as within-modal H-H recognition. In Experiment 4, we found a clear advantage for within-modal V-V recognition. On the basis of previous research (e.g., Jones (1981)), we might predict that the V-H condition would produce higher accuracy than the H-V condition. However, in contrast to H-V transfer, we found a cost in transfer from vision to haptics, resulting in cross-modal recognition accuracy to be equal in both V-H and H-V conditions. This suggests that information transfer across modalities might be asymmetric and limited by haptic face processing. Taken together these results suggest that the way in which information is shared across modalities is different for haptically and visually learned faces.

One reason for our observed asymmetric transfer may be due to qualitative differences in information processing for haptics and vision. In an old/new recognition task, these differences may benefit V-H over H-V performance: earlier studies have shown, for example, that faces are encoded holistically in vision (Tanaka and Sengco (1997); Maurer et al. (2002)) whereas haptic encoding is limited to serial exploration of an object, i.e. it is not holistic but involves a feature-by-feature analysis (Loomis et al. (1991)). As shown by Tanaka and Sengco (1997), it is difficult to identify a feature of a face when it is presented in isolation or outside the context of the whole face. 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 from a holistic, visual representation. Conversely, haptic recognition quite likely does not benefit from the holistic information encoded by vision and therefore be limited by the use of feature-based information.

Another reason for the clear advantage for within-modal V-V recognition compared to equally accurate cross-modal recognition performance, could be the different levels of expertise for visual and haptic face recognition. In fact, expertise in visual face processing takes many years to develop and requires training and experience (LeGrand et al. (2003)). Inasmuch as we have little to no training in haptic face recognition throughout life, differences in the way faces are processed visually and haptically might, at least partly, be attributed to different levels in expertise. Expertise in visual face perception and face information processing may affect a modality encoding bias, resulting in more efficient processing of face information within this modality. In an unexpected cross-modal transfer task, performance could then potentially be limited by haptic processing performance in either direction V-H and H-V equally, resulting in equally accurate cross-modal recognition. As a partial control against such effects we did, however, ensure that faces were learned equally well in both modalities, in that participants had to reach a performance criterion of 100% in an identification task following the learning session.

Moreover, Molander and Garvill (1979) reported two studies in which they tested cross-modal transfer as a function of the amount of first modality training with non-face objects. They found transfer in the haptic-visual order to be consistently superior to transfer in the visual-haptic order. As the participants received more training, they performed better in absolute terms in a subsequent cross-modal transfer task. The relative size of the transfer effect, however, remained quite unaffected, even after overlearning. These findings were taken to indicate that learning factors such as amount of training and level of learning are of minor importance in cross-modal performance and that it would be more fruitful to concentrate on the perceptual aspects of the modalities. We therefore believe that the influence of expertise in this regard plays only a minor role and that the observed differences in performance might rather be due to the nature of information represented by each modality.

Along these lines, Norman et al. (2008) who recently investigated discrimination of 3D object shape through vision and touch reported a similar asymmetry for cross-modal transfer using naturally shaped non-face stimulus objects. More specifically, while they found within-modal visual performance to be superior to within-modal haptic performance, performance for the two cross-modal conditions (V-H and H-V) was not statistically different from that of the within-modal haptic condition. These results demonstrated that effective cross-modal shape comparisons can be made between vision and touch, but that information transfer is incomplete.

In summary, we found cross-modal transfer in face recognition in both directions, H-V and V-H, with a cost in cross-modal (V-H) relative to within-modal (V-V) recognition of a set of faces learned visually, but no such cost for cross-modal (H-V) compared to within-modal (H-H) recognition of the same faces learned haptically. The fact that participants were able to accurately recognize faces in the cross-modal conditions, despite the novelty of the task, suggests that haptic and visual face recognition share commonalities that allow for efficient, 'default' information transfer about faces. Investigating cross-modal transfer in face recognition may therefore be a good tool to study not only the nature of haptic face processes but also the similarities and differences in the corresponding visual face processing conditions, as well as enable a deeper understanding of the mental representation of faces and object recognition in general.

Chapter 5

Serial exploration of faces: Comparing vision and touch

Abstract- Even though we can recognize faces by touch surprisingly well, haptic face recognition performance is still worse than for visual exploration. One possibility for this performance difference might be due to different encoding strategies in the two modalities, namely holistic encoding in vision versus serial encoding in haptics. Here, we tested this hypothesis by promoting serial encoding in vision, using a novel, gaze-restricted display that limited the effective field-of-view in vision to resemble that of haptic exploration. First, we compared haptic with gaze-restricted and unrestricted visual face recognition. Secondly, we used the face-inversion paradigm to assess how encoding differences might affect processing strategies (featural vs. configural). By promoting serial encoding in vision we found equal face

recognition performance in vision and haptics with a clear switch from configural to featural processing, suggesting that performance differences in visual and haptic face recognition are due to modality-specific encoding strategies.

5.1 Introduction

The visual information provided by human faces is of strong ecological significance, for example, for communication, identification, and mate-selection. Consequently, face processing has received a lot of attention in vision research providing evidence for specific processing strategies that evolve with perceptual expertise. Not surprisingly, a considerable amount of vision research has been devoted to investigating the basis of such expertise, and has long been attributed to the use of configural as opposed to featural processing (processing of individual face parts). Three types of such configural processing have been defined (Maurer et al. (2002)): (i) sensitivity to first order relations (recognizing the stimulus as a face), (ii) holistic processing (gluing the features together into a holistic whole or Gestalt), and (iii) sensitivity to second order relations (perceiving inter-feature distances).

Regardless of the specific form of configural processing involved, one highly consistent finding is that face perception is orientation-specific, i.e. faces are processed more accurately when they are presented in the normal upright position than when they are inverted (for reviews, see Searcy and Bartlett (1996); Valentine (1988)). The predominant explanation for this so called 'face-inversion' effect, first demonstrated by Yin (1969), is that vertical inversion selectively impairs our ability to extract configural information from faces, while leaving featural processing largely intact (Leder and Bruce (2000); Schwaninger et al. (2006)). In fact, using gaze-contingent stimulation van Belle et al. (2010) recently showed that the face inversion effect is, indeed, caused by an inability to perceive the individual face as a whole rather than as a collection of specific features, thus supporting the view that observers' expertise at upright face recognition is due to the ability to perceive an individual face holistically.

Interestingly the inversion effect is not present in infants. In fact, a number of studies have shown that this pattern of results (i.e. a remarkable deficit in recognition of inverted faces, but no such deficit for inverted images of non-face objects) takes many years to develop and is, therefore, seen as one of the hallmarks of visual face processing expertise (Carey and Diamond (1977); Dahl et al. (2009); Hay and Cox (2000); Maurer et al. (2002); Mondloch et al. (2003); Pellicano and Rhodes (2003); Schwarzer (2000)).

Face processing, however, is not limited solely to vision. Haptically accessible information about structurally invariant 3D contours accounts for the recent discovery that humans are capable of identifying individual faces at levels well above chance using only their sense of touch; demonstrated for the first time by Kilgour and Lederman (2002). Since then, other studies have confirmed this result using 3D face masks (Casey and Newell (2007); Dopjans et al. (2009); Kilgour et al. (2004); Kilgour and Lederman (2006); Pietrini et al. (2004)), raising the question whether the visual and haptic modality encode similar information and share the hallmarks of face processing, i.e. whether the use of expert face processing strategies, for example, is modality independent.

Few studies have since investigated a haptic face inversion effect with contradictory results. While Kilgour and Lederman (2006), for example, previously used a haptic face inversion paradigm to study orientation-sensitivity of haptic face processing and found a strong inversion effect for faces, Baron (2008) failed to find such a haptic face inversion effect during haptic classification of facial expressions of emotion. Further research is, therefore, necessary to thoroughly understand orientation-sensitivity of haptic face processing.

In a previous study, we provided further evidence that both the haptic and visual system have the capacity to process faces, and that face-relevant information can be shared across sensory modalities, using a fully controlled stimulus set (Dopjans et al. (2009)). Interestingly, we found this information transfer across modalities to be asymmetric and limited by haptic face processing. More specifically, while we observed a clear advantage for within-modal visual recognition in an old/new recognition task, we found cross-modal recognition accuracy to be equal in both vision-to-haptics and haptics-to-vision conditions, due to a cost in transfer from vision to haptics with no such cost in haptics-to-vision transfer. Moreover, we found that haptic face recognition performance was significantly improved when haptic memory was refreshed during the experiment, indicating high memory demands due to the serial encoding process of haptic exploration. We suggested that the observed asymmetric transfer may be due to differences in visual and haptic information processing. While visual face processing has been shown to involve configural processing, haptics might rely more on featural processing. The different processing strategies might in turn be introduced by qualitative differences in information encoding in haptics and vision. Earlier studies have shown, for example, that 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)). While faces are encoded holistically in vision (Maurer et al. (2002)), haptic encoding is limited to serial exploration of an object, i.e., it is not holistic but involves a feature-by-feature analysis (Loomis et al. (1991); Loomis and Lederman. (1986)) due to its narrow effective field of 'view'. Therefore, haptic information of an object has to be integrated over time in order to take in the same amount of information, imposing higher working memory demands on haptic exploration of objects, which might also be a reason why haptic memory capacity appears to be more limited and variable than that of visual working memory (Bliss and Hämäläinen (2005); Dopjans et al. (2009)). To date, however, relatively few studies have attempted to address the characteristics and functioning of people's memory for haptically perceived objects (e.g., see Walk and Pick (1981), for an extensive early review; Knecht et al. (1996); Millar and Al-Attar (2004)) - as compared to the large number of studies that have addressed people's memory for visually presented objects (e.g., Luck and Vogel (1997); Squire (1992); Alvarez and Cavanagh (2004); Vogel et al. (2001); Desimone (1996)). In addition, to our knowledge, no study has been published that has specifically addressed the question of whether and how the way in which haptic information is gathered affects the encoding and storage of that information in the brain (i.e., a question that is related to the serial vs. holistic manner of perceiving haptic stimuli), and consequently how it affects performance in a high-level cognitive task such as face recognition. If either of these factors, serial encoding and/or higher working memory demands, limits performance, as seems likely, then touch must be at a disadvantage in such a task.

Loomis et al. (1991), for example, showed that recognition performance across the visual and haptic sense can be equated in

2D picture recognition by reducing the visual window to the narrowness of the effective field-of-'view' in haptics. Since the near equivalence of tactual picture perception and narrow-field vision suggested that the difficulties of tactual picture recognition must be largely due to the narrowness of the effective field-of-view, the authors concluded that recognition performance when the field of view is restricted is, indeed, impeded by limitations in working memory or in the integration process.

In a recent study, Casey and Newell (2007) addressed the question how encoding differences across the visual and haptic modalities might affect face recognition performance. The authors proposed to make encoding in vision and haptics more similar by limiting visual encoding to a feature-by-feature procedure, i.e. using scrambled faces to enforce serial encoding. More specifically, visual face images were divided into four parts comprising nose area, eye and brow area, mouth and chin area, and external features. Using an old/new recognition task, participants were either presented with whole face or part-based face images (one face part at a time) in the learning session and whole face images during the recognition task. Since they found the same pattern of results for visual and haptic face recognition, independent of the encoding procedure manipulations in the visual modality, the authors suggested that encoding differences did not account for differences in visual and haptic face recognition. One major shortcoming of this study, however, lies in the visual encoding procedure: while using scrambled faces and presenting only one feature at a time certainly enforces perceptual integration, it does not resemble haptic encoding, as it severely alters/restricts participants' exploratory procedures. More importantly, however, Loomis et al. (1991) reported a clear effect of visual field size on performance. Doubling the width of the field from resembling information gained through one fingertip to two already produced a substantial increment in visual recognition performance. Disassembling a face into four parts yields a much larger field size than possible for haptic exploration and might, therefore, not resemble serial encoding in haptics.

Following this line of inquiry, we here present results that directly test the effect of encoding differences on visual and haptic face recognition performance. This was achieved by using a gaze-restricted display used here to constrain the visual system to sequential, self-directed exploration promoting serial encoding in vision in much the same way that the haptic system encodes objects (Loomis (1981)). The gaze-restricted display limited the effective field-of-view in vision to resemble that of haptic exploration using two fingertips. For this, an aperture was moved over a face image by the participants, resembling serial, haptic exploratory procedures. Subjects were given control of the movements of the aperture, such that they could control the information input continuously through time, which is also similar to haptic exploration. The only constraint of this gaze-restricted design was, therefore, that *only one feature*, determined by the observer him/herself, was available at any given time on a face. Thus, the specific aperture-viewing procedure used in these experiments allowed a fair comparison with haptic recognition. In a first series of experiments, we compared haptic, gaze-restricted and unrestricted visual face recognition. Secondly, we tested the effect of serial encoding of faces on working memory demands and, ultimately, recognition performance for haptic and gaze-restricted face recognition. Finally, we used the face-inversion paradigm to assess how encoding differences might affect face processing strategies (featural vs. configural face information processing).

5.2 Experiment 1

We assessed the effect of modality-specific encoding differences on visual and haptic face recognition testing haptic (H), gaze-restricted (GRV) and unrestricted (UV) visual face recognition.

5.2.1 Methods

Subjects

54 experimentally naïve participants (18 per condition) were paid 8 Euros an hour to perform the respective experiment. All participants reported normal or corrected-to-normal vision and had no sensory impairment.

Stimuli

Stimuli for the haptic and unrestricted-visual face recognition experiments consisted of nineteen white plastic face masks. For this, the 3D models of 19 faces were taken from the MPI-Face-Database (Troje and Bülthoff (1996)) and edited for printing using the graphics package 3D Studio Max (Autodesk). Three-dimensional face masks were printed with the use of an Eden 250 printer (Objet Geometries Ltd.) and weighed about 138 ± 5 g each and measured 89 ± 5.5 mm wide, 120 ± 7.5 mm high and 103.5 ± 5.5 mm deep. See *Figure 5.1* for an example of the stimuli used. [1]

---

[1] Due to technical constraints the experiments were conducted with smaller-than-life face masks. We have, however, previously shown that stimulus size does not significantly affect haptic face recognition performance (Dopjans et al. (2009)).

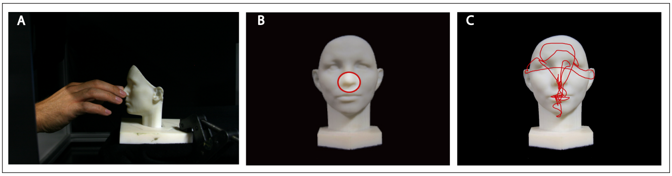

Figure 5.1: ((A) Experimental Setup used for haptic and unrestricted visual face recognition. (B) Demonstration of the gaze-restricted display: The red circle indicates the size of the aperture. Only the part of the image inside the aperture was visible as indicated by the difference in brightness of the images inside and outside of the aperture. The aperture of 2° visual angle was moved over the frontal photograph of the face mask. (C) Example of a recorded trajectory during gaze-restricted face recognition.

Visual stimuli for the gaze-restricted face recognition experiments were generated and presented under Matlab 7.11 using the Psychophysics Toolbox (Brainard (1997); Pelli (1997)). Each gaze-restricted stimulus was created using two images. The first was one of 19 photographs from a frontal view of the white plastic face masks previously used in haptic experiments. The faces spanned 14.7 ± 1.2° visual angle in the vertical plane and 9.1 ± 0.5° visual angle in the horizontal plane and were presented on a black background spanning 25.85° visual angle in the horizontal plane and 19.52° visual angle in the vertical plane. The second was a black image that was superimposed on the photograph. These two images were blended into each other via a Gaussian weight mask (an aperture). The mask was centred at the centre of gaze, allowing for a smooth transition between the two images and uncovering a window of 2° visual angle of the underlying photograph. The visual aperture uncovered an area equivalent to two fingers at arm's length, reflecting the most commonly used exploratory procedure by participants in the haptic face recognition experiments. See *Figure 5.1 (B,C)* for examples of the stimuli used for gaze-restricted visual face recognition.

Procedures and Experimental Designs

Participants performed the experiment in only one condition. Learning-phase, identification task and testing-phase were conducted in the same modality, i.e. either in the haptic, gaze-restricted-visual or unrestricted-visual modality. The general design of the experiments was the same to ensure comparability. In the haptic and unrestricted-visual experiments, the faces were positioned on a platform that was placed horizontally, on top of a fixed table. All faces could be rigidly fixed to this platform and were always presented from a frontal view. Participants used a chin rest that was placed 30 cm away from the stand on which the objects were presented. An opaque curtain that could be slid back to reveal faces for the visual experiments separated the participants from the stand. During haptic exploration of the faces, an arm-rest was provided to prevent exhaustion.

In the gaze-restricted experiments, participants were seated about 60 cm away from a computer screen (21-inch CRT) resting their chin on a chin rest and used a mouse to move a Gaussian window which uncovered 2° of the photograph of the 3D face. Participants were instructed not to move the mouse rapidly back and forth, for such a method would have produced an effective visual field much larger than intended, since very rapid scanning differs little from simultaneous dull display (Ikeda and Uchikawa (1978)) by virtue of screen and visual persistence. Since we recorded the trajectories we were able to control for this confound. No trials had to be deleted.

Before performing the experiment in the haptic and gaze-restricted conditions, we presented one stimulus and asked the naïve participants to explore it and to report what kind of object they were dealing with. Every participant identified the stimulus correctly as a face. Participants were then familiarized with three upright faces (out of 19 total) that were randomly chosen from six sets of three faces each. We labeled each face with a short first name. They were told to explore the face masks carefully and to learn their names because they would be asked to recognize those particular faces later. No further information was given about the nature of the following experiment during the familiarization.

In the subsequent identification task, participants had to name each randomly presented face mask after exploration. Feedback was provided in that participants were told whether the face was recognized correctly or not. Each face mask had to be identified correctly twice before the experiment continued.

The old/new recognition task immediately followed the identification task and consisted of 3 blocks of 19 trials, corresponding to 3 old (learned) and 16 new faces (each object was shown once per block). This asymmetric design was chosen because of time constraints for haptic learning. Face masks were shown one at a time in random order with an ISI of 10 sec in which the faces were exchanged. Participants were asked to explore each face mask and to report whether it was one of the three faces they had learned (old) or not (new). Although exploration time was unrestricted, they were instructed to respond as quickly and accurately as possible by pressing an 'old' or 'new' labeled key on a keyboard with their left hand. Participants took about 10 min to complete a haptic or gaze-restricted test-block and 4 min for an unrestricted visual test-block. No feedback was provided for the old/new recognition task.

5.2.2 Results

Responses were converted to standard d' scores and averaged across test-blocks. The means and standard error for each condition are shown in *Figure 5.2 (A)*. One-tailed t-tests showed that performance was above chance for each condition (H: $t_{53} = 6.64, p < 0.001$; GRV: $t_{53} = 6.91, p < 0.001$; UV: $t_{53} = 13.98, p < 0.001$). The results were further analyzed using two-tailed t-tests to compare performance in the three conditions. We found that face recognition performance was best in the unrestricted visual condition (UV vs. H: $t_{53} = -7.52, p < 0.001$; UV vs. GRV: $t_{53} = -6.04, p < 0.001$), i.e. visual face recognition accuracy was significantly reduced by promoting serial encoding to levels of haptic face recognition accuracy since we found no significant difference in performance for haptic and gaze-restricted face recognition (H vs. GRV: $t_{53} = -0.58, p = 0.56$).

5.2.3 Discussion

We found a clear advantage for unrestricted visual face recognition over gaze-restricted visual and haptic face recognition with no difference between the latter. More specifically, face recognition performance across the visual and haptic sense was equated by reducing the visual window to the narrowness of the effective field-of-view in haptics (due to a decrease in visual face recognition accuracy as compared to unrestricted visual face recognition). However, we previously reported a memory effect for haptic face recognition in a cross-modal transfer task where haptic face recognition performance was significantly improved by refreshing memory (Dopjans et al. (2009)). Since we found haptic and gaze-restricted face recognition to be at the same level, we suggested that haptic and gaze-restricted face recognition might

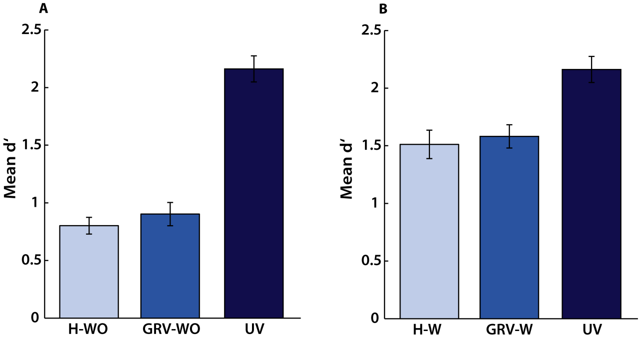

Figure 5.2: Plots comparing face recognition performance (A) in Experiment 1 for haptic (H-WO), gaze-restricted (GRV-WO), and unrestricted visual face recognition without refreshing memory, and (B) in Experiment 2 for haptic (H-W) and gaze-restricted (GRV-W) face recognition with refreshing memory (data for unrestricted visual face recognition as shown for comparability). Data are measured in mean d' ± 1 Standard Error of the Mean (SEM).

be equally impeded by limitations in working memory due to higher memory demands for serial encoding of faces and might be aided by refreshing memory.

5.3 Experiment 2

To test the hypothesis that haptic and gaze-restricted face recognition are limited by restrictions in working memory due to higher memory demands for serial encoding of faces, we repeated the experiment using a 'with refreshing memory' version of the experimental design introduced in Experiment 1.

5.3.1 Methods

Subjects

36 experimentally naïve participants (18 per condition) were paid 8 Euros an hour to perform the respective experiment. All participants reported normal or corrected-to-normal vision and had no sensory impairment.

Stimuli

The same stimuli were used as in Experiment 1.

Procedures and Experimental Designs

Participants performed the experiment in only one condition. Learning-phase, identification task and testing-phase were conducted in the same modality, i.e. either in the haptic or gaze-restricted visual modality. The general design of the experiments was the same to ensure comparability.

The design was the same as in Experiment 1 except that memory was refreshed by repeated exposure to the three learned faces. That is, the identification task was conducted *before each test-block.*

This experiment, therefore, comprised two different conditions: haptic (H-W) and gaze-restricted visual (GRV-W) face recognition with refreshing memory while the data for unrestricted (UV) visual face recognition without refreshing memory from Experiment 1 provided a baseline performance measure.

5.3.2 Results

Responses were converted to standard d' scores and averaged across test-blocks. The means and standard error for each condition are shown in *Figure 5.2 (B)*. One-tailed t-tests showed that performance was above chance for each condition (H-W: $t_{53} = 8.71, p < 0.001$; GRV-W: $t_{53} = 11.54, p < 0.001$). The results were further analyzed using two-tailed t-tests to assess the effect of refreshing memory on haptic and gaze-restricted face recognition performance in comparison to baseline performance (unrestricted visual face recognition without refreshing memory) and whether face recognition performance was affected differently by refreshing memory in the two modalities. Again, we found that face recognition performance was best in the unrestricted visual condition (UV vs. H-W: $t_{53} = -3.31, p < 0.01$; UV vs. GRV-W: $t_{53} = -2.81, p < 0.01$) with no difference between haptic and gaze-restricted performance (H-W vs. GRV-W: $t_{53} = -0.33, p = 0.74$). Finally, we compared the results to those from Experiment 1 (haptic (H-WO) and gaze-restricted vision (GRV-WO) without refreshing memory) using a 2x2 ANOVA with factors Modality (haptics and gaze-restricted vision) and Memory (with and without refreshing memory) to test directly how face recognition performance was affected by refreshing memory in the two modalities. We found a significant main effect for Memory ($F_{1,212} = 23.95, p < 0.001$) for which post-hoc tests revealed that face recognition performance was significantly better 'with-refreshing memory' ($p < 0.001$), but no significant interaction Memory x Modality ($F_{1,212} = 0.01, p = 0.91$) indicating modality-independent higher memory demands for serial encoding of faces. Moreover, we found no significant effect for Modality ($F_{1,212} = 0.36, p = 0.55$).

Taken together, we therefore found the same pattern of results for haptic and gaze-restricted visual face recognition.

Furthermore, reaction time data demonstrated the sequential nature of restricted viewing, as RTs for gaze-restricted visual face recognition were significantly slower than for unrestricted vision and at the same level as for haptic face recognition (H-WO: 19.96 ± 0.80 sec; GRV-WO: 16.60 ± 0.97 sec; UV: 4.29 ± 0.35 sec; H-W: 20.03 ± 0.92 sec; GRV-W: 16.30 ± 0.84 sec). It is important to note, however, that long response times as observed in this haptic and gaze-restricted recognition task are difficult to interpret and only allow for limited conclusions. The data here is thus reported only for completeness' sake.

5.3.3 Discussion

Similarly to Experiment 1, we found that haptic and gaze-restricted exploration resulted in the exact same recognition pattern. Compared to the previous results, memory refreshing led to a similar increase in performance in both modalities. These results suggest *modality-independent* higher working memory demands for serial encoding of faces equally affecting gaze-restricted and haptic face recognition performance.

Taken together, we found the same pattern of results for haptic and visual face recognition performance using a gaze restricted display. Given these modality-specific encoding differences the question arises how they might affect face processing strategies.

5.4 Experiment 3

We used the face-inversion paradigm to assess the effect of encoding differences on face processing strategies (featural vs. configural) in haptic (H), gaze-restricted (GRV) and unrestricted visual (UV) face recognition, by comparing recognition performance for upright (-U) vs. inverted (-I) faces in each modality.

5.4.1 Methods

Subjects

54 experimentally naïve participants (18 per modality) were paid 8 Euros an hour to perform the respective experiment. All participants reported normal or corrected-to-normal vision and had no sensory impairment.

Stimuli

The same stimuli were used as in Experiment 1.

Procedures and Experimental Designs

Participants performed the experiment in only one modality (unrestricted visual, haptic or gaze-restricted visual). There was no switch in modality between learning and testing-phase. The general design of all experiments was the same to ensure comparability.

The same set-ups were used as described above for Experiment 1 for presenting upright as well as inverted (upside-down) faces. Furthermore, the learning and identification phase were conducted with upright faces as described above for Experiment 1. Using a 'with-refreshing memory' design, the old/new recognition task immediately followed the identification task and consisted of 3 blocks of 19 trials, corresponding to 3 old (learned) and 16 new faces (each object was shown once per block). In block 1 participants were presented with upright faces, while inverted (upside-down) faces were presented in the second and third test-block.

In the old/new recognition task face masks were shown one at a time in random order with an ISI of 10 sec in which the faces were exchanged. Participants were asked to explore each face mask and to report whether it was one of the three faces they had learned (old) or not (new). Although exploration time was unrestricted, they were instructed to respond as quickly and accurately as possible by pressing an 'old' or 'new' labeled key on a keyboard with their left hand. Participants took about 10 min to complete a haptic or gaze-restricted test-block and 4 min for an unrestricted visual test-block. No feedback was provided for the old/new recognition task.

Again, participants performed the experiment in only one modality, i.e. participants in the haptic modality learned faces haptically, proceeded to the identification task in the haptic modality, followed by the haptic old/new recognition task with upright (block 1) and inverted (blocks 2 and 3) faces.

5.4.2 Results

Responses were converted to standard d' scores and averaged across inverted test-blocks. The means and standard error for each condition are shown in *Figure 5.3*. One-tailed t-tests were used to assess that performance was above chance for each modality and orientation (H-U: $t_{17} = 4.29, p < 0.001$; H-I: $t_{17} = 4.22, p < 0.001$; GRV-U: $t_{17} = 5.33, p < 0.001$; GRV-I: $t_{17} = 5.43, p < 0.001$; UV-U: $t_{17} = 9.13, p < 0.001$; UV-I: $t_{17} = 3.38, p < 0.01$). As a second step, we performed a 3x2 factorial ANOVA with the between-subjects factor being Modality (haptic, gaze-restricted, unrestricted visual) and the within-subjects factor being Face-Orientation (upright, inverted). While we failed to find a significant main effect for Modality ($F_{2,102} = 0.118, p = 0.89$) and Face-Orientation ($F_{1,102} = 3.37, p = 0.70$), we found a highly significant interaction between the two factors ($F_{2,102} = 11.16, p < 0.001$). Two-tailed t-tests revealed that face recognition performance in the upright orientation was significantly

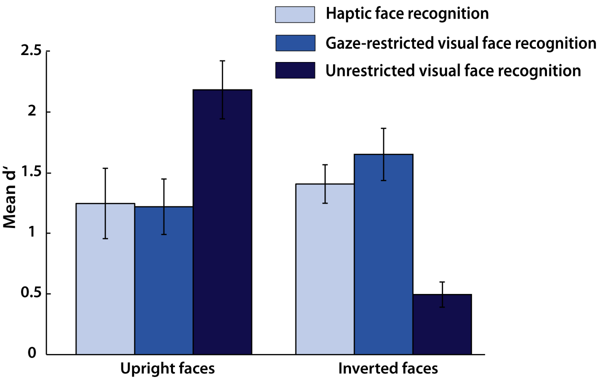

Figure 5.3: Plot comparing face recognition performance for upright and inverted faces for haptic, gaze-restricted and unrestricted visual recognition. Data are measured in mean d' ± 1 SEM.

better in the unrestricted visual modality than in the haptic and gaze-restricted conditions (UV-U vs. H-U: $t_{34} = 2.49, p < 0.05$; UV-U vs. GRV-U: $t_{34} = 2.91, p < 0.01$; H-U vs. GRV-U: $t_{34} = 0.07, p = 0.94$). More interestingly however, we found a clear inversion effect in the unrestricted visual modality while no decrease in recognition accuracy was found in either the haptic or gaze-restricted modality (UV-U vs. UV-I: $t_{34} = 6.02, p < 0.001$; H-U vs. H-I: $t_{34} = -0.44, p = 0.66$; GRV-U vs. GRV-I: $t_{34} = -1.13, p = 0.26$). Consequently, in the inverted orientation, performance was significantly better in the haptic and gaze-restricted modalities than in the unrestricted visual modality (UV-I vs. H-I: $t_{34} = -3.40, p < 0.01$; UV-I vs. GRV-I: $t_{34} = -3.43, p < 0.01$; H-I vs. GRV-I: $t_{34} = -0.06, p = 0.52$).

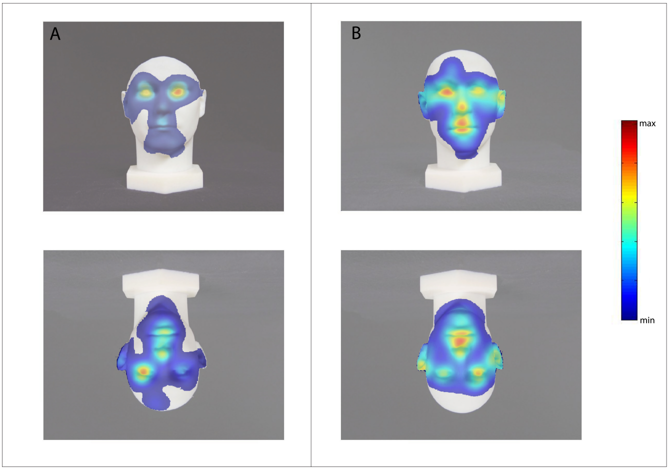

Figure 5.4: Heat maps illustrating the relative number of fixation per trial on a given screen position on the stimuli for an upright and an inverted trial for column (A) a single subject and (B) averaged across all subjects for one stimulus. Fixation positions were smoothed using a Gaussian filter with a sigma corresponding to 2° visual angle in order to account for the fixation position variability when fixating a certain point.

5.4.3 Discussion

We found a strong face inversion effect for unrestricted visual but none for haptic or gaze-restricted face recognition using our 3D face masks. While we found evidence for sensitivity to first order relations, as all participants recognized the stimulus as a face, the lack of a face inversion effect seems to indicate that (i) face inversion effects may rely on simultaneous processing of facial features and (ii) participants adopted a feature-based strategy for haptic and gaze-restricted face recognition. In the unrestricted visual modality, however, participants clearly seem to have benefited from the use of a configural processing strategy in the old/new recognition task.

Analysis of gaze-restricted exploration

Exploratory patterns from the gaze-restricted visual condition strongly support the use of a feature-based strategy in vision when serial encoding is enforced (*Figure 5.4*). For an initial analysis we pooled the trajectories smoothed with a Gaussian window to create heat maps resembling those used to visualize eye tracking data. The heat maps shown in (*Figure 5.4*) demonstrate (1) that participants fixated single features rather than moving the window around a lot - a strategy more consistent with obtaining second order relations such as inter-featural distances, and (2) that there is a general similarity of exploratory patterns in the upright and inverted conditions, i.e. there were no clear and systematic difference between conditions and orientations in terms of the location of gaze fixations. The latter finding is well illustrated by *Figure 5.4 (B)* in which averaged exploration trajectories are shown for one stimulus: eyes, nose, mouth, and ears result in much more interest than other parts of the face for both conditions.

Considering the small differences between conditions, we subjected the fixation data to a simple analysis in terms of duration and number. We calculated fixations from the recorded trajectories. Fixations were defined as consecutive data points lying within the 2° visual angle window (40 pixel per degree) for at least 500 msec, following a 'gaze-restricted saccade' (velocity difference between two consecutive windows of duration exceeding a pre-defined acceleration criterion defined as window size (40 pixel per degree) * acceleration threshold (4 degrees per second^2)). We used two-sided t-tests to compare duration of fixation and number of fixations for gaze-restricted face recognition for the *without-* (number: 3.97 ± 0.18; duration: 9.62 ± 0.28 sec) and *with-refreshing memory* conditions (number: 3.44

± 0.1; duration: 8.07 ± 0.24 sec), as well as for upright (number: 5.56 ± 0.17; duration: 12.68 ± 0.6 sec) and inverted (number: 5.31 ± 0.14; duration: 12.2 ± 0.53 sec) faces. We found that participants made significantly more fixations in the *without-* than *with-refreshing* memory condition ($t_{2050} = -3.68, p < 0.001$). Likewise, fixations in the *without-refreshing* memory condition were significantly longer ($t_{2050} = -3.4, p < 0.001$). Given that face recognition accuracy was significantly lower in the *without-refreshing* memory condition, the higher number along with the longer duration of fixations seems to reflect participants' difficulties to remember the learned faces and extract the right information for subsequent face recognition.

Similarly to face recognition accuracy, we found no significant difference for the number of fixations ($t_{682} = -1.13, p = 0.26$) or the duration ($t_{682} = 0.59, p = 0.55$) for upright and inverted faces supporting the interpretation that participants used the same, *feature-based*, strategy for recognizing upright and inverted faces in gaze-restricted face recognition.

Further analysis of the exploration pattern and comparison with eye tracking data will be necessary to go into more depth about the characteristic differences between serial and 'holistic' exploration.

5.5 General Discussion

Research on the nature of haptic encoding has highlighted an important distinction regarding the relative effectiveness with which the visual and haptic systems encode objects, more specifically, faces. While faces are encoded holistically in vision (all aspects of an image are processed in parallel; Maurer et al. (2002)), haptic encoding is limited to serial exploration of an object (involves a feature-by-feature analysis; Loomis et al. (1991); Loomis and Lederman. (1986)) due to its narrow effective field-of-view. Given these modality-specific encoding differences the question arises how they might affect high-level cognitive tasks such as face recognition. As discussed in the introduction, visual expert face processing is characterized by configural processing. If information gained through serial haptic encoding can be accurately integrated into a more global representation (e.g., Lakatos and Marks (1999)), haptic face processing might also benefit from configural processing. If not, haptic face processing should rely more on featural information.

Here, we studied the effects of modality-specific encoding differences in visual and haptic face recognition in terms of recognition accuracy as well as information processing strategies. We found that (1) face recognition was equally disrupted using gaze-restricted vision and haptics as compared to unrestricted vision, indicating that face processing is impeded by serial encoding even when participants have control over the information that they view through the aperture. (2) Using gaze-restricted vision increased the time necessary to recognize a face relative to unrestricted visual face recognition. More specifically, the time required to recognize faces using gaze-restricted vision was similar to that for haptic face recognition. (3) Participants' gaze-restricted exploration of faces focused mainly on single features which is consistent with previous investigations of the relative importance of different internal face features for recognition (Haig (1986); Schyns et al. (2002); Sekuler et al. (2004); Yarbus (1967)) on the time scale of feature integration (i.e., simultaneous versus sequential).

Our experiments yield the exact same pattern of results for haptic and visual face recognition performance using a gaze restricted display. Not only was face recognition performance across the visual and haptic sense equated by reducing the visual window to the narrowness of the effective field-of-view in haptics (due to a decrease in visual face recognition accuracy as compared to unrestricted visual face recognition), but we also observed modality-independent higher working memory demands for serial encoding of faces. That is, recognition performance is low in the without-refreshing memory condition in haptic and gaze-restricted face recognition and significantly improved when memory to the three learned faces is refreshed before each test-block. Since refreshing memory is not necessary in the unrestricted visual condition, the higher working memory demands are likely to be attributed to temporal integration of serially gained information for face recognition in both modalities. This finding agrees with a previous study in which recognition performance across the visual and haptic sense was equated in 2D picture recognition by reducing the visual window to the narrowness of the effective field-of-'view' in haptics (Loomis et al. (1991)). The authors concluded that recognition performance in restricted field-of-view conditions was, indeed, impeded by limitations in working memory or in the integration process. Our study, however, is the first, to our knowledge, to directly assess the effect of modality-specific encoding differences on working memory demands and consequently their effect on performance in a high-level cognitive task such as face recognition.

Furthermore, with faces presented in the unrestricted visual condition, we replicated the well-known face inversion effect in our task, as previously observed in several behavioral studies: With the exact same stimuli to recognize, participants performed significantly better with upright than inverted faces. More interestingly, however, we failed to find such an inversion effect for haptic or gaze-restricted face recognition. While we found evidence for sensitivity to first order relations, as all participants recognized the stimulus as a face, the lack of a face inversion effect indicates that participants adopted a feature-based strategy for haptic and gaze-restricted face recognition. In the unrestricted visual modality, however, participants clearly seem to have benefited from the use of a configural processing strategy in the old/new recognition task.

Our findings agree well with a recent study by van Belle et al. (2010), in which the authors used a similar gaze-contingent stimulus presentation method to study the visual face inversion effect. They compared participants' face discrimination performance on (1) faces presented in full view, (2) with only the central window of vision revealed and, (3) with only the fixated feature masked by means of an eye-contingent mask. Similar to our results, they found a face inversion effect for faces presented in full view but none when observers had their vision constrained such that they could see only through a small central window. The authors, therefore, concluded that the inversion effect is not caused primarily by a difficulty in perceiving local detailed facial features but by the observers' inability to simultaneously extract diagnostic information at different locations on an inverted face, i.e. that holistic face perception is impaired for inverted faces. Our results confirm and extend these findings by providing the first direct comparison between serial exploration in visual and haptic processing of faces.

As mentioned in the introduction, Kilgour and Lederman (2006) previously used a haptic face inversion paradigm to study orientation-sensitivity of haptic face processing and found a strong inversion effect for faces. An important difference to our study, however, lies in the task chosen to investigate orientation specificity for haptically explored faces. While we used an old/new recognition task, Kilgour and Lederman (as well as Lakatos and Marks (1999) who studied configural processing in haptics for non-face objects) chose a 2AFC same/different face discrimination task. Given that configural processing on the basis of serial encoding requires that serial information is accurately integrated into a global representation, the working memory demands in an old/new recognition task are much higher than in a same/different task, which taps mainly into short-term memory. These differences in memory load might favor the use of featural over configural information in an old/new recognition task - future studies are needed to assess the differences between the two tasks in more detail.

Moreover, McGregor et al. (2010) recently investigated the use of configural versus feature-based processing in haptic identity classification of upright versus inverted versus scrambled faces using 2D raised-line displays. While performance was low overall, they found that upright and scrambled faces produced equivalent accuracy, and both were identified more accurately than inverted faces. While this finding was taken to indicate that the upright orientation was 'privileged' in the haptic representation the authors suggested that the effect of scrambling argued against the use of configural information. Scrambling faces alters an object-centered description of a face. It also changes a body-centered description of the face as a configuration, while maintaining the local features in their normal, upright orientation within this body-based frame of reference. The fact that there was no statistical difference in accuracy between upright and scrambled faces was thus taken to indicate that configural information was not used to haptically process facial identity in raised-line drawings. Our findings are, therefore, interesting in light of our recent study investigating cross-modal transfer in visual and haptic face recognition where we observed asymmetric cross-modal face information transfer (a cost in vision-to-haptics but none in haptics-to-vision transfer resulting in equal cross-modal recognition performance, Dopjans et al. (2009)). 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 from a holistic, visual representation. Conversely, haptic recognition quite likely does not benefit from the holistic information encoded by vision and might, therefore, be limited by the use of feature-based information. The observed performance differences in visual and haptic face recognition might, therefore, be attributed to qualitative differences in information processing due to modality-specific encoding differences. Further research, such as the comparison of exploratory procedures in unrestricted vision, haptics and gaze-restricted vision using eye-tracking and motion-capture, is, however, necessary to thoroughly study differences in encoding strategies in these modalities.

Furthermore, our results may lend indirect evidence to the idea that information processing is expertise dependent in the ongoing face-specificity vs. expertise debate in configural information processing. While many studies argue that faces represent a special class of objects in that certain types of information processing, in particular configural processing of facial features, are specific to face perception (e.g., Kanwisher (2000)), other findings suggest that they are a matter of expertise (e.g., Diamond and Carey (1986); Gauthier and Tarr (1997)). Our results suggest that not only haptics but also gaze-restricted vision are limited to serial information encoding and that both process feature-based face information, while unrestricted vision processes configural information.

Finally, inasmuch as we have little to no training in haptic or gaze-restricted visual face recognition throughout life, it is possible that participants might be able to develop strategies to compensate for processing differences introduced by serial encoding. Further research is required to elucidate the role of expertise in face recognition such as a training study to investigate whether encoding-process-dependent differences in information processing strategies can be overcome with the acquisition of, for example, gaze-restricted face recognition expertise. In general, continued investigation of information encoding and processing of faces using gaze-restricted vision and haptic exploration will continue to increase our understanding of the cognitive and neural processes underlying human face recognition.

Chapter 6

Learning to recognize faces through serial exploration

Abstract- Human observers are experts at visual face recognition due to specialized visual mechanisms for face processing that evolve with perceptual expertise. Such expertise has long been attributed to the use of configural processing, enabled by fast, parallel information encoding and impeded by serial encoding. Here we tested whether participants can learn to efficiently recognize faces that are serially encoded. Ten participants were trained to be experts at gaze-restricted face recognition. Tests comparing expert with novice performance revealed: (1) a marked improvement in terms of speed and accuracy, (2) a gradual development of configural processing strategies (3) experts' ability to rapidly learn and accurately recognize new exemplars. Taken together, post-training participants displayed hallmarks

of expertise for gaze-restricted face recognition suggesting that they were able to learn new strategies to compensate for the serial nature of information encoding. The results are also discussed in terms of their relevance for other sensory modalities that rely on serial information encoding.

6.1 Introduction

Human observers are experts at visual face recognition. Consequently, face processing has received a lot of attention in vision research providing evidence for specialized visual mechanisms that evolve with perceptual expertise. Amongst the hallmarks of face processing expertise are:

- the use of configural as opposed to featural processing (processing of individual face parts). Three types of such configural processing have been defined (Maurer et al. (2002), see also Gauthier and Tarr (2002)): (i) sensitivity to first order relations (individual object parts are better recognized in the context of other parts than isolation), (ii) holistic processing (ability or tendency to consider all parts of an object simultaneously, regardless of the exact configuration of parts), and (iii) sensitivity to second order relations (perceiving inter-feature distances; individual object parts are placed in the context of the other individual parts from the same object). Strictly speaking, however, only holistic and relational processing (perceiving inter-feature distances) have been shown to increase with expertise (Gauthier and Tarr (2002)).
- that face perception is *orientation-specific*, i.e. faces are processed more accurately when they are presented in the

> normal upright position than when they are inverted (for reviews, see Searcy and Bartlett (1996); Valentine (1988)). According to a long-standing and influential hypothesis of this so called 'face-inversion' effect, first demonstrated by Yin (1969), vertical inversion selectively impairs our ability to extract configural information from faces, while leaving featural processing largely intact (Leder and Bruce (2000); Schwaninger et al. (2006)), i.e. while an upright face is processed at a global level, i.e. that of the whole face, an inverted face would have to be processed at a more local level, feature by feature (but also see Sekuler et al. (2004)).

In fact, van Belle et al. (2010) recently showed that the face inversion effect is, indeed, caused by an inability to perceive the individual face as a whole rather than as a collection of specific features, thus supporting the view that observers' expertise at upright face recognition is due to the ability to perceive an individual face holistically. Using a gaze-contingent stimulus presentation method to study the visual face inversion effect, they compared participants' face discrimination performance on (1) faces presented in full view, (2) with only the central window of vision revealed, and (3) with only the fixated feature masked by means of an eye-contingent mask. They found a face inversion effect for faces presented in full view but none when observers had their vision constrained such that they could see only through a small central window. The authors, therefore, concluded that the inversion effect is not caused primarily by a difficulty in perceiving local detailed facial features but by the observers' inability to simultaneously extract diagnostic information at different locations on an inverted face, i.e. that holistic face perception is impaired for inverted faces. The authors, however, did not specifically discuss the effect of constraining the effective field of view on encoding strategies. While unrestricted vision can process all aspects of an image in parallel, so that both local facial features and their global configuration can be rapidly processed (Tanaka and Sengco (1997)), constraining the effective field of view limits participants to serial exploration of an object, i.e., it is not holistic but involves a feature-by-feature analysis (Loomis et al. (1991); Loomis and Lederman. (1986)).

We previously tested the effect of encoding differences on face recognition performance in unrestricted vision, gaze-restricted vision and haptics (the latter as a sensory modality that is 'naturally' limited to serial encoding; Dopjans et al. (2010)). This was achieved by using a gaze-restricted display which promoted serial encoding in vision. The gaze-restricted display limited the effective field-of-view in vision such that only one feature, determined by the observer him/herself, was available *at the same time* on a face. In a first series of experiments, we compared haptic, gaze-restricted and unrestricted visual face recognition. Secondly, we used the face-inversion paradigm to assess how encoding differences might affect face processing strategies (featural vs. configural face information processing).

By promoting serial encoding in vision, we found the same pattern of results for haptic and visual face recognition performance using a gaze restricted display. Not only was face recognition performance across the visual and haptic sense equated by reducing the visual window to the narrowness of the effective field-of-view in haptics (due to a decrease in visual face recognition accuracy as compared to unrestricted visual face recognition), but we also found a strong face inversion effect for unrestricted visual but none for gaze-restricted face recognition. Taken together, our results suggest that configural processing is enabled by fast and parallel information encoding and impeded by restricted, serial encoding.

Given these observed effects of encoding differences on information processing strategies, the question arises whether participants can learn to efficiently recognize faces that are serially encoded. It is well established that hallmarks of expert face processing, such as orientation-sensitivity and configural processing, take many years to develop, with processing starting out as largely featural or element by element based and gradually developing into configural processing as displayed in adults (Carey and Diamond (1977); Dahl et al. (2009); Hay and Cox (2000); Maurer et al. (2002); Mondloch et al. (2003); Pellicano and Rhodes (2003); Schwarzer (2000)). Inasmuch as we have little to no training in gaze-restricted visual face recognition throughout life, it is possible that participants might be able to develop strategies to compensate for processing differences introduced by serial encoding. In particular, if participants were able to learn to accurately integrate information gained through serial encoding into a more global representation (e.g., Lakatos and Marks (1999)), gaze-restricted face processing might also benefit from configural processing.

While it is not expected to equate laboratory-trained expertise to real-world expertise, as the latter occurs on the scale of years (e.g., Carey and Diamond (1977); Maurer et al. (2002); Mondloch et al. (2003)), whereas typical laboratory training studies require only hours of training (Gauthier and Tarr (1997, 2002); Malpass et al. (1973)), training studies allow for the manipulation of different factors that may contribute to the acquisition of expertise, providing better control over variables influencing this process. Nevertheless, only few experimental studies of laboratory-acquired perceptual expertise have been reported (Gauthier et al.

(1999a); Gauthier and Tarr (1997, 2002); Gauthier et al. (1999b, 1998); Scott et al. (2006b, 2008)). Gauthier and colleagues have used training studies to examine the acquisition of perceptual expertise using novel objects called 'Greebles' (Gauthier et al. (1999a); Gauthier and Tarr (1997, 2002); Gauthier et al. (1999b, 1998)). The first of these investigations found that training not only led to faster and more accurate responses, but training also increased the configural (and thus facelike) processing of Greebles (Gauthier and Tarr (1997)), evidenced by increased reaction time to trained Greeble configurations (studied parts, in a studied configuration) compared to transformed Greeble configurations (studied parts, in a different configuration). Tests of generalization of learning after Greeble training suggested that learning generalized to Greebles that were structurally similar to the training set, but did not generalize to Greebles that were less similar to the training set (Gauthier et al. (1998)).
Taken together, the following hallmarks of the acquisition of perceptual expertise have been proposed (Palmeri and Cottrell (2010)):

- A marked improvement in terms of speed and accuracy for expert recognition compared to novices. One important aspect of the speedup is the so-called 'entry-level shift' whereby subordinate-level categorizations are made as quickly as basic-level categorizations in experts, while novices are faster at basic-level than subordinate-level categorizations.

- The gradual development of configural processing strategies, more accurately, holistic and relational processing, as, for example, measured using the face inversion effect. While novices are largely unaffected by inversion, expert recognition is significantly impaired when faces are presented upside-down.

- The ability to rapidly learn and accurately recognize new exemplars. That is, expertise allows generalization to previously unknown members of an expert object class. More specifically, generalization of learning occurs when performance improvements with a specific set of trained exemplars generalize or transfer to previously unlearned exemplars.

Here we present results that directly test the effect of training on information processing of serially encoded faces using a gaze-restricted display. Participants were trained in gaze-restricted face recognition on five consecutive days. The pre-test on day one established baseline face recognition performance of novices on upright and inverted faces, followed by two days of gaze-restricted training on upright faces. The post-test on day four again tested gaze-restricted face recognition performance on upright and vertically inverted faces. Face-inversion effects were used as a marker of face expertise. Intuitively, the most obvious way to assess expertise with a perceptual category is to determine how well experts learn new exemplars of the category. If our trained participants were truly gaze-restricted face recognition experts, we would expect them to be able to transfer expertise derived from our training procedure to new stimuli. We, therefore, tested their *generalization* ability for the learned stimulus class on day five. Finally, we assessed the *permanence* of any learning effect by repeating the post-test in a final session seven days after training was completed.

6.2 Methods

Because so few experimental studies of perceptual expertise have been reported, little is known about the best methods for manipulating the level of expertise. It is obvious that experts are generally more experienced than novices but it is not clear exactly how much experience is necessary to produce significant 'expertise effects' and how to determine the reached level of expertise. There are, therefore, essentially two types of expertise training methods. (1) All participants receive a fixed amount of training that the experimenters think will be enough to result in expertise, or (2) each participant receives a variable amount of training, until he or she reaches a pre-specified criterion, assuming that this criterion is an adequate measure of expertise. The advantage of the latter is that one can be reasonably confident that at the conclusion of training, all participants are 'experts' in terms of a globally defined performance criterion. The disadvantage, however, is that the criterion training method makes it harder to assess the results of tasks performed during the training procedure itself, since different participants will go through different amounts of training. Here, we, therefore, chose to have each participant perform exactly the same training procedure until a group average criterion of $d' > 3.6$ (100% hit rate and 10% false alarm rate; in the experiment described later this translates to only 2 incorrect 'recognitions' per subject and block) was reached. For this, all participants were tested on the same day. Moreover, we opted for an unsupervised learning task in which participants performed the previously used gaze-restricted old/new recognition task. The unsupervised learning design was chosen as previous training studies have stressed the importance of unsupervised exposure in the formation of expert perceptual abilities (Scott et al. (2006b, 2008)).

6.2.1 Subjects

10 experimentally naïve participants were paid 8 Euros an hour to perform the experiment. All participants reported normal or corrected-to-normal vision and had no sensory impairment.

6.2.2 Stimuli

Stimuli were two sets of 19 face images. The faces for the two sets were chosen on the basis of visual similarity ratings to ensure comparability. Stimuli were generated and presented under Matlab 7.11 using the Psychophysics Toolbox (Brainard (1997); Pelli (1997)). Each gaze-restricted stimulus was created using two images. The first was one of 19 photographs from a frontal view of the white plastic face masks previously used in haptic experiments (Dopjans et al. (2009, 2010)). The faces spanned 14.7 ± 1.2° visual angle in the vertical plane and 9.1 ± 0.5° visual angle in the horizontal plane and were presented on a black background spanning 25.85° visual angle in the horizontal plane and 19.52° visual angle in the vertical plane. The second was a black image that was superimposed on the photograph. These two images were blended into each other via a Gaussian weight mask (an aperture). The mask was centred at the centre of gaze, allowing for a smooth transition between the two images and uncovering a window of 2° visual angle of the underlying photograph. The visual aperture uncovered an area equivalent to two fingers at arm's length, reflecting the most commonly used exploratory procedure by participants observed in previous

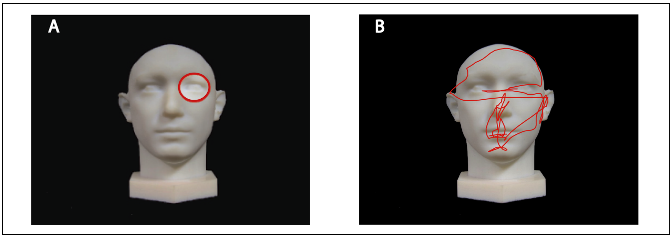

Figure 6.1: (A) Demonstration of the gaze-restricted display: The red circle indicates the size of the aperture. Only the part of the image inside the aperture was visible as indicated by the difference in brightness of the images inside and outside of the aperture. The aperture of 2° visual angle was moved over the frontal photograph of the face mask. (B) Example of a recorded trajectory during gaze-restricted face recognition.

haptic face recognition experiments. See *Figure 6.1* for examples of the stimuli used.

6.2.3 Experimental Design

Participants were trained in gaze-restricted face recognition using an old/new recognition task on 5 consecutive days. The pre-test on day 1 assessed participants' ability to recognize upright and vertically inverted faces using a gaze-restricted display. During training on days 2 and 3, participants performed the old/new recognition task on upright faces only. As participants reached our pre-defined criterion after three days of training the post-test on day 4 assessed gaze-restricted face recognition performance of upright and vertically inverted faces. Participants saw stimulus set 1 on days 1 to 4. On day 5, stimulus set 2 was used for upright and inverted gaze-restricted face recognition to rule out that possible learning effects might be solely due to familiarity with the faces. Finally, six participants returned for a final session 7 days later (on day 12) to perform the old/new recognition task with upright and inverted faces from stimulus set 1.

6.2.4 Procedure

Participants were seated about 60 cm away from a computer screen (21-inch CRT) resting their chin on a chin rest and used a mouse to move a Gaussian window which uncovered 2° of the photograph of the 3D face. Participants were instructed not to move the mouse rapidly back and forth, for such a method would have produced a much larger effective visual field, since very rapid scanning differs little from simultaneous dull display (Ikeda and Uchikawa (1978)) due to screen and visual persistence. In addition, we recorded the trajectories to check for this confound. No trials had to be deleted.

Participants were familiarized with three upright faces (out of 19 total) that were randomly chosen from six sets of three faces each. We labeled each face with a short first name. They were told to explore the face masks carefully and to learn their names because they would be asked to recognize those particular faces later. No further information was given about the nature of the following experiment during the familiarization.

In the subsequent identification task, participants were randomly presented with the three learned (upright) faces and had to name each face after exploration. Feedback was provided in that participants were told whether the face was recognized correctly or not. Each face mask had to be identified correctly twice before the experiment continued. This identification task with feedback was repeated before each test block.

The old/new recognition task immediately followed the identification task and consisted of 4 blocks of 19 trials, corresponding to 3 old (learned) and 16 new faces (each object was shown once per block). This asymmetric design was chosen because of time constraints for gaze-restricted learning. Face masks were shown one at a time in random order. Participants were asked to explore each face mask and to report whether it was one of the three faces they had learned (old) or not (new). Although exploration time was unrestricted, they were instructed to respond as quickly and accurately as possible by pressing an 'old' or 'new' labeled key on a keyboard with their left hand. No feedback was provided for the old/new recognition task.

The training sessions on days 2 and 3 consisted of 4 test-blocks with upright faces. On the other days, the experiment comprised two test-blocks with upright and two test-blocks with vertically inverted (upside-down) faces. During the learning phase, as well as the identification task upright faces were used.

6.3 Results

6.3.1 Behavioral data

Responses were converted to standard d' scores and averaged across test-blocks. The means and standard error for day and orientation are shown in *Figure 6.2*. One-tailed t-tests showed that performance was above chance for each face-orientation on each day (day 1: upright: $t_9 = 3.54, p < 0.01$; inverted: $t_9 = 5.31, p < 0.001$; day 2 (upright): $t_9 = 10.38, p < 0.001$; day 3 (upright): $t_9 = 12.35, p < 0.001$; day 4: upright: $t_9 = 16.31, p < 0.001$; inverted: $t_9 = 11.30, p < 0.001$; day 5: upright: $t_9 = 14.06, p < 0.001$; inverted: $t_9 = 13.93, p < 0.001$; day 12: upright: $t_5 = 11.33, p < 0.001$; inverted: $t_5 = 10.86, p < 0.001$). Our pre-defined criterion of $d' >= 3.6$ was reached after only 3 days of training.

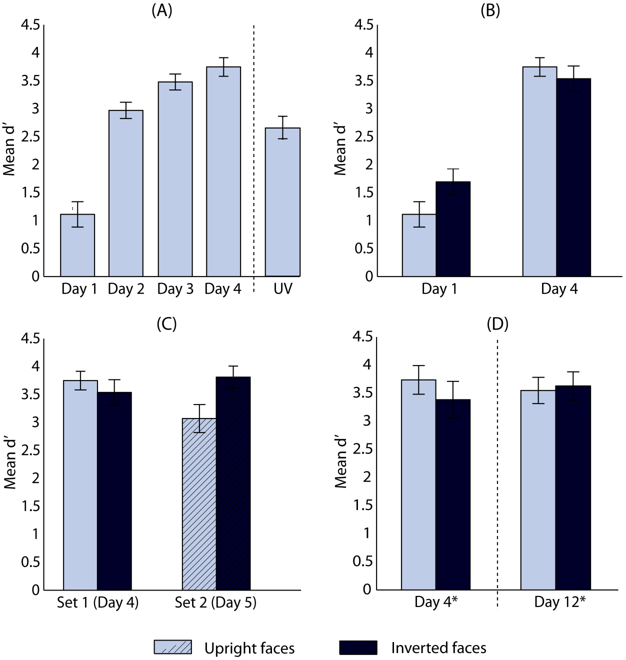

Figure 6.2: Plots comparing face recognition performance (A) for upright faces on days 1 to 4 and compared to unrestricted visual face recognition, (B) for upright and inverted faces for gaze-restricted face recognition novices (day 1) and 'experts' (day 4), (C) for upright and inverted faces for experts using the training set (set 1, day 4) and a new set of faces (set 2, day 5), and (D) for upright and inverted faces for experts at the end of training (day 4*) and after 7 days (day 12*; * indicates data from only 6 subjects). Data are measured in mean d' ± 1 Standard Error of the Mean (SEM).

We further analyzed the results for the upright face orientation on days 1 to 4 using a repeated-measures ANOVA to test for a learning effect. We found a significant main effect for days ($F_{1,9} = 167.01, p < 0.001$) indicating that face recognition performance significantly improved through training, so much so that post-training performance was significantly better than for unrestricted visual face recognition ($t_{26} = 3.44, p < 0.01$; data taken from a previous experiment using the same task and stimuli testing *unrestricted visual* face recognition, Dopjans et al. (2009)). Secondly, we tested for inversion effects on days 1, 4, 5 and 12 using two-tailed, paired t-tests to compare performance for upright and vertically inverted faces on the respective day. Interestingly, we failed to find an inversion effect for gaze-restricted visual face recognition on any day (day 1: $t_9 = -1.36, p = 0.21$; day 4: $t_9 = 0.70, p = 0.50$; day 5: $t_9 = -1.68, p = 0.11$; day 12: $t_5 = -0.87, p = 0.42$), indicating that the observed improvement in face recognition performance was not accompanied by a change in information processing strategies (replicating and extending Dopjans et al. (2010)). Thirdly, we used two-tailed paired t-tests to compare face recognition performance on upright faces for stimulus set 1 (on day 4) and set 2 (on day 5) to test generalizability of the observed learning effect to a new set of stimuli. We found no significant difference between performances on days 4 and 5 ($t_9 = 2.23, p = 0.06$), showing that participants were indeed able to *generalize* newly learned strategies for efficient gaze-restricted face recognition to a new set of stimuli. Finally, we compared recognition performance on upright faces on days 4 and 12 to test permanence of the observed learning effect. Since we were only able to test six participants for the experiment on day 12, we re-analyzed performance on day 4 for those six participants only and used a paired t-test to compare performances. Again, we found no significant difference between results on day 4 and 12 ($t_5 = -0.11, p = 0.92$), indicating that the newly acquired perceptual skill persisted for at least one week.

6.3.2 Response times

Response times of all trials were averaged across test-blocks. The means and standard error for each day and orientation are shown in *Figure 6.3*. We analyzed response times for the upright face orientation on days 1 to 4 using a repeated-measures ANOVA to test for a learning effect. We found a significant main effect for days ($F_{1,9} = 144.44, p < 0.001$) indicating that participants became faster at gaze-restricted face recognition performance through training. More interestingly, we compared response times on day 4 (after training) to those from a previous experiment using the same task and stimuli testing *unrestricted visual* face recognition (Dopjans et al. (2009)) using two-tailed t-tests. We found *no* significant difference ($t_{26} = 1.80, p = 0.08$), indicating that through training, response times improved to levels of unrestricted visual face recognition - a result that can be taken as further evidence for developing expertise in gaze-restricted face recognition. Secondly, we tested for inversion effects (longer response times for inverted than upright faces) on days 1, 4, 5 and 12. A repeated-measures ANOVA revealed a significant interaction for factors Training and Orientation ($F_{1,9} = 14.92, p < 0.01$). While we failed to find an inversion effect on day 1 with participants being -almost significantly- faster for inverted than upright faces ($t_9 = 2.26, p = 0.05$), post-training participants were significantly faster at recognizing upright than inverted faces on day 4 ($t_9 = -2.42, p < 0.05$). Furthermore, two-tailed, paired t-tests revealed that participants were significantly faster

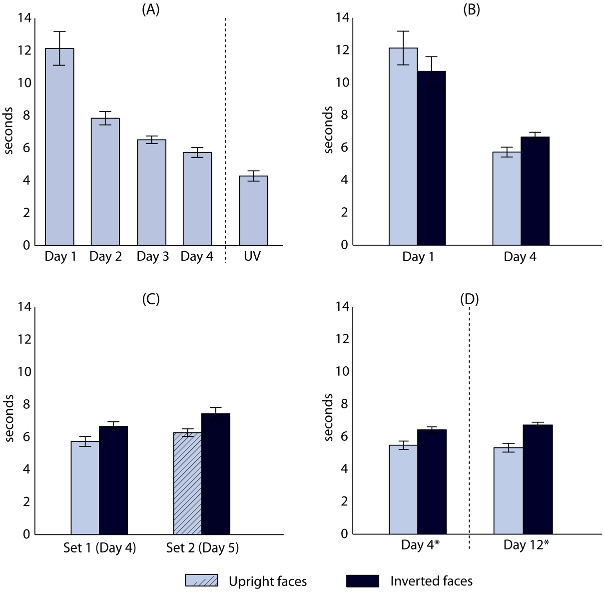

Figure 6.3: Plots comparing response times for gaze-restricted face recognition performance (A) for upright faces on days 1 to 4 and compared to unrestricted visual face recognition, (B) for upright and inverted faces for gaze-restricted face recognition novices (day 1) and 'experts' (day 4), (C) for upright and inverted faces for experts using the training set (set 1, day 4) and a new set of faces (set 2, day 5), and (D) for upright and inverted faces for experts at the end of training (day 4*) and after 7 days (day 12*; * indicates data from only 6 subjects). Data are measured in mean seconds ± 1 Standard Error of the Mean (SEM).

recognizing upright than inverted faces on day 5 and -just reaching significance- on day 12 (day 5: $t_9 = -2.34, p < 0.05$; day 12: $t_5 = -2.26, p = 0.05$). Taken together, these results indicate a potential switch in processing strategies through training. Thirdly, we used two-tailed paired t-tests to compare response times for upright faces of stimulus set 1 on day 4 and set 2 on day 5 to test generalizability of the observed learning effect to a new set of stimuli. We found no significant difference between response times on days 4 and 5 ($t_9 = -1.23, p = 0.25$), showing that participants were indeed able to generalize newly learned strategies for efficient gaze-restricted face recognition to a new set of stimuli. Finally, we compared response times on upright faces on days 4 and 12 to test permanence of the observed learning effect. We re-analyzed performance on day 4 for the six participants who did the experiment on day 12 and used a paired t-test to compare performances. Again, we found no significant difference between results on day 4 and 12 ($t_5 = 0.47, p = 0.66$).

6.3.3 Analysis of gaze-restricted exploration

For an initial analysis we pooled the aperture trajectories smoothed with a Gaussian window to create heat maps resembling those used to visualize eye tracking data. The heat maps for day 1 shown in *Figure 6.4 (A)* clearly demonstrate (1) that participants fixated single features rather than, for example, long-range distances between features and (2) that there is a general similarity of exploratory patterns in the upright and inverted conditions, i.e. there were no clear and systematic difference between orientations in terms of the location of gaze fixations: eyes, nose, mouth, and ears result in much more interest than other parts of the face for both conditions. Heat maps for day 4 shown in *Figure 6.4 (B)*, however, suggest a change in exploratory patterns

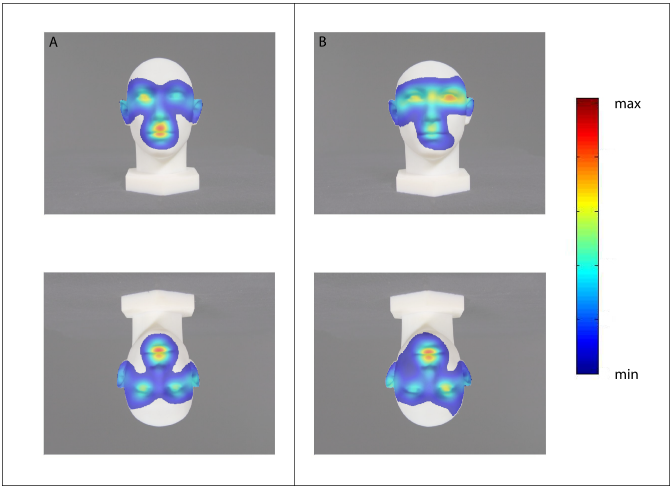

Figure 6.4: Heat maps illustrating the relative number of fixations per trial (averaged across all subjects for one stimulus) at a given screen position for an upright and an inverted trial before expertise training on day 1 (A) and after expertise training on day 4 (B). Fixation positions were smoothed using a Gaussian filter with a sigma corresponding to 2° visual angle in order to account for the fixation position variability when fixating a certain point.

through training in that (1) seemingly fewer fixations are made in the upright orientation which in turn are less feature-oriented and (2) a noticeable difference between orientations in terms of the pattern and number of fixations made.

Consequently, we further analyzed the fixations quantitatively in terms of duration and number before and after expertise training. For this, we first calculated gaze fixations from the recorded trajectories based on the same logic as for eye movement fixations as follows: Fixations were defined as consecutive data points lying within the 2° visual angle window (40 pixel per degree) for at least 500 msec, following a 'gaze-restricted saccade'. A saccade occurred whenever the velocity difference between two consecutive windows of duration exceeded a pre-defined acceleration criterion (defined as window size (again, 40 pixel per degree) * acceleration threshold (4 degrees per second^2)). See *Table 6.1* and *Table 6.2* for the number and duration of fixations for each day.

Participants made significantly fewer and shorter fixations after training than before (number: $t_{379} = 5.38, p < 0.001$; duration: $t_{379} = -4.22, p < 0.001$) - a finding that is also evident in the shorter, overall response times we observed.

Face Orientation	**Day 1**	**Day 4**	**Day 5**	**Day 12**
upright	4.97 ± 0.15	3.24 ± 0.11	3.63 ± 0.1	3.22 ± 0.09
inverted	4.68 ± 0.12	3.82 ± 0.14	3.91 ± 0.11	3.19 ± 0.93

Table 6.1: Number of fixations made on upright and inverted faces on days 1, 4, 5 and 12. Data are given as mean ± 1 SEM.

Face Orientation	**Day 1**	**Day 4**	**Day 5**	**Day 12**
upright	5.58 ± 0.33	3.09 ± 0.36	3.75 ± 0.22	2.35 ± 0.15
inverted	5.00 ± 0.35	5.26 ± 0.13	3.83 ± 0.30	3.39 ± 0.25

Table 6.2: Duration of fixations made on upright and inverted faces on days 1, 4, 5 and 12. Data are given in seconds ± 1 SEM.

Similarly to face recognition accuracy and response times for day 1, we found no significant difference for the number of fixations ($t_{379} = 1.8, p = 0.07$) or the duration ($t_{379} = 1.42, p = 0.15$) for upright and inverted faces supporting the interpretation that novices on day 1 used the same, *feature-based*, strategy for recognizing upright and inverted faces. More interestingly, however, we found that participants made significantly fewer and shorter fixations on upright than inverted faces after expertise training on day 4 (number: $t_{379} = -3.94, p <$

0.001; duration: $t_{379} = -3.1, p < 0.01$). Similarly, our experts made significantly fewer fixations on upright than inverted faces on day 5 ($t_{379} = -2.03, p < 0.05$), with no significant difference in duration ($t_{379} = -0.25, p = 0.81$) and significantly shorter fixations on upright than inverted faces on day 12 ($t_{379} = -3.82, p < 0.001$), with no significant difference in number of fixations ($t_{379} = 0.29, p = 0.77$). Taken together, these results support the response time data, indicating that the information used for recognizing upright faces cannot be easily extracted for inverted faces, lending further support towards a potential change in information processing strategies through expertise training. It is important to note that this potential change did, however, not lead to more accurate performance.

Participants made more fixations on the new set of faces on day 5 than on the faces from the training set on day 4 ($t_{379} = -2.57, p < 0.05$) with no significant difference in duration ($t_{379} = -1.53, p = 0.13$) and no significant difference in the number ($t_{379} = 0.13, p = 0.89$) or duration of fixations ($t_{379} = 1.88, p = 0.06$) on days 4 and 12. The increased number of fixations made on upright faces on day 5 compared to day 4 can most likely be attributed to the novelty of the stimulus set and is not reflected in the behavioral or response time data.

6.4 Discussion

We have previously shown that serial encoding similarly impedes visual and haptic face recognition. In both cases we failed to find a face inversion effect which could be due to promoting featural versus configural processing of facial information, the latter being widely considered *the* hallmark of expert face processing. This finding raised the question whether modalities that rely on serial encoding of information actually allow for expert face processing. Consequently, here we trained participants in gaze-restricted face recognition to assess whether they can learn to efficiently recognize faces that are serially encoded. In summary, we found (1) that participants became significantly faster (to levels of unrestricted visual face recognition) and better (exceeding levels of unrestricted visual face recognition) at gaze-restricted face recognition through short training, as our pre-defined performance criterion was met after three days of training, (2) an inversion effect for response times (and fixations, but not accuracy) for 'experts' but not novices, (3) improvement in performance did not arise from familiarity with the faces but transferred to novel faces and (4) expertise lasted at least one week. Our results, therefore, suggest that our experts did indeed learn how to learn and recognize faces during the training procedure. This learning took place using a gaze-restricted display that promotes serial encoding in vision, which we will discuss in more detail in the context of the following three points: (1) a significant improvement in performance and response times, (2) a qualitative shift in processing strategies and (3) the expertise derived from our training procedure apparently having transferred well to a second test-set.

First, the significant improvement in performance and response times that we observed during expertise acquisition, while only changes in magnitude, represent a learning effect. Although participants received very little feedback during training (only during the identification task) performance improved quickly and passed our pre-defined criterion after only three days of training (at least 36 trials of the identification task, 228 trials of the old/new recognition task). Interestingly, post-training recognition of serially encoded faces exceeded levels of unrestricted visual recognition. Since participants still displayed hallmarks of face processing expertise such as the inversion effect (Dopjans et al. (2010)), the comparably poor performance for unrestricted visual recognition might simply be attributed to the lack of pigmentation cues in our 3D face masks. Both, pigmentation and shape cues, have been shown to contribute equally to face recognition (O'Toole et al. (1999); Yip and Sinha (2002)), such that a lack of either cue might worsen recognition performance. Thus, as gaze-restricted participants might have also learned compensatory strategies for the lack of pigmentation cues during expertise acquisition in order to achieve high levels of recognition accuracy, it is reasonable to assume that unrestricted visual face recognition of our 3D face masks might as well improve through training.

Moreover, participants became significantly faster at recognizing serially encoded faces to the level of unrestricted visual face recognition. The latter is of interest as Tanaka and Taylor (1991) suggest that one hallmark of expertise processing, the so-called 'entry-level shift', is that experts are as fast to recognize objects of their expertise at the subordinate level (for example, 'robin' for bird experts) as they were at the basic level ('bird'). In contrast, non-experts are consistently faster on basic-level discriminations. Similarly, because humans are face experts, judgments of face identity (subordinate level) are as fast as judgments that are more categorical, for instance gender, or whether the shown stimulus was a face, or not. While we did not measure response times for these types of tasks, the finding that participants showed response times for gaze-restricted face recognition at the level of unrestricted visual face recognition in a subordinate-level recognition task was taken as indirect evidence that participants displayed this hallmark of face processing expertise. Importantly, this speed advantage for experts did not result in a speed-accuracy trade-off as they were faster *and* more accurate, indicating that experts and novices differ in the knowledge of what kinds of information are helpful in recognizing faces. Indeed, a comparison of fixation patterns before and after training suggests a change in preferred features with training (*Figure 6.4*). While normal observers in unrestricted visual face recognition usually show a preference for the eye region when viewing upright faces (see below; Tanaka and Farah (1993); Sergent (1984); McKelvie (1976); Goldstein and Mackenberg (1966)), gaze-restricted novices show a preference for the mouth area for upright and inverted faces. Interestingly, similar preferences for the mouth region in face exploration have been reported for haptic exploration (Lederman et al. (2010)) as well as patient groups with face recognition deficits such as prosopagnosic patient L.R. (Bukach et al. (2006)) and individuals with autism (Joseph and Tanaka (2003); Klin et al. (2002); Langdell (1978)). Gaze-restricted experts, on the other hand, show the 'normal' preference for the eye region for upright faces but, interestingly, not for inverted faces. Indeed, this pattern of fixation behavior for gaze-restricted experts agrees well with eye-tracking studies of unrestricted visual face recognition in which eyes are looked at more frequently than any other facial part when faces are presented in a natural (upright) way (Dahl et al. (2009); Tanaka and Farah (1993); Sergent (1984); Barton et al. (2006); Emery (2000); Farroni et al. (2002)). In contrast to upright faces, face inversion has been shown to lead to a drastic loss of eye preference in human faces (Dahl et al. (2009)). The saliency of the eye region has, therefore, been argued not to be due to low-level appearance, but to be driven by higher-level expectations based on the spatial configuration of the face (Guo et al. (2003)). Additionally, a high proportion of eye fixations is, to some extent, believed to be indicative of holistic face processing (Dahl et al. (2007)). These findings might, therefore, indicate a switch in information processing strategies from featural to configural, through expertise training. (A more detailed analysis of the exploratory patterns for gaze-restricted face recognition is currently in preparation). Secondly, it seems clear that expertise should be more than simply a practice effect involving a qualitative shift in processing strategies. As mentioned above, we used the inversion effect as a measure to evaluate the nature of the abilities acquired by experts in processing faces (Diamond and Carey (1986); Sergent (1984); Yin (1969)): For novices, we replicated previous results as we failed to find an inversion effect in terms of accuracy as well as response times, indicating the use of featural processing strategies (Dopjans et al. (2010)). Interestingly, novices even showed a 'paradoxical' inversion effect for response times as they were slower on upright than inverted faces. While this effect has previously been shown in prosopagnosic patients its cause remains unclear (Farah et al. (1995b); Gelder and Rouw (2000)). For experts, however, response times were significantly faster on upright than inverted faces. Crucially, this difference represents a qualitative change in the recognition behavior of experts indicating that the expertise manipulation produced a speed advantage for upright over inverted faces. This speed advantage was reflected in a similar shift in fixation behavior: While the number and duration of fixations was generally reduced through training, again indicating that experts and novices differ in the knowledge of what kinds of information are helpful in recognizing faces, the qualitative switch in fixation behavior for upright and inverted faces in experts is crucial: experts but *not* novices made significantly fewer and shorter fixations on upright than inverted faces. The reduced number and duration of fixations together with the observed inversion effect suggests that experts fixated less on single features and rather moved the window around a lot - a strategy more consistent with obtaining second order relations such as inter-featural distances- while still covering roughly the same area of a face pre- and post-training (*Figure 6.4 (B)*). This finding is consistent with Leder and Bruce (2000) hypothesis that relational information is processed locally, rather than derived from a holistic template. Taken together, response time and fixation results might suggest that the information used for recognizing upright faces cannot be easily extracted for inverted faces, possibly due to a change in information processing strategies from featural to configural, through expertise training.

What is not entirely clear is why our participants showed this sensitivity in response times and not in accuracy. We previously found a strong face inversion effect for unrestricted visual but none for haptic or gaze-restricted face recognition novices using our 3D face masks (Dopjans et al. (2010)). This inversion effect was found for recognition accuracy, as well as response times (response time data not published; for similar results see van Belle et al. (2010)). Of course, psychophysical models rarely allow one to predict a priori whether a difference between conditions will manifest itself in one dependent measure or another (Gauthier and Tarr (1997); Tanaka and Farah (1993); Tanaka and Sengco (1997)). Alternatively, novices might not abruptly switch from one type of processing to another during expertise acquisition but rather undergo a more continuous shift of the type of processing. Studying the acquisition of perceptual expertise with 'Greebles', Gauthier and Tarr (2002), for example, found that holistic and relational processing- both types of configural processing that are affected by inversion- develop on different time scales and appear to be very strongly related to the amount of expertise. This gradual shift in strategies might have manifested itself in the response time measure first whereas an advantage of configural processing for recognition accuracy might only be established over a longer period of time with more training. After all, the inversion effect takes many years to first develop in children (Carey and Diamond (1977); Dahl et al. (2009); Hay and Cox (2000); Maurer et al. (2002); Mondloch et al. (2003); Pellicano and Rhodes (2003); Schwarzer (2000)). While participants showed sensitivity to configural information (taking into account the precise relations between different parts of objects as well as the parts themselves), they might still have relied on featural information for recognition. The shift in processing strategies as indicated by the qualitative change in response times and fixation behavior might, nonetheless, reflect the initial steps of adapting a configural processing strategy for serially encoded faces.

Our results agree well with a previous study in which Gauthier and Tarr (1997) addressed the question whether the putatively face-specific use of configural processing strategies might be part of a more general recognition mechanism fine-tuned by experience with homogeneous stimuli. Adapting the experimental paradigm that Tanaka and colleagues had previously used to illustrate sensitivity to configural changes in face processing (Tanaka and Farah (1993); Tanaka and Sengco (1997)), Gauthier and Tarr tested sensitivity to configural transformations for novices and experts with a homogeneous set of nonface stimuli ('Greebles'). They found that (i) they could create experts with about 7-10 hours of training using the 'entry-level shift' criterion (Tanaka and Taylor (1991)) and (ii) Greeble experts, but not novices, exhibited sensitivity to configural changes similar to those observed in face processing but only for upright Greebles. More specifically, posttraining, moving certain parts of a Greeble affected the processing of its other, untransformed, parts. Interestingly, participants showed this sensitivity in response time while Tanaka's participants had previously shown it with faces in accuracy, supporting our hypothesis that short-term, laboratory-acquired expertise might manifest itself in terms of sensitivity to configural information more easily in response times than accuracy. Future studies with a more extensive training phase are needed to assess the evolution of processing strategies during acquisition of expertise for serially encoded faces and their manifestation in terms of response times and accuracy, in more detail.

Alternatively, the observed decrease in response time and fixations might reflect the optimization of exploratory strategies as indicated by the above mentioned change of preferred features from the mouth to the eye region. While the face inversion effect has been shown to be a robust marker of face expertise (Diamond and Carey (1986); Sergent (1984); McKone et al. (2007)), further studies with our setup using other experimental paradigms to test configural processing such as the composite-effect (e.g., Hole (1994)), configural changes in features (e.g., Freire et al. (2000)), the Thatcher Illusion (e.g., Boutsen and Humphreys (2003)), or scrambling facial features (e.g., Collishaw and Hole (2000)) are necessary to fully investigate the use of processing strategies in gaze-restricted face recognition. Moreover, as we trained participants on upright faces, one might also argue that the observed inversion effect could be attributed to a mere exposure effect. It would, therefore, be interesting to train participants on inverted faces. Robbins and McKone (2003) have shown that orientation specificity of unrestricted visual face processing is highly stable against practice when participants failed to learn holistic processing for inverted faces (in contrast to the situation for objects; Tarr and Pinker (1989)). Whether or not participants would be able to learn holistic processing for inverted faces using gaze-restricted vision would shed further light on the question of orientation-specificity of gaze-restricted face recognition.

Thirdly, the most obvious way to assess expertise with a perceptual category is to determine how well experts learn new exemplars of the category (Gauthier et al. (1998)). After their training had been completed, our experts performed the task on a new set of faces. Since we found no significant difference in accuracy, response times or fixation behavior between recognition of faces from the training set and novel faces, the expertise derived from our training procedure appeared to have transferred well to a novel stimulus set. This finding, again, suggests that through training participants learned what kind of information is helpful in recognizing serially encoded faces and that this knowledge is not based on familiarity with the training faces but can be generalized to novel faces.

Finally, participants were as accurate and fast at recognizing serially encoded faces after seven days without practice. More interestingly, the difference in response times for upright and inverted faces just reached significance ($p = 0.05$), indicating that participants might need more training to consolidate and extend their previously acquired expert skills. Therefore, while our participants displayed hallmarks of face processing expertise for serially encoded faces after only 3 hours of training, more training might be necessary to consolidate and extend their level of expertise.

6.4.1 Relevance for haptic processing

Although we usually use vision when attempting to recognize faces, recent studies have shown that humans are capable of identifying individual faces at levels well above chance using only their sense of touch (Kilgour and Lederman (2002); Casey and Newell (2007); Dopjans et al. (2009); Kilgour et al. (2004); Kilgour and Lederman (2006); Pietrini et al. (2004)). This similarity between the sensory modalities seems to stem partially from the fact that both systems are able to process the 3D geometrical structure of objects. In fact, there is evidence that the two systems may share a common neural substrate that is responsible for computing object form (Amedi et al. (2002); James et al. (2002)), raising the question whether the visual and haptic modality encode similar information and share the hallmarks of face processing, i.e. whether the use of expert face processing strategies, for example, is modality independent.

As mentioned in the introduction, however, vision and haptics involve qualitatively different information encoding strategies. While faces are encoded holistically in vision (Maurer et al. (2002)), haptic encoding is limited to a feature-by-feature analysis of an object such that haptic information has to be integrated over time in order to take in the same amount of information (Loomis et al. (1991); Loomis and Lederman. (1986)). These modality-specific encoding differences affect information processing strategies, in that serial encoding impairs the use of expert face processing strategies such as configural processing (Dopjans et al. (2010)). Given, however, that we have little to no training win haptic face recognition the question arises whether participants can learn to efficiently recognize faces through haptic exploration. Studying haptic perceptual expertise not only informs us about the learning capabilities of the haptic system, but also has the potential to constrain hypotheses about how the haptic and visual systems process similar information.

Gaze-restricted viewing was used here to constrain the visual system to serial, self-directed exploration of object features in much the same way that the haptic system explores objects (Loomis (1981)). James and colleagues previously trained participants for 10 to 12 hours in order to be able to haptically identify faces and Greebles with 100% accuracy (James et al. (2006)). Unfortunately, their study only assessed the influence of familiarity on brain activation during haptic exploration of 3D face masks and did not test for hallmarks of face expertise for haptic face processing. As a next step, it will, therefore, be important to test the acquisition of haptic face processing expertise as we did here for gaze-restricted visual face recognition.

The present study shows how practice with a previously novel way of perceiving faces, i.e. through serial encoding, can lead to some of the recognition effects typically associated with unrestricted visual face recognition, indicating that -at least for vision- serial encoding of information might allow for expert face processing.

Chapter 7

Visual experience is necessary for efficient haptic face recognition

Abstract- Several studies have shown that humans (as well as monkeys) become experts for face recognition with dedicated neural machinery that subserves face processing. Here, we describe an initial experiment with a novel approach to the question how perceptual expertise shapes efficient face processing strategies by studying face recognition in the blind. More specifically, we investigated haptic face recognition in the sighted, congenitally blind, and acquired blind. If visual experience were necessary for haptic face recognition, one would expect the congenitally blind participants to perform worse than acquired blind or sighted participants. Visually impaired individuals, however, are more familiar than sighted people with the use of touch for pattern perception, potentially resulting in superior haptic face

recognition performance. Using an old/new recognition task we found a pronounced advantage for acquired blind (visual and haptic expertise) and sighted (visual but no haptic expertise) participants over the congenitally blind (no visual but haptic expertise) in haptic face recognition. This data, thus, provides behavioral evidence for the crucial role of visual input for the development of efficient face processing capabilities. The study, therefore, directly contributes to the central question of how faces are processed in humans and how perceptual expertise shapes these processing strategies.

7.1 Introduction

Face processing has received a lot of attention in vision research providing evidence for specific processing strategies that evolve with perceptual expertise (Maurer et al. (2002); Schwarzer et al. (2007)). Prior studies have shown that humans can also recognize faces by touch alone (Kilgour and Lederman (2002)). In previous experiments, we provided further evidence that both the haptic and visual system have the capacity to process faces, and that face-relevant information can be shared across sensory modalities (Dopjans et al. (2009)). Interestingly, we found this information transfer across modalities to be asymmetric and limited by haptic face processing. While visual face perception relies on specific processes that evolve with perceptual expertise, sighted people have little to no training in haptic face recognition throughout life. We, therefore, suggested that the observed asymmetry in visual and haptic face processing might be attributed to different levels of perceptual expertise in the two modalities. It is, therefore, crucial to understand which prerequisites are necessary for efficient haptic face recognition, i.e. is visual or haptic expertise more important for haptic face recognition?

In order to interact with their environment more efficiently, blind people are usually extensively trained to use their sense of touch to encode the shape of Braille letters, familiar objects, etc. (see Saito et al. (2007)). Given this haptic expertise, one would, therefore, also expect a behavioral benefit in haptic face recognition for the blind. An alternative hypothesis, however, might be that haptic face recognition requires previous visual exposure to faces - especially given the complexity of the object class and the fact that specialized areas for face processing have been found in the visual processing pathway (Kanwisher et al. (1997)). In this case one would expect acquired blind (having visual and haptic expertise) to be better at haptic face processing than congenitally blind (having only haptic expertise).

Here, we set out to assess the importance of visual experience with faces versus perceptual haptic expertise by investigating haptic face recognition in the congenitally blind, acquired blind and sighted.

7.2 Methods

7.2.1 Stimuli

Three-dimensional (3D) models of nineteen faces were taken from the MPI-Face-Database (Troje and Bülthoff (1996)) and edited for printing using the graphics package 3D Studio Max (Autodesk). Face masks were printed with the use of an Eden 250 printer (Objet Geometries Ltd.). In our case, face masks weighed about 138 ± 5 g each and measured 89 ± 5.5 mm wide, 120 ± 7.5 mm high and 103.5 ± 5.5 mm deep. Due to technical constraints

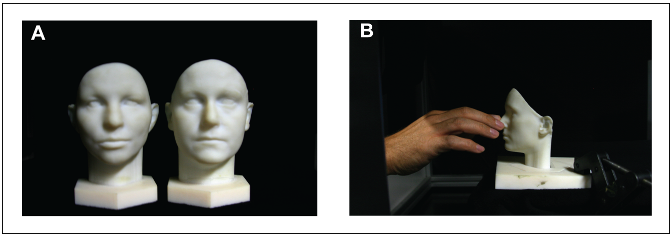

Figure 7.1: (A) Two example stimuli used for haptic face recognition. (B) Experimental setup.

the experiments were conducted with these smaller-than-life face masks. We have, however, previously shown that stimulus size does not significantly affect haptic face recognition performance (Dopjans et al. (2009)). While the use of plastic face masks deprives participants of 'natural' material information, Kilgour and Lederman (2002) have shown that while haptic face encoding involves both geometric and material properties, greater emphasis is placed on the geometric cues. However, since depriving faces of material information resulted in a significant decrease in performance in their study, we asked participants in debriefing questionnaires following the experiment about how they were affected by the object material in solving the task. All participants replied not to have been affected by the artificial material. Moreover, all participants reportedly recognized the stimuli as faces and treated them as such. For an example of the stimuli used see *Figure 7.1 (A)*.

7.2.2 Apparatus

The faces were positioned on a platform that was placed horizontally, on top of a fixed table. All faces could be rigidly fixed to this platform and were always presented from a frontal view.

Participants used a chin rest that was placed 30 cm away from the stand on which the objects were presented. An opaque curtain separated the participants from the stand. During haptic exploration of the faces, an arm-rest was provided to prevent exhaustion (*Figure 7.1 (B)*).

7.2.3 Participants

Eighteen sighted (mean age 23 years, 9 male, 9 female), nine acquired blind (mean age 19.4 years) and nine congenitally blind participants (mean age 18.6 years) volunteered and gave informed consent for an experimental protocol that was conducted under general ethical guidelines. All participants were right-handed (For a sample description of the blind participants see *Table 7.1*). All the blind participants in our study were rated in terms of their cognitive abilities by their supervisors and teachers. We found no obvious correlation of cognitive capabilities such as grades or performance in school with haptic performance, e.g., one individual with reported learning disabilities had the best haptic face recognition performance whereas another participant with reportedly normal cognitive abilities scored lowest.

Participant No	**Gender**	**Age**	**Age at Onset of Blindness**	**Cause of Blindness**
CB 1	M	20	-	Lebers cong Amaurosis
CB 2	M	17	-	Bilateral Anophthalmia
CB 3	M	18	-	Bardet-Biedl Syndrom
CB 4	M	20	-	Bilateral Leukoma cornea
CB 5	M	21	-	Retinopathia praematurorum GIII, Optic atrophy
CB 6	F	17	-	Bilateral optic atrophy (septo-optic dysplasia)
CB 7	F	17	-	Bilateral Anophthalmia
CB 8	M	19	-	Lebers optic atrophy
CB 9	M	18	-	Bilateral Retinopathia praematurorum
AB 1	M	23	6	Bilateral optic atrophy
AB 2	M	19	5	Hypoplasia papillae
AB 3	M	18	13	High myopia, retinal detachment
AB 4	F	17	5	Lebers cong Amaurosis
AB 5	M	20	13	Lebers cong Amaurosis
AB 6	F	20	12	Retinits pigmentosa
AB 7	F	23	17	Retinitis pigmentosa
AB 8	F	18	8	Bilateral optic atrophy
AB 9	F	17	5	Bilateral enucleation due to retinoblastoma

Table 7.1: Sample description of blind participants. - CB= congenitally blind, AB= acquired blind.

7.2.4 Experimental Design

Participants were seated in front of a table, behind an opaque curtain such that they did not see the face masks. Before performing the experiments, we presented one face mask and asked the naïve participants to explore it haptically and to report what kind of object they were dealing with. All participants were then haptically familiarized with three faces (out of 19 total) that were randomly chosen from six sets of three faces each. We labeled each face with a short first name. Participants were allowed to explore the faces only haptically using the right hand, with no constraint on either the exploratory procedure or the duration of exploration. They were told to explore the face masks carefully and to learn their names because they would be asked to recognize those particular faces later. No further information was given about the nature of the following experiment during the familiarization. Haptic learning of the three faces took 4 min on average. In the subsequent identification task, participants had to name each randomly presented face mask after haptic exploration. Feedback was provided in that participants were told whether the face was recognized correctly or not. Each face mask had to be identified correctly twice before the experiment continued. This identification task was repeated before each test block.

The old/new recognition task immediately followed the identification task and consisted of 3 blocks of 19 trials, corresponding to 3 old (learned) and 16 new faces (each object was shown once per block). This asymmetric design was chosen because of time constraints for haptic learning. Face masks were shown one at a time in random order with an ISI of 10 sec in which the faces were exchanged. Participants were asked to explore each face mask haptically and to report whether it was one of the three faces they had learned (old) or not (new). Audio signals indicated begin and end of the exploration. As before, participants were free to use their own exploratory strategy to explore the faces. Although exploration time was unrestricted, they were instructed to respond as quickly and accurately as possible by pressing an 'old' or 'new' labeled key on a keyboard with their left hand. Participants took about 10 min to complete a haptic block. No feedback was provided for the old/new recognition task.

In addition to recognition accuracy we measured the number of repetitions needed in the identification task until each face was identified correctly twice and response times (the time between the onset of stimulus presentation and the participant pressing one of the response keys).

7.3 Haptic face recognition performance

Before performing the experiments, we presented one face mask and asked the naive participants to explore it haptically and to report what kind of object they were dealing with. While all participants were able to correctly identify the presented object as a face it is important to understand which prerequisites are necessary for efficient haptic face recognition, i.e., is visual or haptic expertise more important for haptic face recognition?

Again, if visual expertise was more important for efficient haptic face recognition than haptic expertise, we would expect the acquired blind to perform as well as, if not better than, the sighted and both better than the congenitally blind. If, on the other hand, haptic expertise was more important than visual expertise, the congenitally blind should perform as well as, if not better than, the acquired blind and both better than the sighted.

While there is no data on how often blind people engage in haptic face processing, all blind participants in both groups told us that they touch faces as often as the social norms allow them to. They also mentioned that they would like to touch faces more often as this would give them additional information about the person they would be interacting with. None of the participants mentioned that the plastic-like material of our faces was disturbing or 'unnatural' when asked during the debriefing.

7.3.1 Results

Responses were measured in d'. Performance was evaluated for the last block (3) in which all participants had become sufficiently acquainted with the difficult task (also see *numrep*, Number of repetitions in the identification task). One-tailed t-tests were used to assess whether performance was above chance for the respective group (sighted, acquired blind, congenitally blind). Haptic face recognition performance was significantly above chance for all three groups (sighted: $t_{17} = 6.42, p < 0.001$; acquired blind: $t_8 = 7.62, p < 0.001$; congenitally blind: $t_8 = 2.16, p < 0.05$; *Figure 7.2*). More interestingly, however, a One-Way-ANOVA followed by Bonferroni post-hoc tests revealed that recognition accuracy was significantly better in acquired blind and sighted than congenitally blind participants ($F_{2,33} = 4.29, p = 0.02$, sighted vs. congenitally blind: $p < 0.05$; acquired vs. congenitally blind: $p < 0.01$) with no significant difference between these two groups ($p > 0.05$).

7.3.2 Discussion

The results reported here are consistent with the hypothesis that visual expertise is more important for efficient haptic face recog-

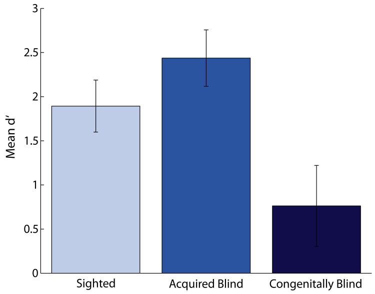

Figure 7.2: Plot comparing haptic face recognition performance in block 3 in the sighted, acquired blind and congenitally blind, measured in d'. Data are represented as mean ± 1 SEM.

nition than haptic expertise as we found a clear advantage in haptic face recognition for acquired over congenitally blind. Hence, it seems that a lack of visual expertise cannot be compensated for by purely perceptual haptic expertise as the low performance for the congenitally blind shows. These findings, therefore, directly contribute to the central question of how faces are processed in humans and how perceptual expertise shapes these processing strategies.

7.4 Number of repetitions in the identification task

Before each test-block participants had to perform a haptic identification task on the initially learned faces in which they had to name each presented face mask after haptic exploration. Each repetition consisted of the subsequent presentation of the three initially learned faces in random order. To pass the haptic identification task and move on to the next test-block of the old/new

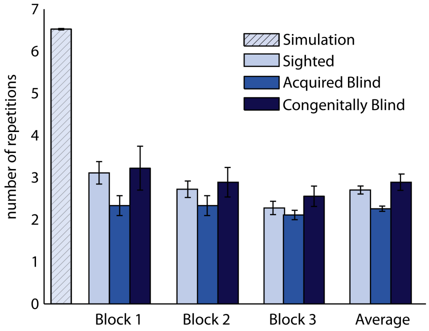

Figure 7.3: Plot showing number of repetitions needed in the identification task until each face was identified correctly twice for the sighted, acquired and congenitally blind groups. This data was taken as a measure of how well the three faces could be discriminated by the participants. Error bars represent ± 1 SEM.

recognition task, each face had to be identified correctly twice (i.e. there was a minimum of 2 repetitions). For each participant, we recorded how many repetitions were needed until each face was identified correctly twice. While the old/new recognition task was very difficult and possibly susceptible to factors such as the use of cognitive strategies, the number of repetitions in the identification task directly assessed how well participants were able to discriminate the faces, i.e., how well they had learned them (Results are shown in *Figure 7.3*). In addition, we calculated the number of repetitions needed to identify each face correctly twice by chance. Using a random observer yielded a chance performance of 6.5 repetitions.

7.4.1 Results

As a first analysis step, we used two-tailed t-tests comparing the simulation data to performance in Block 1 to assess whether participants performed better than chance in the identification task. We found that performance was highly significantly better than chance in each group (Sim vs (1) sighted: $t_{10017} = 5.81, p < 0.001$, (2) acquired blind: $t_{10007} = 5.05, p < 0.001$, (3) congenitally blind: $t_{10007} = 3.98, p < 0.001$), indicating that all three groups learned the faces correctly and were able to discriminate between them.

As a second step, we performed a 3x3 factorial ANOVA with the within-subjects factor being test-Blocks (1-3) and the between-subjects factor being Group (congenitally blind, acquired blind, sighted). We found significant main effects of Block ($F_{2,99} = 3.23, p < 0.05$) and Group ($F_{2,99} = 3.48, p < 0.05$) with no significant interaction between them ($F_{4,99} = 0.35, p = 0.84$).

Thirdly, we performed a post-hoc analysis for effects of the factors Block and Group separately, using two-sided t-tests with Bonferroni correction for multiple comparisons. Here we found that the number of repetitions needed in the identification task was significantly lower in Block 3 than Block 1 ($p < 0.05$) with no significant difference between Blocks 1 and 2 ($p > 0.05$) and 2 and 3 ($p > 0.05$), respectively. Furthermore, acquired blind participants needed significantly fewer trials (repetitions) than congenitally blind ($p < 0.05$), with no significant difference between sighted and acquired blind ($p > 0.05$), as well as sighted and congenitally blind participants ($p > 0.05$).

Finally, we used a One-Way-ANOVA to evaluate performance before the last block (3). We failed to find a significant main effect of Group ($F_{2,33} = 1.19, p = 0.32$) indicating that all three groups performed equally well in the identification task before the third test-block (for which we evaluated performance in the old/new recognition task), i.e., had learned the faces equally well.

7.4.2 Discussion

Taken together, these results indicate that while the sighted and congenitally blind had slightly more difficulties at discriminating between the three learned faces in the beginning of the experiment, they (1) had still learned the faces correctly and were able to discriminate them at levels well above chance, and (2) got significantly better during the experiment and performed as well as the acquired blind before the last block. We, therefore, suggest that the differences in performance in the identification task can be attributed to the novelty of the task of haptic face discrimination and varying levels of expertise in the three participant groups. Sighted participants have little to no haptic expertise, while the congenitally blind have little visual experience with faces which might be beneficial in using face information in any modality. The acquired blind, on the other hand, have experience in both, haptically recognizing objects as well as visually recognizing faces. They might, therefore, have less to no difficulty to get acquainted to the novel and unusual task of haptic face recognition. Once both the sighted and congenitally blind have become sufficiently acquainted with the task, they perform as well as the acquired blind in the identification task.

7.5 Response times

It is important to note that long response times as observed in this haptic recognition task are difficult to interpret and only allow for limited conclusions. The data here is thus reported only for completeness' sake.

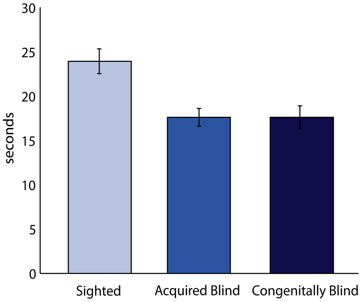

Figure 7.4: Plot comparing response times for haptic face recognition in the sighted, acquired blind and congenitally blind, measured in seconds. Error bars represent ± 1 SEM.

7.5.1 Results

We analyzed response times for the three participant groups using a One-Way ANOVA followed by Bonferroni post-hoc tests (*Figure 7.4*). We found a significant main effect of Group ($F_{2,2008} = 71.73, p < 0.001$) for which post-hoc tests revealed that congenitally and acquired blind were significantly faster than sighted participants (congenitally blind vs. sighted and acquired blind vs. sighted: $p < 0.05$ congenitally vs. acquired blind: $p > 0.05$). (In addition, we looked at the response times across test-blocks and found that congenitally and acquired blind participants were consistently faster than sighted in haptic face recognition, although the group of sighted participants became significantly faster across test-blocks.)

7.5.2 Discussion

This response time advantage in both blind groups over the sighted might reflect haptic expertise in the blind. Blind people

are usually extensively trained to use their sense of touch to encode the shape of Braille letters, familiar objects, etc. and might therefore be faster in haptic tasks (for example, Heller (1989)). Sighted participants, on the other hand, became faster as they became acquainted with the unusual task.
In contrast, the observed differences in response times could also reflect the use of different cognitive strategies, for example, the use of visual imagery in the sighted during haptic exploration which, in turn, could have resulted in longer exploration times. Debriefing questionnaires that were filled out by the participants right after the experiment, however, revealed that both sighted and acquired blind participants frequently tried to imagine what the faces looked like during haptic exploration. Since the acquired blind were significantly faster than the sighted throughout the experiment this explanation is less likely.

7.6 General Discussion

Our experiment provides the first demonstration of a pronounced advantage for acquired blind (visual and haptic expertise) and sighted (visual but no haptic expertise) participants over congenitally blind (no visual but haptic expertise) in haptic face recognition. The fact that a lack of visual experience affects haptic recognition is consistent with the cross-modal calibration hypothesis (Gori et al. (2008, 2009)). It states that information from the more robust sense (for each specific task) is used to calibrate the other senses during development. Different tasks within each modality are, therefore, differently affected by impaired cross-modal calibration due to a lack of good vision at birth.

Are our findings specific for face recognition? Heller et al. (2003) showed that visual experience is not necessary for the development of perceptual selectivity using tangible versions of the embedded figures test. Here, congenitally blind people performed as well as blindfolded participants. Acquired blind participants, however, showed significantly higher levels of performance, suggesting that a combination of pictorial experience and haptic skill can facilitate perceptual selectivity in touch. Conversely, we found that congenitally blind participants performed significantly *worse* in haptic face recognition than acquired blind and sighted participants, stressing the importance of prior visual experience with faces. In other words, a lack of relevant visual experience cannot be compensated for or improved by purely perceptual haptic expertise as the results for the blind groups show. This finding is, therefore, consistent with the hypothesis that efficient face processing relies on specific strategies requiring visual expertise, which, in turn, requires visual input to the right hemisphere during infancy (LeGrand et al. (2003)). LeGrand et al. showed that early deprivation of visual input to the right hemisphere severely impairs the development of expert face processing, indicating that the neural circuitry responsible for adults' face expertise is not pre-specified but requires early visual experience. Along these lines, a recent study showed that visual brain areas specialized in face processing, once established through visual experience, can be recruited for comparable haptic tasks in the blind (Goyal et al. (2006)). In this study, however, no behaviorally relevant task, e.g., a face recognition task, was used. While performance levels for the acquired blind group in our study might as well suggest the use of specialized face processing areas for haptic face recognition, further studies will be necessary to investigate whether the reported activation in visual face areas might actually be related to the measured benefit in behavioral performance.

Taken together, the results reported here provide behavioral evidence for the crucial role of visual input for the development of efficient face processing capabilities and thus contribute directly to the central question of how faces are processed in humans and how perceptual expertise shapes face processing strategies.

Chapter 8

General Discussion

The experiments described in this thesis examined various aspects of visual and haptic face recognition. In this section we summarize the findings reported in the four studies which comprise this thesis (a more detailed discussion of each study can be found in the respective chapters) and discuss future avenues of (haptic) face recognition research.

8.1 Cross-modal transfer in visual and haptic face recognition

The aim of study 1 was to investigate whether visual and haptic modalities encode similar information about faces to allow for efficient cross-modal transfer. In summary, we first showed that our 3D face stimuli can be learned and recognized using touch alone (Experiment 2). Above-chance cross-modal recognition, however, was only possible when haptic memory was refreshed, suggesting a memory effect in cross-modal face recognition (Experiment 3). Here, cross-modal haptics-to-vision recognition was as accurate as within-modal haptic recognition. In Experiment 4, we found a clear advantage for within-modal visual recognition. However, in contrast to haptics-to-vision transfer, we found a cost in transfer from vision to haptics, resulting in cross-modal recognition accuracy to be equal in both vision-to-haptics and haptics-to-vision conditions. This suggests that information transfer across modalities might be asymmetric and limited by haptic face processing. Indeed, while we have shown that the haptic modality is capable of fine-level discriminations, as participants could easily discriminate between the learned faces, the question remained whether it is also capable of configural processing, i.e., an important hallmark of visual expert face processing. Since the haptic modality is limited to serial exploration of an object (involves a feature-by-feature analysis) it might also rather process this featural information instead of integrating it into a more global representation. We, therefore, suggested that the observed asymmetric transfer may be due to differences in visual and haptic information processing which might in turn be introduced by qualitative differences in information encoding in vision and haptics (holistic in vision versus featural in haptics).

8.2 Serial exploration of faces: Comparing vision and touch

In study 2, we investigated the effects of modality-specific encoding differences in visual and haptic face recognition in terms of recognition accuracy as well as information processing strategies. This was achieved by using a gaze-restricted display to constrain the visual system to sequential, self-directed exploration, promoting serial encoding in vision in much the same way than the haptic system encodes objects. We found that face recognition was equally disrupted using gaze-restricted vision and haptics as compared to unrestricted vision. First, face recognition performance across the visual and haptic sense was equated due to a decrease in gaze-restricted visual face recognition accuracy as compared to unrestricted visual face recognition (Experiments 1,2). Secondly, we also observed modality-independent higher working memory demands for serial encoding of faces (Experiment 2). Finally, we found a strong face inversion effect for unrestricted visual but none for haptic or gaze-restricted face recognition (Experiment 3).

Taken together, these results indicate that face processing is impeded by serial encoding even when participants have control over the information that they view through the aperture. Specifically, participants in the haptic and gaze-restricted modalities seemed unable to use expert face processing strategies displayed in unrestricted vision, i.e., configural processing is enabled by fast and parallel information encoding and constrained by restricted, serial encoding. This finding raised the question whether modalities that rely on serial encoding of information actually allow for expert face processing. On the one hand, the inability to use configural processing could be due to intrinsic constraints, such as sensory limitations or working memory capacity. Since serial information of an object has to be integrated over space and time, participants might not be able to precisely measure second order relations between features. Moreover, the memory load for complex stimuli such as faces might simply be too immense to allow for accurate integration of spatial information across the whole. On the other hand, it could merely be due to a lack of experience in a modality that relies on serial encoding of information, i.e., a lack of perceptual expertise. If the latter is the case, participants might be able to develop strategies to compensate for processing differences introduced by serial encoding.

Despite any modality-specific differences though, study 1 has shown that face information can still be shared across the visual and haptic modalities. Given that (1) haptic processing seems to rely on featural shape processing, (2) cross-modal recognition, independent of direction of transfer, is limited to levels of haptic face recognition, and (3) we previously failed to find an inversion effect for visual recognition of haptically learned faces (Dopjans et al. (2008)), our results suggest that the nature of this shared information is *featural*. This is not to say, however, that face information could not be shared at a different level if, for example, acquisition of haptic face expertise involved configural processing.

8.3 Learning to recognize faces through serial exploration

Study 3 examined whether participants can learn to efficiently recognize faces that are serially encoded. For this, we trained participants in gaze-restricted face recognition. Here, we found several hallmarks of expertise after short training: (1) a significant improvement in terms of response times (to levels of unrestricted visual face recognition) and accuracy (exceeding levels of unrestricted visual face recognition), (2) an inversion effect for response times for 'experts' but not novices, (3) a shift in fixation patterns from the bottom to the top half of the face, (4) generalization of expertise to a novel set of stimuli, and (5) that expertise lasted at least one week. (For a more detailed analysis, e.g., of the fixation patterns see chapter 6.4.) Taken together, study 3 shows how practice with a previously novel way of perceiving faces, i.e., through serial encoding, can lead to some of the recognition effects typically associated with unrestricted visual face recognition, indicating that -at least for vision- serial encoding of information might allow for expert face processing. The question remains, however, whether haptic training will display the same pattern of expertise acquisition. Participants required surprisingly little training in the gaze-restricted modality-3 hours at most- to reach accuracy levels exceeding those of unrestricted visual face recognition. While gaze-restricted vision is certainly a novel way of perceiving faces it is still based on visual processes. Since humans are undisputedly visual animals, used to building representations from visual input, and are experts at unrestricted visual face recognition, both factors might have facilitated the development of compensatory strategies and, consequently, the acquisition of expertise. If this was the case and inasmuch as we have little to no perceptual haptic expertise the question arises whether haptic training will display the same pattern of face expertise acquisition. Studying the role of perceptual haptic expertise, whether and how it might affect haptic face recognition, might, therefore, prove highly informative.

8.4 Visual experience is necessary for efficient haptic face recognition

In study 4, we used a different approach to the question how perceptual expertise shapes efficient face processing strategies by studying face recognition in the blind. More specifically, we investigated haptic face recognition in the sighted (visual but no haptic expertise), congenitally blind (no visual but haptic expertise), and acquired blind (visual and haptic expertise) to assess the roles of perceptual haptic expertise and visual input for the development of efficient face processing capabilities. In summary, we found a response time advantage in both blind groups over the sighted, reflecting haptic expertise in the blind, and a pronounced advantage in terms of accuracy for acquired blind and sighted participants over congenitally blind in haptic face recognition. These findings indicate that a lack of relevant visual experience cannot be compensated for or improved by purely perceptual haptic expertise as the results for the blind groups show.

8.5 Summary and future directions

This thesis investigated the mechanisms underlying (haptic) face processing as a tool for studying (1) face processing itself, (2) the role of perceptual expertise, and (3) the differences and commonalities in information processing in the visual and haptic modalities. Taken together, the studies presented in this thesis demonstrate how modality-specific differences in information acquisition affect processing strategies in high-level cognitive tasks, such as face recognition, and how those differences can be compensated for by perceptual expertise. More specifically, the way in which information is processed in the brain depends strongly on the way this information is encoded, i.e., visual face information is not processed configurally *per se* but relies on fast, parallel information encoding. The limitations imposed on face recognition by serial encoding, however, are not insurmountable but can be overcome with the acquisition of perceptual expertise. Here, one of the most compelling examples is that -posttraining-recognition accuracy for serially encoded faces exceeded levels of holistically encoded faces- even with novel stimuli. On the other hand, evidence was found that an inherent lack of relevant visual experience cannot be improved by purely perceptual haptic expertise.

To further test how perceptual expertise shapes face processing in the visual and haptic modalities the work presented in this thesis could be extended in several ways. The following paragraphs outline studies on three topics: (1) tests of whether the haptic modality allows for expert face processing and its relation to vision, (2) tests whether haptic face recognition deficits in the blind are face-specific, and (3) a search for the neural underpinnings of multisensory face expertise.

First, while we have shown for vision that- given enough training- serial encoding of information might allow for expert face processing, a respective haptic training study would provide useful insight into the limitations and capabilities of the haptic modality. Moreover, similar to gaze-restricted exploration, the dynamic study of manual exploration of faces using motion capture might prove highly informative. Especially compared to eye-tracking data from visual exploration of faces, such scan patterns will show how complex face information is acquired in the visual and haptic modalities and which the salient features are in each modality.

A second series of experiments should further investigate the cause of the haptic face recognition deficit observed in congenitally blind individuals. Since we only tested the blind in haptic *face* recognition, it is necessary to determine whether the observed deficit is due to a lack of visual experience specifically with faces or with objects in general. A broader range of objects classes need to be tested, including every-day objects to test the importance of stimulus experience and different exemplars of an object class to determine fine-level discrimination ability. If the observed deficit were indeed face-specific and congenitally blind individuals were not able to develop expert face processing strategies, this would suggest that faces and objects are processed by independent mechanisms in the haptic modality. Moreover, given that expertise leading to face-like processing can occur at any age (Diamond and Carey (1986); Gauthier and Tarr (1997)), it would be of particular interest to train congenitally blind in haptic face recognition to test whether they can develop expert strategies. The duration and outcome of training compared to sighted individuals would further be useful in order to determine the role of visual expertise with faces on haptic face processing, i.e., whether haptic face processing can benefit from existing visual face expertise mechanisms. Furthermore, these studies will undoubtedly shed more light on the 'domain-specificity' versus 'expertise' debate.

It is, indeed, a long-standing and ongoing debate whether the specialized cognitive and neural mechanisms for face recognition constitute a domain-specific modular system dedicated to face recognition (referred to as 'domain-specificity', e.g., Kanwisher et al. (1997); Kanwisher (2000); McKone et al. (2007)) or the endpoint on the continuum of perceptual category learning (referred to as 'expertise hypothesis', e.g., Gauthier and Tarr (1997); Bukach et al. (2006)). Our finding that expert face processing is impaired for serially encoded faces and that these encoding-specific limitations can be compensated for by perceptual expertise could be interpreted as evidence for the 'expertise hypothesis'. On the other hand it could be interpreted as tuning other sensory modalities within face-specific mechanisms once compensatory strategies for serial encoding have been learned. A third series of experiments on the physiological underpinnings of multisensory face and object expertise will further investigate these questions. An extensive amount of literature on physiological processing has argued for and against 'domain-specificity' using brain imaging (e.g., Grill-Spector et al. (2004); Rhodes et al. (2004); Gauthier et al. (1999b); Bukach et al. (2006)) and monkey single-unit recording (e.g., Deaner et al. (2005); Tsao et al. (2006)). Here, the site of primary interest is the fusiform face area (FFA), for several reasons, due to its much stronger response to faces than objects, its location being consistent with the critical lesion site for loss of face recognition ability and its sensitivity to differences between individual upright and inverted faces and holistic processing. Several studies have investigated the neural correlates of haptic face recognition (James et al. (2006); Kilgour et al. (2005); Kitada et al. (2009)). In addition, we recently used fMRI to study haptic recognition of our face masks and LEGO houses (Azulay et al. (2009, 2010)). We found distinctive patterns of activation for the two stimulus classes. For haptic face recognition, marked activation was observed in the right FFA and bilateral face-LOC; both regions were also activated for a visual face localizer. Since neural mechanisms underlying visual and haptic face recognition are similar, the next step should be to investigate whether neural mechanisms dedicated to face perception still exist without visual experience. An fMRI study with congenitally blind individuals may be able to answer this question. For this, brain areas activated during visual and haptic face recognition in sighted could be localized and compared to areas activated during haptic face recognition in the congenitally blind. Second, studying brain activation in the acquired blind might reveal whether the measured behavioral benefit is reflected in the use of specialized visual processing areas. Finally, using a classic adaptation paradigm for identity adaptation across modalities in sighted participants could be used to study the neural correlates of cross-modal transfer in visual and haptic face recognition. For this, participants would first be presented with a face visually followed by a second, haptic presentation (and vice versa) of the same face. If an adaptation effect for the second, cross-modal presentation were found, this would provide strong evidence for a shared neural representation for faces.

These extensions would help to further our understanding of how complex information is processed in multiple sensory systems and how they might benefit from each other. For example, we have shown clear cross-modal information transfer and quick learning of expertise with a novel exploration strategy. These results might have important implications for people with low visual acuity such as macular degeneration as one could envision the development of training schemes for re-acquiring perceptual expertise using different exploratory strategies and/or sensory modalities.

References

Allen, S. and Brooks, L. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General*, 120:3–19.

Alvarez, G. and Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, 15(2):106–111.

Amedi, A., Jacobson, G., Hendler, T., Malach, R., and Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cerebral Cortex*, 12:1202–1212.

Azulay, H., Dopjans, L., and Amedi, A. (2009). Patterns of activity during haptic face recognition– an fmri study. *Journal of Molecular Neuroscience*, 39 (Supplement 1):S11.

Azulay, H., Dopjans, L., Bubic, A., Wallraven, C., and Amedi, A. (2010). Neural correlates of humans performing tactile encoding and retrieval of 3d face masks. *FENS Abstracts*, 5:114.1.

Baron, M. (2008). Haptic classification of facial expressions of emotion on 3d facemasks and the absence of a haptic face-inversion effect. *Unpublished BSc (honors) thesis, Queen's University.*

Barton, J. J. S., Radcliffe, N., Cherkasova, M. V., Edelman, J., and Intriligator, J. M. (2006). Information processing during face recognition: the effects of familiarity, inversion, and morphing on scanning fixations. *Perception*, 35(8):1089–1105.

Behrmann, M. and Ewell, C. (2003). Expertise in tactile pattern recognition. *Psychological Science*, 14:480–486.

Benson, P. and Perrett, D. (1991). Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images. *European Journal of Cognitive Psychology*, 3(1):105–135.

Bliss, I. and Hämäläinen, H. (2005). Different working memory capacity in normal young adults for visual and tactile letter recognition task. *Scandinavian Journal of Psychology*, 46:247–251.

Boster, J. and Johnson, J. (1989). Form or function: A comparison of expert and novice judgements of similarity among fish. *American Anthropologist*, 91:866–889.

Boutsen, L. and Humphreys, G. (2003). The effect of inversion on the encoding of normal and 'thatcherized' faces. *Quaterly Journal of Experimental Psychology*, 56A(6):955–975.

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision*, 10:433–436.

Brennan, S. (1985). The caricature generator. *Leonardo*, 18:170–178.

Brooks, L. R., Norman, G. R., and Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General*, 120(3):278–287.

Bruce, V., Hanna, E., Dench, N., Healey, P., and Burton, M. (1992). The importance of 'mass' in line drawings of faces. *Applied Cognitive Psychology*, 6:619–628.

Bruce, V. and Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77 (Pt 3):305–327.

Bruce, V. and Young, A. (1998). *In the eye of the beholder: The science of face perception.* Oxford University Press.

Bukach, C., Bub, D., Gauthier, I., and Tarr, M. (2006). Perceptual expertise effects are not all or none: Spatially limited perceptual expertise for faces in a case of prosopagnosia. *Journal of Cognitive Neuroscience*, 18(1):48–63.

Burton, A. and Vokey, J. (1998). The face-space typicality paradox: Understanding the face-space metaphor. *The Quaterly Journal of Experimental Psychology*, 51A(3):475–483.

Burton, A., Wilson, S., Cowan, M., and Bruce, V. (1999). Face recognition in poor-quality video. *Psychological Science*, 10:243–248.

Busey, T. A. and Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision Research*, 45(4):431–448.

Bushnell, E. and Baxt, C. (1999). Children's haptic and cross-modal recognition with familiar and unfamiliar objects. *Journal of Experimental Psychology: Human Perception and Performance*, 25:1867–1881.

Carey, S. and Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science*, 195(4275):312–314.

Carey, S. and Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition*, 1:253–274.

Casey, S. J. and 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.

Casey, S. J. and Newell, F. N. (2007). Are representations of unfamiliar faces independent of encoding modality? *Neuropsychologia*, 45(3):506–513.

Chance, J. E., Turner, A. L., and Goldstein, A. G. (1982). Development of differential recognition for own- and other-race faces. *Journal of Psychology*, 112(1st Half):29–37.

Chase, W. and Simon, H. (1973). Perception in chess. *Cognitive Psychology*, 4:55–Ű81.

Collishaw, S. and Hole, G. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception*, 29(8):893–909.

Cooke, T., Jäkel, F., Wallraven, C., and Bülthoff, H. (2007). Multimodal similarity and categorization of novel, three-dimensional objects. *Neuropsychologia*, 45(3):484–495.

Cooke, T., Steinke, F., Wallraven, C., and Bülthoff, H. (2005). A similarity-based approach to perceptual feature validation. *In APGV 2005 - Symposium on Applied Perception in raphics and Visualization*, page ACM SIGGRAPH.

Cooke, T., Wallraven, C., and Bülthoff, H. (2010). Multidimensional scaling analysis of haptic exploratory procedures. *ACM Transactions on Applied Perception*, 7:1–17.

Cunningham, D. W., Billock, V. A., and Tsou, B. H. (2001). Sensorimotor adaptation to violations of temporal contiguity. *Psychological Science*, 12(6):532–535.

Curby, K. C. and Rossion, B. (2010). *Perceptual expertise: bridging brain activity and behavior*, chapter Competition between face and nonface domains of expertise, pages 245–270. Oxford University Press.

Dahl, C. D., Logothetis, N. K., and Hoffman, K. L. (2007). Individuation and holistic processing of faces in rhesus monkeys. *Proceedings of the Royal Society - Biological Sciences*, 274(1622):2069–2076.

Dahl, C. D., Wallraven, C., Bülthoff, H. H., and Logothetis, N. K. (2009). Humans and macaques employ similar face-processing strategies. *Current Biology*, 19(6):509–513.

Damasio, A., Damasio, H., and Hoesen, G. V. (1982). Prosopagnosia: Anatomical basis and behavioral mechanisms. *Neurology*, 32:331–341.

Damasio, A. R. (1990). Category-related recognition defects as a clue to the neural substrates of knowledge. *Trends in Neurosciences*, 13(3):95–98.

Davidoff, J., Matthews, W., and F.Newcombe (1986). *Aspects of face processing*, pages 279–290. Dordrecht: Martinus Nijhoff.

Davies, G., Ellis, H., and Shepherd, J. (1977). Cue saliency in faces as assessed by the 'photofit' technique. *Perception*, 6(3):263–269.

Davies, G., Ellis, H., and Shepherd, J. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology*, 63:180–187.

Deaner, R. O., Khera, A. V., and Platt, M. L. (2005). Monkeys pay per view: adaptive valuation of social images by rhesus macaques. *Current Biology*, 15(6):543–548.

Deibert, E., Kraut, M., Kremen, S., and Hart, J. (1999). Neural pathways in tactile object recognition. *Neurology*, 52(7):1413–1417.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences of the United States of America*, 93(24):13494–13499.

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

Dopjans, L., Bülthoff, H., and Wallraven, C. (2010). Serial exploration of faces: Comparing vision and touch. *Submitted.*

Dopjans, L., Wallraven, C., and Bülthoff, H. (2008). A visual but no haptic face inversion effect indicates modality-specific processing differences. *Perception*, 37(ECVP Abstract Supplement):5.

Dopjans, L., Wallraven, C., and Bülthoff, H. (2009). Cross-modal transfer in visual and haptic face recognition. *IEEE Transaction on Haptics*, 2:236–240.

Dufour, V., Pascalis, O., and Petit, O. (2006). Face processing limitation to own species in primates: a comparative study in brown capuchins, tonkean macaques and humans. *Behavioural Processes*, 73(1):107–113.

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

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, 24(6):581–604.

Ericsson, K., Charness, N., Feltovich, P., and Hoffman, R. (2006). *Cambridge handbook of expertise and expert performance.* Cambridge: Cambridge University Press.

Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433.

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

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

Farah, M., Wilson, K., Drain, M., and Tanaka, J. (1998). What is 'special' about face perception? *Psychological Review*, 105(3):482–498.

Farah, M. J., Wilson, K. D., Drain, H. M., and Tanaka, J. R. (1995b). The inverted face inversion effect in prosopagnosia: evidence for mandatory, face-specific perceptual mechanisms. *Vision Research*, 35(14):2089–2093.

Farroni, T., Csibra, G., Simion, F., and Johnson, M. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences of the United States of America*, 99:9602–9605.

Fraser, I., Craig, G., and Parker, D. (1990). Reaction time measures of feature saliency in schematic faces. *Perception*, 19(5):661–673.

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

Gauthier, I., Behrmann, M., and Tarr, M. J. (1999a). Can face recognition really be dissociated from object recognition? *Journal of Cognitive Neuroscience*, 11(4):349–370.

Gauthier, I. and Curby, K. (2005). A perceptual traffic-jam on highway n170: Interference between face and car expertise. *Current Directions in Psychological Science*, 4:30–33.

Gauthier, I., Curran, T., Curby, K. M., and Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience*, 6(4):428–432.

Gauthier, I., Skudlarski, P., Gore, J. C., and Anderson, A. W. (2000a). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3(2):191–197.

Gauthier, I. and Tarr, M. (1997). Becoming a 'greeble' expert: Exploring the face recognition mechanism. *Vision Research*, 37:1673–1682.

Gauthier, I., Tarr, M., Moylan, J., Skudlarski, P., Gore, J., and Anderson, A. (2000b). The fusiform 'face area' is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience*, 12:495–504.

Gauthier, I. and Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. *Journal of Experimental Psychology: Human Learning and Memory*, 28(2):432–446.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., and Gore, J. C. (1999b). Activation of the middle fusiform ’face area’ increases with expertise in recognizing novel objects. *Nature Neuroscience*, 2(6):568–573.

Gauthier, I., Williams, P., Tarr, M. J., and Tanaka, J. (1998). Training ’greeble’ experts: a framework for studying expert object recognition processes. *Vision Research*, 38(15-16):2401–2428.

Gelder, B. D. and Rouw, R. (2000). Paradoxical configuration effects for faces and objects in prosopagnosia. *Neuropsychologia*, 38:1271–1279.

Ghahramani, Z., Wolpert, D. M., and Jordan, M. I. (1996). Generalization to local remappings of the visuomotor coordinate transformation. *Journal of Neuroscience*, 16(21):7085–7096.

Giese, M. and Leopold, D. (2004). Physiologically inspired neural model for the encoding of face spaces. *Neurocomputing*, 65-66:93–101.

Goldstein, A. and Mackenberg, E. (1966). Recognition of human faces from isolated facial features: A developmental study. *Psychonomic Science*, 6:149–150.

Gori, M., DelViva, M., Sandini, G., and Burr, D. (2008). Young children do not integrate visual and haptic information. *Current Biology*, 18:694–698.

Gori, M., Sandini, G., Christina, M., and Burr, D. (2009). Haptic orientation discrimination is severely impaired in blind and low-vision children. *Perception 38 ECVP Abstract Supplement*, page 161.

Goyal, M. S., Hansen, P. J., and Blakemore, C. B. (2006). Tactile perception recruits functionally related visual areas in the late-blind. *Neuroreport*, 17(13):1381–1384.

Grill-Spector, K., Knouf, N., and Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, 7:555–562.

Guo, K., Robertson, R. G., Mahmoodi, S., Tadmor, Y., and Young, M. P. (2003). How do monkeys view faces?–a study of eye movements. *Experimental Brain Research*, 150(3):363–374.

Hadjikhani, N. and Roland, P. E. (1998). Cross-modal transfer of information between the tactile and the visual representations in the human brain: A positron emission tomographic study. *Journal of Neuroscience*, 18(3):1072–1084.

Haig, N. (1986). Exploring recognition with interchanged facial features. *Perception*, 15:235–247.

Halberstadt, J., Gladstone, R., and Levine, G. (2003). Featural processing in face references. *Journal of Experimental Social Psychology*, 39:270–278.

Hancock, P., Burton, A., and Bruce, V. (1996). Face processing: Human perception and principal components analysis. *Memory & Cognition*, 24:26–40.

Harmon, L. (1973). The recognition of faces. *Scientific American*, 229(5):70–83.

Harmon, L. and Julesz, B. (1973). Masking in visual recognition: Effects of two-dimensional noise. *Science*, 180:1194–1197.

Hay, D. and Cox, R. (2000). Developmental changes in the recognition of faces and facial features. *Infant and Child Development*, 9:199–212.

Heathcote, A., Brown, S., and Mewhort, D. (2000). Repealing the power law: The case for an exponential law of practice. *Psychonomic Bulletin and Review*, 7:185–207.

Heller, M. A. (1989). Picture and pattern perception in the sighted and the blind: the advantage of the late blind. *Perception*, 18(3):379–389.

Heller, M. A., Wilson, K., Steffen, H., Yoneyama, K., and Brackett, D. D. (2003). Superior haptic perceptual selectivity in late-blind and very-low-vision subjects. *Perception*, 32(4):499–511.

Hole, G. (1994). Configurational factors in the perception of unfamiliar faces. *Perception*, 23(1):65–74.

Ikeda, M. and Uchikawa, K. (1978). Integrating time for visual pattern perception and a comparison with the tactile mode. *Vision Research*, 18:1565–1571.

James, T., Servos, P., Kilgour, A., Huh, E., and Lederman, S. (2006). The influence of familiarity on brain activation during haptic exploration of 3d facemasks. *Neuroscience Letters*, 397:269–273.

James, T. W., Humphrey, G. K., Gati, J. S., Servos, P., Menon, R. S., and Goodale, M. A. (2002). Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia*, 40(10):1706–1714.

Janik, S. W., Wellens, A. R., Goldberg, M. L., and Dell'Osso, L. F. (1978). Eyes as the center of focus in the visual examination of human faces. *Perceptual & Motor Skills*, 47(3 Pt 1):857–858.

Johansen, M. K. and Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cognitive Psychology*, 45(4):482–553.

Johnson, K. and Mervis, C. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General*, 126:248–277.

Johnson, M., Dziurawiec, S., Ellis, H., and Morton, J. (1991). Newborns preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40(1-2):1–19.

Johnson, M. and Morton, J. (1991). *Biology and cognitive development.* Blackwell, Oxford.

Jolicoeur, P., Gluck, M. A., and Kosslyn, S. M. (1984). Pictures and names: making the connection. *Cognitive Psychology*, 16(2):243–275.

Jones, B. (1981). *Intersensory perception and sensory integration*, chapter The developmental significance of cross-modal matching, pages 109–136. New York: Plenum.

Joseph, R. and Tanaka, J. (2003). Holistic and part-based face recognition in children with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 44:529–542.

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

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

Kanwisher, N., Tong, F., and Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition*, 68:1–11.

Kanwisher, N. and Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B*, 361:2109–2128.

Keating, C. and Keating, E. (1982). Visual scan patterns of rhesus monkeys viewing faces. *Perception*, 11(2):211–219.

Kilgour, A. R., de Gelder, B., and Lederman, S. J. (2004). Haptic face recognition and prosopagnosia. *Neuropsychologia*, 42(6):707–712.

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

Kilgour, A. R. and Lederman, S. J. (2002). Face recognition by hand. *Perception & Psychophysics*, 64(3):339–352.

Kilgour, A. R. and Lederman, S. J. (2006). A haptic face-inversion effect. *Perception*, 35(7):921–931.

Kim, N. S. and Ahn, W. K. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General*, 131(4):451–476.

Kitada, R., Johnsrude, I., Kochiyama, T., and Lederman, S. (2009). Functional specialization and convergence in the occipitotemporal cortex supporting haptic and visual identification of human faces and body parts: an fmri study. *Journal of Cognitive Neuroscience*, 21(10):2027–2045.

Klatzky, R. and Lederman, S. (1993). Toward a computational model of constraint-driven exploration and haptic object identification. *Perception*, 22:597–621.

Klatzky, R. and Lederman, S. (1995). Identifying objects from a haptic glance. *Perception & Psychophysics*, 57:1111–1123.

Klatzky, R. and Lederman, S. (2000). *Toucher pour connacetre: Psychologie cognitive de la perception tactile manuelle*, chapter L'identification haptique des objets significatifs, pages 109–128. Paris: Presses Universitaires de France.

Klatzky, R., Lederman, S., and Metzger, V. (1985). Identifying objects by touch: An 'expert system'. *Perception & Psychophysics*, 37:299–302.

Klatzky, R., Lederman, S., and Reed, C. (1987). ThereŠs more to touch than meets the eye: The salience of object attributes for haptics with and without vision. *Journal of Experimental Psychology: General*, 116(4):356–369.

Klin, A., Jones, W., Schultz, R., Volkmar, F., and Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, 59(9):809–816.

Knappmeyer, B., Thornton, I., and Bülthoff, H. (2003). The use of facial motion and facial form during the processing of identity. *Vision Research*, 43:1921–1936.

Knecht, S., Kunesch, E., and Schnitzler, A. (1996). Parallel and serial processing of haptic information in man: Effects of parietal lesions on sensorimotor hand function. *Neuropsychologia*, 34:669–687.

Lakatos, S. and Marks, L. (1999). Haptic form perception: relative salience of local and global features. *Perception & Psychophysics*, 61(5):895–908.

Lander, K. and Chuang, L. (2005). Why are moving faces easier to recognize? *Visual Cognition*, 12:429–442.

Langdell, T. (1978). Recognition of faces: an approach to the study of autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 19(3):255–268.

Leder, H. and Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *The Quaterly Journal of Experimental Psychology*, 53A:513–536.

Leder, H., Candrian, G., Huber, O., and Bruce, V. (2001). Configural features in the context of upright and inverted faces. *Perception*, 30(1):73–83.

Lederman, S. and Klatzky, R. (1987). Hand movements: A window into haptic object recognition. *Cognitive Psychology*, 19:342–368.

Lederman, S., Summers, C., and Klatzky, R. (1996). Cognitive salience of haptic object properties: role of modality-encoding bias. *Perception*, 25(8):983–998.

Lederman, S. J. and Klatzky, R. L. (1990). Haptic classification of common objects: knowledge-driven exploration. *Cognitive Psychology*, 22(4):421–459.

Lederman, S. J., Klatzky, R. L., and Kitada, R. (2010). Haptic face processing and its relation to vision. In Kaiser, J. and Naumer, M. J., editors, *Multisensory Object Perception in the Primate Brain*, pages 273–300. Springer New York. 10.1007/978-1-4419-5615-$6_1$5.

LeGrand, R., Mondloch, C. J., Maurer, D., and Brent, H. P. (2003). Expert face processing requires visual input to the right hemisphere during infancy. *Nature Neuroscience*, 6(10)(10):1108–1112.

Levine, S., Banich, M., and Koch-Weser, M. (1988). Face recognition: a general or specific right hemisphere capacity? *Brain and Cognition*, 8(3):303–325.

Lewis, M. (2003). Thatcher's children: development and the thatcher illusion. *Perception*, 32:1415–1421.

Liu, C., Seetzen, H., Burton, A., and Chaudhuri, A. (2003). Face recognition is robust with incongruent image resolution: Relationship to security video images. *Journal of Experimental Psychology: Applied*, 9:33–41.

Loffler, G., Yourganov, G., Wilkinson, F., and Wilson, H. (2005). fmri evidence for the neural representation of faces. *Nature Neuroscience*, 8(10):1386–1390.

Loomis, J. and Lederman., S. (1986). *Handbook of perception and human performances, Vol.2, Cognitive processes and performance*, chapter Tactual perception, pages 1–41. New York: Wiley.

Loomis, J. M. (1981). Tactile pattern perception. *Perception*, 10(1):5–27.

Loomis, J. M., Klatzky, R. L., and Lederman, S. J. (1991). Similarity of tactual and visual picture recognition with limited field of view. *Perception*, 20(2):167–177.

Luck, S. and Vogel, E. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390:279–281.

Lynch, E. B., Coley, J. D., and Medin, D. L. (2000). Tall is typical: central tendency, ideal dimensions, and graded category structure among tree experts and novices. *Memory & Cognition*, 28(1):41–50.

Malpass, R., Lavigueur, H., and Weldon, D. (1973). Verbal and visual training in face recognition. *Perception & Psychophysics*, 13:330–334.

Mangini, M. and Biederman, I. (2004). Making the ineffable explicit: estimating the information employed for face classifications. *Cognitive Science*, 28:209–226.

Marks, D. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, 64:17–24.

Maurer, D., Grand, R., and Mondloch, C. (2002). The many faces of configural processing. *Trends Cogn Sci*, 6(6):255–260.

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

McGregor, T., Klatzky, R., Hamilton, C., and Lederman, S. (2010). Haptic classification of facial identity in 2d displays: Configural versus feature-based processing. *IEEE Transaction on Haptics*, 3:48–55.

McKelvie, S. (1976). The role of eyes and mouth in the memory of a face. *American Journal of Psychology*, 89:311–323.

McKone, E., Kanwisher, N., and Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences*, 11(1):8–15.

Meissner, C. and Brigham, J. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, 7:3–35.

Millar, S. and Al-Attar, Z. (2004). External and body-centered frames of reference in spatial memory: Evidence from touch. *Perception & Psychophysics*, 66:51–59.

Molander, B. and Garvill, J. (1979). The invariance of asymmetric cross-modal transfer effect. *Scandinavian Journal of Psychology*, 20:171–177.

Mondloch, C., LeGrand, R., and Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception*, 31:553–566.

Mondloch, C. J., Geldart, S., Maurer, D., and Grand, R. L. (2003). Developmental changes in face processing skills. *Journal of Experimental Child Psychology*, 86(1):67–84.

Murphy, G. and Wright, J. (1984). Changes in conceptual structure with expertise: Differences between real-world experts and novices. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10:144–155.

Murray, M. M., Foxe, J. J., and Wylie, G. R. (2005). The brain uses single-trial multisensory memories to discriminate without awareness. *Neuroimage*, 27(2):473–478.

Murray, M. M., Michel, C. M., de Peralta, R. G., Ortigue, S., Brunet, D., Andino, S. G., and Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *Neuroimage*, 21(1):125–135.

Myles-Worsley, M., Johnston, W. A., and Simons, M. A. (1988). The influence of expertise on x-ray image processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14(3):553–557.

Nagel, S. K., Carl, C., Kringe, T., Märtin, R., and König, P. (2005). Beyond sensory substitution–learning the sixth sense. *Journal of Neural Engineering*, 2(4):R13–R26.

Newell, A. and Rosenbloom, P. (1981). *Cognitive skills and their acquisition*, chapter Mechanisms of skill acquisition and the law of practice, pages 1–55. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Newell, F. N., Ernst, M. O., Tjan, B. S., and Bülthoff, H. H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological Science*, 12(1):37–42.

Nodine, C. F. and Krupinski, E. A. (1998). Perceptual skill, radiology expertise, and visual test performance with nina and waldo. *Academic Radiology*, 5(9):603–612.

Nodine, C. F., Kundel, H. L., Mello-Thoms, C., Weinstein, S. P., Orel, S. G., Sullivan, D. C., and Conant, E. F. (1999). How experience and training influence mammography expertise. *Academic Radiology*, 6(10):575–585.

Norman, J. F., Clayton, A. M., Norman, H. F., and Crabtree, C. E. (2008). Learning to perceive differences in solid shape through vision and touch. *Perception*, 37(2):185–196.

Ostry, D. J. and Romo, R. (2001). Tactile shape processing. *Neuron*, 31(2):173–174.

O'Toole, A., Roark, D., and Abdi, H. (2002). Recognizing moving faces: a psychological and neural synthesis. *Trends in Cognitive Sciences*, 6:261–266.

O'Toole, A., Vetter, T., and Blanz, V. (1999). Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: an application of three-dimensional morphing. *Vision Research*, 39:3145–3155.

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., and Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition*, 22(2):208–224.

Palmeri, T. (1999). Theories of automaticity and the power law of practice. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25:543–551.

Palmeri, T. and Cottrell, G. (2010). *Perceptual expertise: bridging brain activity and behavior*, chapter Modeling perceptual expertise, pages 197–244. Oxford University Press.

Palmeri, T. J. and Blalock, C. (2000). The role of background knowledge in speeded perceptual categorization. *Cognition*, 77(2):B45–B57.

Pascalis, O. and Bachevalier, J. (1998). Face recognition in primates: A cross species study. *Behavioral Processes*, 43:87–96.

Pascalis, O., de Haan, M., and Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, 296(5571):1321–1323.

Pearson, D. and Robinson, J. (1985). Visual communication at very low data rates. *Proceedings of the IEEE*, 74(4):795–812.

Pelli, D. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10:437–442.

Pellicano, E. and Rhodes, G. (2003). Holistic processing of faces in preschool children and adults. *Psychological Science*, 14(6):618–622.

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W.-H. C., Cohen, L., Guazzelli, M., and Haxby, J. V. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proceedings of the National Academy of Sciences*, 101(15):5658–5663.

Proffitt, J. B., Coley, J. D., and Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26(4):811–828.

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

Redding, G. M., Rossetti, Y., and Wallace, B. (2005). Applications of prism adaptation: a tutorial in theory and method. *Neuroscience and Behavioral Reviews*, 29(3):431–444.

Rhodes, G. (1988). Looking at faces: first-order and second-order features as determinants of facial appearance. *Perception*, 17(1):43–63.

Rhodes, G. (1996). *Superportraits: Caricatures and Recognition.* East Sussex, UK.: Psychology Press Publishers.

Rhodes, G., Byatt, G., Michie, P. T., and Puce, A. (2004). Is the fusiform face area specialized for faces, individuation, or expert individuation? *Journal of Cognitive Neuroscience*, 16(2):189–203.

Rickard, T. (1997). Bending the power law: A cmpl theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General*, 126:288–311.

Roark, D., O'Toole, A., and Abdi, H. (2003). Human recognition of familiar and unfamiliar people in naturalistic video. *IEEE International Workshop on Analysis and Modeling of Faces*, Nice, France:36–43.

Robbins, R. and McKone, E. (2003). Can holistic processing be learned for inverted faces? *Cognition*, 88(1):79–107.

Robbins, R. and McKone, E. (2006). No face-like processing for objects-of-expertise in three behavioral tasks. *Cognition*, 103:34–79.

Rosch, E., Mervis, C., Gray, W., Johnson, D., and P. Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8:382–439.

Rossion, B., Collins, D., Goffaux, V., and Curran, T. (2007). Long-term expertise with artificial objects increases visual competition with early face categorization processes. *Journal of Cognitive Neuroscience*, 19(3):543–555.

Rossion, B. and Curran, T. (2010). Visual expertise with pictures of cars correlates with rt magnitude of the car inversion effect. *Perception*, 39(2):173–183.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., and Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychological Science*, 13(3):250–257.

Rossion, B., Kung, C.-C., and Tarr, M. J. (2004). Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex. *Proceedings of the National Academy of Sciences*, 101(40):14521–14526.

Russell, R., Sinha, P., Biederman, I., and Nederhouser, M. (2004). The importance of pigmentation for face recognition. *Journal of Vision*, 4:418.

Sadato, N., Pascual-Leone, A., Grafman, J., Deiber, M. P., Ibañez, V., and Hallett, M. (1998). Neural networks for braille reading by the blind. *Brain*, 121 (Pt 7):1213–1229.

Sadr, J., Jarudi, I., and Sinha, P. (2003). The role of eyebrows in face recognition. *Perception*, 32:285–293.

Saito, D., Okada, T., Honda, M., Yonekura, Y., and Sadato, N. (2007). Practice makes perfect: The neural substrates of tactile discrimination by mah-jong experts include the primary visual cortex. *BMC Neuroscience*, 7:79.

Scapinello, K. and Yarmey, A. (1970). The role of familiarity and orientation in immediate and delayed recognition of pictorial stimuli. *Psychonomic Science*, 21(6):329–331.

Schiff, W., Banka, L., and de Bordes Galdi (1986). Recognizing people seen in events via dynamic 'mug shots'. *American Journal of Psychology*, 99:219–231.

Schwaninger, A., Wallraven, C., Cunningham, D., and Chiller-Glaus, S. (2006). Processing of identity and emotion in faces: a psychophysical, physiological and computational perspective. *Progress in Brain Research*, 156:321–343.

Schwarzer, G. (2000). Development of face processing: the effect of face inversion. *Child Developmental*, 71(2):391–401.

Schwarzer, G., Zauner, N., and Jovanovic, B. (2007). Evidence of a shift from featural to configural face processing in infancy. *Developmental Science*, 10(4):452–463.

Schyns, P., Bonnar, L., and Gosselin, F. (2002). Show me the features! understanding recognition from the use of visual information. *Psychological Science*, 13(5):402–409.

Scott, L., Shannon, R., and Nelson, C. (2005). Behavioral and electrophysiological evidence of species-specific face processing. *Cognitive, Affective & Behavioral Neuroscience*, 5:405–416.

Scott, L., Shannon, R., and Nelson, C. (2006a). Neural correlates of human and monkey face processing in 9-month-old infants. *Infancy*, 10:171–186.

Scott, L., Tanaka, J., Sheinberg, D., and Curran, T. (2006b). A reevaluation of the electrophysiological correlates of expert object processing. *Journal of Cognitive Neuroscience*, 18:1453–1465.

Scott, L., Tanaka, J., Sheinberg, D., and Curran, T. (2008). The role of category learning in the acquisition and retention of perceptual expertise: A behavioral and neurophysiological study. *Brain Research*, 1210:204–215.

Searcy, J. and Bartlett, J. (1996). Inversion and processing of component and spatial-relational information of faces. *Journal of Experimental Psychology: Human Perception and Performance*, 22:43–47.

Sekuler, A., Gaspar, C., Gold, J., and Bennett, P. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, 14(5):391–396.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology*, 75 (Pt 2):221–242.

Sergent, J., Ohta, S., and MacDonald, B. (1992). Functional neuroanatomy of face and object processing. a positron emission tomography study. *Brain*, 115 Pt 1:15–36.

Simion, F., Cassia, V., Turati, C., and Valenza, E. (2001). The origins of face perception: Specific versus non-specific mechanisms. *Infant and Child Development*, 10:59–65.

Sloman, S. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119:3–22.

Squire, L. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience*, 4(3):232–243.

Tanaka, J. and Farah, M. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, 12:242–248.

Tanaka, J. and Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, 23:457–482.

Tanaka, J. W. (2001). The entry point of face recognition: evidence for face expertise. *Journal of Experimental Psychology: General*, 130(3):534–543.

Tanaka, J. W. and Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, 12(1):43–47.

Tanaka, J. W., Curran, T., and Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychological Science*, 16(2)(2):145–151.

Tanaka, J. W. and Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition*, 25(5):583–592.

Tarr, M. J. and Gauthier, I. (2000). Ffa: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, 3(8):764–769.

Tarr, M. J. and Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21(2):233–282.

Thompson, P. (1980). Margaret thatcher: a new illusion. *Perception*, 9(4):483–484.

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., and Kanwisher, N. (2000). Response properties of the human fusiform face area. *Cognitive Neuropsychology*, 17(1):257–280.

Troje, N. F. and Bülthoff, H. H. (1996). Face recognition under varying poses: the role of texture and shape. *Vision Research*, 36(12):1761–1771.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., and Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, 311(5761):670–674.

Valentine, T. (1988). Upside-down faces: A review of the effects of inversion upon face recognition. *British Journal of Psychology*, 79:471–491.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quaterly Journal of Experimental Psychology*, 43A(2):161–204.

Valentine, T. (1999). *Face-Space Models of Face Recognition.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Valentine, T. and Bruce, V. (1986). The effect of race, inversion and encoding activity upon face recognition. *Acta Psychologica (Amst)*, 61(3)(3):259–273.

van Belle, G., de Graef, P., Verfaillie, K., Rossion, B., and Lefèvre, P. (2010). Face inversion impairs holistic perception: evidence from gaze-contingent stimulation. *Journal of Vision*, 10(5):1–10.

Vogel, E., Woodman, G., and Luck, S. (2001). Storage of features, conjunctions and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1):92–114.

Vokey, J. and Read, J. (1992). Familiarity, memorability and the effect of typicality on the recognition of faces. *Memory & Cognition*, 20:291–302.

Vokey, J. and Read, J. (1995). *Cognitive and computational aspects of face recognition*, chapter Memorability, familiarity and categorical structure in the recognition of faces. London: Routledge.

Walk, R. and Pick, H. J. (1981). *Intersensory perception and sensory integration.* New York: Plenum.

Wallis, G. and Bülthoff, H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences*, 98(8):4800–4804.

Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex*, 15(8):1234–1242.

Yarbus, A. (1967). *Eye movements and vision.* New York: Plenum Press.

Yarmey, A. (1971). Recognition memory for familiar 'public' faces: Effects of orientation and delay. *Psychonomic Science*, 24(6):286–288.

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

Yip, A. and Sinha, P. (2002). Role of color in face recognition. *Perception*, 31:995–1003.

Young, A. W., Hellawell, D., and Hay, D. C. (1987). Configurational information in face perception. *Perception*, 16(6):747–759.

Yovel, G. and Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current Biology*, 15(24):2256–2262.