



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

Copyright statement

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

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

Liability statement

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

Access Agreement

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

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

Sensory Cue Interactions in Person Perception: Insights from Ageing and Developmental Prosopagnosia

A thesis submitted in fulfilment of the requirements for the degree of Doctor
of Philosophy to the University of Dublin, Trinity College

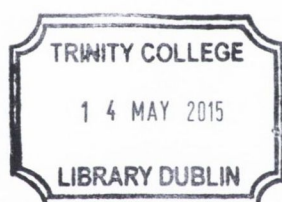
Corrina Maguinness
M.A.

Under the supervision of Professor Fiona Newell
School of Psychology

First Submitted in October 2014



Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

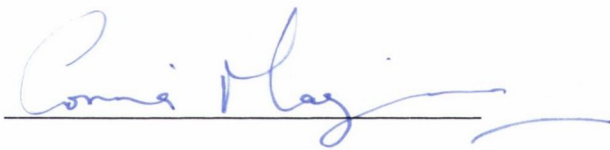


Thesis 10528

Declaration

I hereby declare that this thesis has not been submitted as an exercise for a degree at this or any other University, and it is entirely my own work.

I agree that the Library may lend or copy the thesis upon request. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.



Corrina Maguinness

January 2015

Student number: 10261956

Summary

A large body of behavioural and neuropsychological research, using static images of faces, has informed our understanding of the mechanisms underlying identity processing. Indeed, pioneering studies revealed dedicated modules for the processing of facial form in the human brain. Yet interactions with people in the real world are accompanied by a greater variety of sensory information beyond visual images, for example, people move and speak and are thus highly dynamic, multisensory stimuli. While these additional cues (e.g. facial motion, body motion and the voice) have received attention from researchers in the field, they have largely been studied in isolation, possibly driven by assumptions that each identity cue was processed independently. However, recent findings that facial form modules in the brain also respond to facial movement and to voice information (in the absence of a facial image) have sparked a re-examination of how these sensory cues may *interact* to support person perception. To date, assertions as to how these cues may interact have been largely based on observations from neurotypical younger adult studies (reviewed in detail in Chapter 1).

The aim of this thesis is to extend our understanding of how multiple sensory cues may interact in person perception through the study of three different cohorts: younger adults, older adults and developmental prosopagnosics (DP). This thesis argues that the unique approach of examining cohorts where there is a decline (older age) or a disorder (DP) in static face processing, may give rise to a more concrete understanding of how static and dynamic cues interact to support normal person perception. For example, it is poorly understood how impaired face processing, as found in cases of DP, may impact on the processing of other perceptual information e.g. perceiving the voice, or whether perception in an older person may benefit from multisensory cue combination, as processing in one modality declines. Such findings would clarify how the interactions between the senses may shape our ability to represent a person in memory. A number of behavioural studies, described in this thesis, examined this issue by centring on three main themes: 1) the interaction of form and motion cues in face perception; 2) the interaction of facial motion and the voice in person perception; and 3) the interaction between dynamic body cues and the voice in person perception.

The first theme involved studies which confirmed and expanded on previous findings of poorer static face processing in ageing and developmental prosopagnosia, using a battery of standardised face processing tests (Chapter 2). The tests revealed that older adults showed poor performance for recognising facial identity across changes in the visual appearance of the face; although this pattern was particularly exaggerated for the two cases of DP. Using a series of

matching experiments reported in Chapter 3, the results demonstrated that learning a face in motion can lead to more accurate matching of facial identity across changes in the viewpoint and expression of the face, in older, relative to younger adults. This finding demonstrates that these two sensory cues, form and motion, are readily integrated and can, as a consequence, enhance the representation of the face in memory. The results of the experiments reported in Chapter 4 suggest evidence that certain types of facial motion resulted in poor matching of facial identity in the DP group, suggesting that motion may have distracted from the encoding of static form cues in faces. This finding adds support to the idea that the interactions between form and motion in face processing may be mandatory, as the ability to use motion cues may also be compromised in disorders of face recognition.

The second theme focused on interactions between the face and voice for person perception. Using a series of matching tasks, evidence reported in Chapter 5 revealed that relative to younger adults, older adults were poorer at matching identity from vocal cues alone. Interestingly, this pattern was exaggerated in one case of DP, suggesting that atypical face processing may impact on how the voice is processed to support identity. In addition, the results of Chapter 5 demonstrated that older adults', and to some extent younger adults', ability to represent the face and the voice as a single coherent percept, was likely mediated by the presence of dynamic, relative to static, cross-modal cues. These results support cross-modal sensory interactions for the purpose of representing identity.

The third theme concentrated on interactions between the voice and the body in person perception. Using a sex categorisation task reported in Chapter 6, the results demonstrated that sex cues present in the voice could modulate the perception of 'gender-ambiguous' biological motion displays in younger adults and in DP. However, in general older adults did not appear to benefit from voice information in the same way as younger adults, suggesting that sensory integration for less bound couplings e.g. the body and the voice, compared to the face and the voice, may be less likely to be observed in older age.

Taken together the main findings from this collection of studies support and inform emerging models of person perception and suggest that dynamic, cross-modal cues are integrated for the purpose of person perception at earlier stages of processing than previously assumed. As a consequence person perception can be enhanced or at least modulated by these interactions. The theoretical implications of these results are discussed in detail in Chapter 7.

Publications arising from this thesis (to date)

Maguinness, C. & Newell, F. N. (2014). Recognising others: Adaptive changes to person recognition throughout the lifespan. In Schwartz, B. L., Howe, M. L., Toglia, M. P., & Otgaar, H. (Eds). *What is adaptive about adaptive memory?* Oxford University Press. ([Chapter 1](#))

Maguinness, C. & Newell, F. N. (2014). Motion facilitates face perception across changes in viewpoint and expression in older adults. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 2266-2280. ([Chapter 3](#))

Barrett, M., Cullen, B., Maguinness, C., Merriman, N., Roudaia, E., Stapleton, J., Stienen, B. M. C., & Newell, F. N. (2012). A glance back on 50 years of research in perception. *The Irish Journal of Psychology*, 34, 1-7. ([Chapter 1](#))

Conference peer-reviewed abstracts

Maguinness, C. & Newell, F. N. (2014). Audio-visual interactions in face and voice perception in older adults. *Perception 43 ECVF Abstract Supplement*, 93. ([Chapter 5](#))

Maguinness, C. & Newell, F. N. (2014). Still moving: form and motion integration for face perception revealed by prosopagnosia. *Proceedings of Experimental Psychology Society Meeting*, Kent, UK. ([Chapter 4](#))

Maguinness, C. & Newell, F. N. (2013). Dynamic information benefits unfamiliar face perception in older adults. *Perception 42 ECVF Abstract Supplement*, 139. ([Chapter 3](#))

Maguinness, C. & Newell, F. N. (2012). Changing views: the role of ageing in the matching of unfamiliar faces across dynamic or static changes in viewpoint. *Perception 41 ECVF Abstract Supplement*, 182. ([Chapter 3](#))

Maguinness, C., Rogers, D., & Newell, F. N. (2011). The role of multisensory motion in the perception of gender and deployment of attention. *Perception 40 ECVF Abstract Supplement*, 183. ([Chapter 6](#))

Acknowledgements

First and foremost I would like to thank my supervisor Fiona Newell. I am deeply grateful for all your invaluable support and guidance throughout the PhD process. I have always been inspired by our meetings and conversations about results and their theoretical implications.

To my partner Michael, I thank you for all your love and continuous support. Also to my parents whom I love so dearly. You are the center of my world and have grounded me throughout this process. I will be eternally grateful for all your kindness and support.

Thanks are also due to Bradley Duchaine, Niko Troje, Job van der Schalk and Angeta Fischer for providing the visual stimuli which were used in many of the experiments reported in this thesis. Many thanks are also due to Ciara Devine and Hanni Kiiski for recoding the stimuli used in a number of studies reported here and to David Mc Govern for his help and guidance with data analysis.

I would also like to thank all the lab members whom I have worked with over the past years, particularly Eugenie Roudaia, Annalisa Setti and Jason Chan. I have learned many things throughout the years from working with you. Also thanks are due to Rose Anne Kenny for providing me with a number of invaluable opportunities during the PhD process.

Finally, I would also like to thank all of the older adults who gave up their time to participate in the studies reported throughout. I am particularly appreciative to UM and PL who are two of loveliest ladies I have ever met. I feel privileged to have met both of you and for you to have shared your experiences about living with your face recognition disorder with me. I hope someday we will understand the etiology of the disorder well enough to offer some avenue of rehabilitation. If that day comes, both of you and all the other prosopagnosics who take part in research studies throughout the world, should know that it is the gracious giving of your time which will have been one of the main driving forces of this success.

Table of Contents

Chapter 1 Introduction to Person Perception and Recognition 2

1.1. Overview 2

1.2. Recognising Faces 3

1.2.1. Individual differences and genetic underpinnings 3

1.2.2. Environmental influence 5

1.2.3. Disorders of face recognition: developmental prosopagnosia 5

1.2.4. Changes in face recognition: the effect of ageing 7

1.2.5. ‘Face-space’ and false alarms: changes in face representations in older age 8

1.2.6. Face processing: underlying neural processes 10

1.2.6.1 The fusiform face area 10

1.2.6.2. The core face processing network 11

1.3. Recognising Others: Beyond the Static Face 12

1.3.1. Dynamic faces 12

1.3.1.1. Accommodating motion for recognition in models of face processing 12

1.3.2. Body and biological motion perception 16

1.3.2.1. Body and biological motion processing: underlying neural processes 18

1.3.3. Voice perception 18

1.3.3.1 Representing voices in the brain: underlying neural processing 19

1.3.4. Interactions between the face and voice in person perception 19

1.4. Outline and Scope of the Thesis 23

Chapter 2 Characterising Face Processing Abilities in Neurotypical Older Adults and Age-Matched Developmental Prosopagnosics 26

2.1. Introduction 26

2.2. Face Processing Battery *Part I* 27

2.3. Face Processing Battery *Part II* 29

2.4. Object Processing Battery 30

2.5. General Method	31
2.5.1 Developmental prosopagnosia case reports	31
2.5.2. Older adult participants (age-matched control group)	31
2.5.3. Tests for general sensory acuity and cognitive impairment	32
2.5.4. Comparison analysis	33
2.6. Face Processing Battery <i>Part I</i>	34
2.6.1. Cambridge Face Memory Test (Test 1).....	34
2.6.1.1. Method	34
2.6.1.2. CFMT Results	36
2.6.1.3. CFMT Discussion	37
2.6.1.4. CFMT-Aus Results	38
2.6.1.5. CFMT-Aus Discussion.....	39
2.6.2. Cambridge Face Perception Test (Test 2).....	39
2.6.2.1. Method	39
2.6.2.2. CFPT Results.....	41
2.6.2.3. CFPT Discussion	41
2.7. Face Processing Battery <i>Part II</i>	43
2.7.1. Famous Faces Test (Test 3).....	43
2.7.1.1. Method	43
2.7.1.2. FFT Results	44
2.7.1.3. FFT Discussion.....	44
2.7.2. Benton Face Recognition Test (Test 4).....	45
2.7.2.1. Method	45
2.7.2.2. BFRT Results	46
2.7.2.3. BFRT Discussion	46
2.8. OBJECT PROCESSING BATTERY	47
2.8.1. Cambridge Car Memory Test (Test 5).....	47
2.8.1.1. Method	47

2.8.1.2. CCMT Results 47

2.8.1.3. CCMT Discussion 48

2.8.2. Birmingham Object Recognition Battery (Tests 6A-6D)..... 48

2.8.3. Picture Naming Task (Test 6A) 48

2.8.3.1. Method..... 48

2.8.4. Object Decision Task (Test 6B) 49

2.8.4.1. Method..... 49

2.8.5. Association Match Task (Test 6C) 50

2.8.5.1. Method..... 50

2.8.6. Minimal Feature View Task (Test 6D) 51

2.8.6.1. Method..... 51

2.8.7. BORB Results 51

2.8.7.1. BORB Discussion..... 52

2.9. Relationship between measures of face processing in older adults..... 52

2.10. Explaining older adult variance in face processing performance 55

2.11. Summary 56

Chapter 3 The Effect of Motion on Face Perception in Younger and Older Adults 59

3.1. Abstract 59

3.2. Introduction 60

3.2.1. Outline of the Experiments 64

3.3. EXPERIMENT 1A 64

3.3.1. Method..... 66

3.3.2. Results 69

3.3.3. Discussion..... 73

3.4. EXPERIMENT 1B 74

3.4.1. Method..... 74

3.4.2. Results 76

3.4.3. Discussion..... 77

3.5. EXPERIMENT 2 77

 3.5.1. Method 78

 3.5.2. Results 80

 3.5.3. Discussion 83

3.6. EXPERIMENT 3 84

 3.6.1. Method 85

 3.6.2. Results 86

 3.6.3. Discussion 87

3.7. General Discussion 88

Chapter 4 The Effect of Motion on Face Perception in Developmental Prosopagnosia 95

 4.1. Abstract 95

 4.2. Introduction 96

 4.2.1. Outline of the Experiments 100

 4.3. EXPERIMENT 4 100

 4.3.1. Method 101

 4.3.2. Viewpoint Change Cost Results 103

 4.3.3. Discussion 106

 4.4. EXPERIMENT 5 106

 4.4.1. Method 107

 4.4.2. Expression Change Cost Results 108

 4.4.3. Discussion 111

 4.5. EXPERIMENT 6 111

 4.5.1. Method 112

 4.5.2. Results and Discussion 112

 4.6. General Discussion 113

Chapter 5 Interactions between the Face and Voice in Person Perception 120

 5.1. Abstract 120

5.2. Introduction 122

5.2.1. Interactions between the face and voice in ageing 125

5.3. Outline of the Experiments 126

5.4. EXPERIMENT 7 127

5.4.1. Method..... 127

5.4.2. Results 131

5.4.3. Discussion..... 132

5.5. EXPERIMENT 8..... 133

5.5.1 Method..... 133

5.5.2. Results 138

5.5.3. Discussion..... 140

5.6. EXPERIMENT 9..... 141

5.6.1. Method..... 142

5.6.2. Results 145

5.6.3. Discussion..... 147

5.7. EXPERIMENT 10..... 148

5.7.1. Face and voice interactions in developmental prosopagnosia 148

5.7.2. Method..... 149

5.7.3. Analysis and Results 150

5.7.4. Discussion..... 151

5.8. General Discussion 152

5.8.1. Voice processing in developmental prosopagnosia 155

Chapter 6 Vocal Cues Can Modulate the Perception of Gender in Point-Light Walker Displays
..... 158

6.1. Abstract 158

6.2. Introduction 160

6.3. Outline of the Experiments 163

6.4. EXPERIMENT 11..... 163

6.4.1. Method	164
6.4.2. Results	168
6.4.3. Discussion	168
6.5. EXPERIMENT 12A.....	169
6.5.1. Method	169
6.5.2 Results	172
6.5.3. Discussion	173
6.6. EXPERIMENT 12B.....	174
6.6.1. Method	174
6.6.2. Results	175
6.6.3. Discussion	177
6.7. EXPERIMENT 13	177
6.7.1. Method	178
6.7.2. Results	179
6.7.3. Discussion	180
6.8. EXPERIMENT 14	180
6.8.1. Method	181
6.8.2. Results	182
6.8.3. Discussion	183
6.9. EXPERIMENT 15A.....	184
6.9.1. Sex classification in developmental prosopagnosia	184
6.9.2. Method	185
6.9.3. Analysis and Results	186
6.10. EXPERIMENT 15B	187
6.10.1. Method	187
6.10.2. Analysis and Results	187
6.10.3. Discussion	188
6.11. General Discussion	190

6.11.1. Sex classification in developmental prosopagnosia..... 193

Chapter 7 General Discussion 196

7.1. Introduction 196

7.1.1. Summary of the main findings 197

7.2. Implications of the findings..... 200

7.2.1. Integrating the face, voice and body into models of person perception 200

7.2.2. Representing faces in the brain: Reforming ‘face-space’ 205

7.3. Insights into outstanding issues in the field..... 208

7.3.1. Dynamic enhancements over extended intervals 208

7.3.2. Expression and identity processing..... 209

7.4. Limitations of current studies and future directions 211

7.5. Synopsis..... 213

8. References..... 214

9. Appendices..... 241

9.1. Appendix A 241

9.2. Appendix B 242

List of Figures

Figure 1.1. Although faces 1 and 2 share more visual similarity than faces 1 and 3, the perceptual system must become tuned to recognising the face of the same individual over large changes in visual appearance. Therefore, recognising faces 1 and 3, rather than 1 and 2, as common. Face images are taken from the Radboud Faces Database (RaFD) (Langner et al., 2010). 4

Figure 1.2. Illustration of norm-based encoding of identity, adapted from Valentine (1991). Each blue dot represents a facial identity. Dots which are closer together indicate faces which are similar in visual appearance. Note that identities which are further from the centre may be recognised more readily, as they do not compete with the identities represented in the clustered centre. Valentine stated that face space likely has *n* dimensions, although for the purpose of illustration Valentine offered a 2 dimensional representation. 9

Figure 1.3. A schematic illustration of the Bruce and Young (1986) functional model of face recognition. The model proposes that static-based form cues (blue text) support face recognition, while dynamic facial cues are processed independently to support social signal analysis (red text). The revised model by Burton et al. (1990) assumes that facial familiarity decisions would occur in Person Identity Nodes (PIN), rather than in a ‘face recognition unit’. Burton and colleagues argued that units in the PIN could support familiarity decisions across different modalities, such as the voice (green text). Note this assumes that face and voice cues can only interact to support recognition at supramodal, i.e. post perceptual, stages of processing. The illustration is adapted from Bruce and Young (1986). 14

Figure 1.4. A schematic illustration of the working model for face recognition adapted from O’Toole et al. (2002). This model uses Haxby et al. (2000) extended neural face processing system as a framework for understanding how static and dynamic cues may support face recognition. Like Bruce and Young (1986), the model highlights that static based form cues (processed in FFA) are used to support recognition (blue text). However, unlike Bruce and Young (1986), this model suggests that dynamic cues (red text) are also relevant to face recognition. Motion cues may support recognition independently of structural form information (processed in STS), or they may enhance the structural representation of the face through interactions between STS and FFA (dashed red line). 16

Figure 2.1. Illustration of each test stage of the CFMT. Images adapted from Duchaine and Nakayama (2006)..... 36

Figure 2.2. Illustration of a correctly ordered sample trial for the CFPT. Images on the lower panel are morphed to contain varying percentages of the target face image, which must be arranged in order of similarity to the target identity (upper panel). Images adapted from Duchaine et al. (2007). 40

Figure 2.3. Plot showing the correlation between performance on the CFPT and the CFMT-Aus, for the older adult group (N=16). Note that better performance on the CFPT (lower errors) was associated with increased accuracy on the CFMT (higher scores). 53

Figure 3.1. An example of a trial which illustrates the face images shown in the static (upper panel) and rigid motion (lower panel) training conditions used in Experiment 1A. The pair of images shown on the right depicts the test phase and includes an image of the target and the distractor face. The upper panel depicts an example of a novel test view of the target face, whereas the lower panel depicts a learned face image. 69

Figure 3.2. Plot showing the mean cost in accuracy for matching a novel relative to a learned viewpoint of the target face, for the older and younger adults tested in Experiment 1A. The results are shown for each of the static or dynamic (rigid motion) target face learning conditions. Error bars show ± 1 standard error of the mean (SEM). 72

Figure 3.3. Plot showing the mean reaction times for younger adults to match a learned, relative to novel, view of the face across static fluid and static jumbled learning conditions. Error bars show ± 1 SEM. 77

Figure 3.4. An example of a trial which illustrates the face images shown in the static (upper panel) and non-rigid motion (lower panel) training conditions used in Experiment 2. The pair of images shown on the right depicts the test phase and includes an image of the target and the distractor face. The upper panel depicts an example of a novel test expression of the target face, whereas the lower panel depicts a learned face image. 80

Figure 3.5. Plot showing the mean cost in accuracy for matching a novel relative to a learned expression of the target face, for the older and younger adults tested in Experiment 2. The data are presented for each of the static or dynamic (non-rigid motion) target face learning conditions. Error bars show ± 1 SEM. 82

Figure 3.6. Plot showing the interaction between the test expression and learned target face on the mean response times in Experiment 3, in which all face images were inverted. Error bars show ± 1 SEM. 87

Figure 4.1. Plot showing the mean cost in accuracy for matching a novel relative to a learned viewpoint of the target face, across each of the prosopagnosics and their age-matched controls tested in Experiment 4. The results are shown for each of the static or dynamic (rigid motion) target face learning conditions. Error bars show ± 1 standard error of the mean (SEM). 104

Figure 4.2. Plot showing the cost in accuracy for matching a novel relative to a learned viewpoint of the target face, for each of the age-matched controls (n=14) and both prosopagnosics tested in Experiment 4 (rigid motion). The Y axis represents each participant tested. The results are shown separately for a) the static (upper plot) or b) dynamic (lower plot) target face learning conditions. 105

Figure 4.3. Plot showing the mean cost in accuracy for matching a novel relative to a learned expression of the target face, for age-matched controls and prosopagnosics tested in Experiment 5. The results are shown for each of the static or dynamic (non-rigid motion) target face learning conditions. Error bars show ± 1 standard error of the mean (SEM). 109

Figure 4.4. Plot showing the cost in accuracy for matching a novel relative to a learned expression of the target face, for each of the individual age-matched controls (n=14) and both prosopagnosics tested in Experiment 5 (Non-Rigid Motion). The Y axis represents each participant tested. The results are shown separately for the a) static (upper plot) or b) dynamic (lower plot) target face learning conditions. 110

Figure 4.5. An example of a trial which illustrates the face images shown in the static (upper panel) and non-rigid motion (lower panel) training conditions used in Experiment 6. The pairs of images (A-D) shown on the right depict the different test phases and include an image of the target and the distractor expressions, with different identity combinations. In A both test images depict the same identity as the learning condition, while in B both images depict a different identity to the learning condition. In C and D both test images depict the same and different identities to the learning condition. Note that expression and identity overlap in C and not D. 113

Figure 5.1. Illustration of different trials in Experiment 7. a) Depiction of a same sentence content trial (upper panel) and b) different sentence content trial (lower panel). Note that *both* the target learning and test conditions are auditory (voice) only. 130

Figure 5.2. Plot showing the mean accuracy scores for matching speaker identity across the same or different speech content, for both younger and older adult groups tested in Experiment 7. Error bars show ± 1 SEM. 132

<i>Figure 5.3.</i> An illustration of a sample audio-visual speaker learning trial, where the voice is paired with either its static or dynamic face counterpart during learning. Note that both test options are auditory only. The upper panel (a) represents a same sentence content trial, while the lower panel (b) represents a different sentence content trial.....	137
<i>Figure 5.4.</i> Plot showing the mean reaction times for older adults to match speaker identity across the same, relative to different, sentence content trials for static (left of figure) or dynamic (right of figure) face-voice pairings in Experiment 8. Error bars show ± 1 SEM.....	139
<i>Figure 5.5.</i> Plot showing the interaction between the sentence content and interval on the mean accuracy scores for younger adults in Experiment 8. Error bars show ± 1 SEM.	140
<i>Figure 5.6.</i> An illustration of a sample audio-visual speaker learning trial, where the voice is paired with either its static or dynamic face counterpart during learning. The upper panel a) represents a face matching trial in which the test face images (match or distractor) were always static or dynamic, while the lower panel b) represents a voice matching trial.	145
<i>Figure 5.7.</i> Plot showing the mean accuracy scores for matching the face or the voice, when the target identity was presented as a static face-voice pairing (left of figure) or a dynamic face-voice pairing (right of figure) during learning. Error bars show ± 1 SEM.	146
<i>Figure 5.8.</i> Plot showing the mean reaction times for matching the face or the voice, when the target identity was presented with a static face-voice pairing (left of figure) or a dynamic face-voice pairing (right of figure) during learning. Error bars show ± 1 SEM.....	147
<i>Figure 5.9.</i> Plot showing the mean accuracy for matching speaker identity across same and different sentence content trials, for age-matched controls and prosopagnosics tested in Experiment 10. Error bars show ± 1 SEM.....	151
<i>Figure 6.1.</i> An illustration of the gender continuum described and developed by Troje (2002). Negative values (left of centre) represent walkers which are more female in appearance, while positive values (right of centre) represent walkers which are more male in appearance. For illustration purposes 3 static images, extracted from the dynamic sequences at point -6 exaggerated female, -1 gender-ambiguous and 6 exaggerated male, are depicted.	166
<i>Figure 6.2.</i> Plot showing the mean slope for younger and older adult groups as a function of PLW stimuli positioned along the gender continuum. Visual information only was presented in this experiment.	168
<i>Figure 6.3.</i> The proportion of male responses reported for gender-ambiguous walkers, when presented with no voice, a female voice and male voice in Experiment 12A. The dashed line indicates chance performance. Note that sex judgements only differ from chance performance for AV female voice pairings. Error bars represent ± 1 SEM.	173
<i>Figure 6.4.</i> The proportion of male responses reported for gender-ambiguous walkers, when presented with no voice, a female voice and male voice in Experiment 12B. The dashed line indicates chance performance. Note that sex judgements only differ from chance performance for AV male voice pairings. Error bars represent ± 1 SEM.	176
<i>Figure 6.5.</i> The mean gender ratings for gender-ambiguous walkers, when presented with no voice, a female voice and male voice in Experiment 13. The dashed line indicates a 'neutral', i.e. ambiguous rating. Note that gender ratings only differ from this neutral rating for AV female voice pairings. Error bars represent ± 1 SEM.	180

Figure 6.6. Plot showing the individual slopes for UM and PL and the mean slope for age-matched controls as a function of the gender continuum. 186

Figure 6.7. Plot showing the proportion of male responses reported for gender-ambiguous walkers, when presented with no voice, a female voice and male voice for UM and the age-matched controls. The results are shown separately for a) UM’s performance in relation to *all* older adult controls (n=16) and b) UM’s performance in relation to the 5 age-matched older adults who showed an effect of AV presentation. The dashed line indicates chance performance. Error bars represent +1 SEM. 189

Figure 7.1. A schematic illustration of the Bruce and Young (1986) model, revised in line with the findings from the current thesis. This model is updated to suggest that dynamic facial cues (red text) may interact with the encoding of facial form cues (blue text). In addition, the findings from this thesis propose that the face and voice likely interact at perceptual (green text), rather than exclusively at post-perceptual, stages of processing as previously assumed (see also Burton et al. 1990). 201

Figure 7.2. An illustration of a possible integrated form and motion face-space. Each blue dot represents a facial identity (form cues), the red outline represents that motion has enhanced the representation of the face. Although speculative, it is possible that motion cues may be particularly beneficial for representing typical faces which are represented in the clustered centre, helping differentiate between identities. 207

List of Tables

Table 1.1. Accuracy scores* for younger adults and older adults (with standard deviations) on the face processing battery (part I). Scores in bold indicate a significant effect of age. 42

Table 1.2. Accuracy* scores for developmental prosopagnosics UM and PL and age-matched older controls (with standard deviations) on each of the face and object tests in the complete test battery. Scores in bold indicate impaired performance. 54

Table 2.1. The mean percent accuracy performance (with standard deviations) in matching learned and novel views (Experiment 1A) and expressions (Experiment 2 and 3) of the learned target face, for each of the younger and older adult groups. 70

Table 2.2. The mean reaction time taken in milliseconds (with standard deviations) to match learned and novel views (Experiment 1A) and expressions (Experiment 2 and 3) of the learned target face, for each of the younger and older adult groups. 73

Table 3.1. Individual accuracy scores for UM and PL and mean percent accuracy scores for age-matched controls (with standard deviations) for matching facial identity across learned and novel views (Experiment 4) and expressions (Experiment 5) of the face. 106

Chapter 1

Introduction to Person Perception and Recognition

1 Introduction to Person Perception and Recognition

1.1. Overview

Humans are social and we have evolved in close proximity with one another, surrounded by different identities almost everywhere we turn. The accurate perception of others is a fundamental aspect of social cognition, allowing us to detect the identity, sex and intention of an individual (among other attributes). Indeed, humans have developed an expertise for face perception, which arguably exceeds the perception of other stimuli: we can remember thousands of faces as 'unique' (Bruce & Young, 1986). However, recognising others is not purely dependent on our memory for faces alone as other social information can be used to identify others. For example, like the face, both the voice (Belin, Fecteau, & Bédard, 2004) and the body (Cutting & Kozlowski, 1977) contain unique identity signatures which can be used to support recognition.

This opening chapter to the thesis examines the factors which may mediate our capacity to represent facial identity, exploring evidence that face recognition may reflect a rather unique cognitive ability with genetic underpinnings. Although most research on face perception has been based on the recognition of faces from static images, this chapter also discusses emerging research which highlights our ability to use motion cues in the face to support recognition. In parallel, it also examines our ability to use dynamic information from the body to support person perception and reviews evidence on how dynamic visual cues, particularly those displayed by the face, may interact with auditory cues, such as the voice, to support person perception. Furthermore, the commonalities in how the face, body and voice are represented and processed in the brain are also referred to throughout.

Importantly, the findings discussed throughout this chapter are strongly centred on results from studies involving young, neurotypical adults. While this thesis is concerned with how interactions between sensory cues may be affected by older age and in cases of developmental prosopagnosia (DP), the literature reviewed here for these cohorts will focus on static face processing. This is largely due to the fact that the literature pertaining to how other identity cues may be processed in older age and DP is relatively sparse and open for exploration, however the relevant literature for these cohorts will be discussed in detail in individual chapters. Moreover, the experimental work based on studies of young adults contained in this opening chapter will be referred to throughout the thesis as a grounding point for many of the novel studies examined later in the thesis.

Finally, an outline of the current thesis is discussed. Specifically, it is argued that examining how both the ageing process and developmental prosopagnosia modulate interactions between sensory cues (i.e. face, body and voice) in person perception, may offer a unique insight into, and assist in informing, emerging models of person perception.

1.2. Recognising Faces

1.2.1. Individual differences and genetic underpinnings

Although the process appears effortless, recognising and remembering faces as ‘individual’ is a challenge for the perceptual system. Each face we encounter shares the same basic visual-feature organisation as the next, yet we can accurately discriminate between these faces. Furthermore we can recognise the face of a familiar person even over large changes in visual image appearance, such as viewpoint, expression or illumination (Bruce, 1982; Bruce et al., 2001; Burton et al., 1999) (see Figure 1.1). Recent findings suggest that our capacity for remembering faces is robustly separable from other general cognitive abilities, including measures of immediate and delayed recall (Wilhelm, Herzmann, Kunina, & Sommer, 2007). In addition, face recognition performance appears to follow a unique cognitive trajectory, continuing to improve until just beyond the age of 30 years. In contrast other cognitive skills associated with person recognition, for example name recognition, cease to show age improvements almost a decade earlier (Germine, Duchaine, & Nakayama, 2011).

Recent research has suggested that the ability to recognise other faces varies across individuals. This variation is inherently dissociable from individual differences in other general cognitive abilities (Herzmann, Kunina, Sommer, & Wilhelm, 2010; Wilhelm et al., 2010). These differences in face recognition performance are so evident across certain individuals that they have been shown to be *quantitatively* measureable (Russell, Chatterjee, & Nakayama, 2012). For example, recent evidence has suggested the existence of ‘super-recognizers’, individuals with ‘extraordinary’ face recognition ability which greatly surpasses average face recognition performance (Russell, Duchaine, & Nakayama, 2009). In fact, the degree of enhancement observed in these individuals is quantitatively similar to the degree of deficit expressed in face recognition performance in developmental prosopagnosics (Russell et al., 2009). Developmental prosopagnosics are individuals who have a specific impairment in face recognition performance that is not attributable to brain trauma or insult (see further discussion in 1.2.3. below).

Several studies have suggested that rather like the highly heritable and neurobiological correlates that have been observed for general intelligence (for review see Toga & Thompson,

2005), the degree of expertise expressed in face perception may also be largely genetically determined. This genetic component suggests that as humans, face recognition ability may have largely evolved as a phylogenetic adaptation (within species adaptation) (Johnson, 2011). The most compelling evidence that genetics feature as a contributor in face recognition ability come from the results observed through twin studies (Wilmer et al., 2010; Zhu et al., 2010). What is important about these studies in the current context is that the role of genetics can be studied while also controlling for environmental influences, as twins who are raised together largely share similar environmental factors, in the same way as non-identical twins or other siblings. Wilmer et al. (2010) observed that monozygotic (MZ) twins, (i.e. developed from one fertilized oocyte), show highly correlated face recognition performance. In fact the correlation between their performances was more than double the correlation observed for dizygotic (DZ) twins, (i.e. developed from different fertilised oocytes), raised in the same environment (correlation 0.70 versus 0.29, respectively). As the MZ and DZ twin sets both shared similar environmental factors, the observed differences can be assumed to largely reflect genetic underpinnings (Wilmer et al., 2010).



Figure 1.1. Although faces 1 and 2 share more visual similarity than faces 1 and 3, the perceptual system must become tuned to recognising the face of the same individual over large changes in visual appearance. Therefore, recognising faces 1 and 3, rather than 1 and 2, as common. Face images are taken from the Radboud Faces Database (RaFD) (Langner et al., 2010).

Zhu et al. (2010) provided similar evidence for face recognition ability being highly correlated for MZ and less so for DZ twins raised together. Intriguingly, they found no evidence for such a relationship for object recognition abilities across twin sets. They did, however, demonstrate a similar pattern of correlational strength for the ‘face inversion effect’ across twin sets. Image inversion has been shown to disrupt face recognition, yet it is thought to elicit less disruption in the recognition of other objects and is suggested to reflect a method of processing, holistic or configural processing, which is largely specific to face perception (Maurer, Le Grand, & Mondloch, 2002; Schmalz, Palermo, Harris, & Coltheart, 2009; Yin, 1969; Young, Hellawell, & Hay,

1987). Thus the effect Zhu et al. (2010) observed for face recognition performance suggests that it is reflective of the perceptual processes that are recruited specifically for face recognition and not generic low level processing differences across individuals.

1.2.2. Environmental influence

However, it is also important to note that not all of the variance in face recognition performance in the previously mentioned twin studies was explained by genetic similarity. The statistical models used in such twin studies (Rabe-Hesketh, Skrondal, & Gjessing, 2008) take into consideration an error measurement which includes the contribution of non-familial influence i.e. the environmental influence unique to each twin. Thus although the ability to recognise faces may reflect a *unique* and *highly heritable* perceptual ability, environmental influences were also reported to be an important contributor in face recognition performance. This highlights that face recognition is also modulated by ontogenetic individual developmental experience (Johnson, 2011). Indeed, other findings in the face literature have also demonstrated that the degree of exposure to, for example, different face types, can directly impact face recognition ability. Perhaps one of the best examples is highlighted by the *other-race effect* (for a review, see Meissner & Brigham, 2001). Here, face memory is susceptible to a significant degree of error when the face we are to recognise belongs to a different race (ethnic) group. This is mediated by the fact that we perceive individual faces of other-race faces as sharing a large degree of similarity (Valentine & Ferrara, 1991; Vizioli, Rousselet, & Caldara, 2010) and as a consequence exhibit increased 'false alarm' rates in memory for other-race faces in comparison with faces of the race by which we are surrounded (Chiroro, Tredoux, Radaelli, & Meissner, 2008; Chiroro & Valentine, 1995; Meissner & Brigham, 2001). This phenomenon is also known in the literature as the 'own-race bias' and it is argued to emerge during the early stages of development through the process of perceptual narrowing (Kelly et al., 2007).

1.2.3. Disorders of face recognition: developmental prosopagnosia

In addition to twin studies, our understanding of the heritability of face recognition has also been informed by the study of *developmental prosopagnosia*. Prosopagnosia ('face blindness') refers to a condition first documented by Bodamer in 1947 (see Ellis & Florence, 1990 for translated paper) in which patients demonstrated a distinct deficit in recognising individual faces, whereas other visual abilities remained intact. Beyn and Knyazeva (1962) described this condition as reflecting a "failure in recognition at its most highly specific and individual level" (p158). While the first cases documented largely reflected acquired prosopagnosia, which results

from insult to, and dysfunction of, the cortical face processing network (Barton, 2008; Rossion et al., 2003; Schiltz et al., 2006; Tranel, Damasio, & Damasio, 1988), later studies demonstrated the existence of developmental prosopagnosia. Developmental prosopagnosics are impaired in recognising the identity of individual faces, and such deficits emerge in the absence of insult or atypical neuro-development (Duchaine & Nakayama, 2005; Lee, Duchaine, Wilson, & Nakayama, 2010; Susilo & Duchaine, 2013). As such, individuals with developmental prosopagnosia do not lose the ability to recognise faces; rather the ability does not emerge organically during development (Duchaine & Nakayama, 2005). Interestingly, other aspects of face processing including discriminating the sex (Le Grand et al., 2006; Nunn, Postma, & Pearson, 2001) and emotion (Duchaine, Parker, & Nakayama, 2003; Humphreys, Avidan, & Behrmann, 2007) of the face appear to be largely intact in this cohort. Developmental prosopagnosia has been shown to be largely heritable, with multiple cases observed in the same family (Duchaine, Germine, & Nakayama, 2007; Grueter et al., 2007; Lee et al., 2010; Schmalzl, Palermo, & Coltheart, 2008). Furthermore, the disorder is much more prevalent than one might assume. Recent reports indicate that 1 in every 40 individuals, i.e. 2.5 % of the population, likely have this hereditary disorder (Grüter, Grüter, & Carbon, 2008; Kennerknecht et al., 2006; Kennerknecht, Ho, & Wong, 2008). Such findings may implicate the existence of an evolved neural substrate for the recognition of faces (Young, DeHaan, & Bauer, 2008) with genetics arguably playing a role in both typical and atypical development.

Some researchers use the term 'congenital', rather than developmental prosopagnosia (e.g. Behrmann & Avidan, 2005; Humphreys, Avidan, & Behrmann, 2007; Rivolta, Palermo, & Schmalzl, 2013). While the disorder does appear to have a strong heritable component it is important to consider that the term 'congenital' assumes that the disorder is present at birth. Susilo & Duchaine (2013) however highlight that the exact etiology and onset of the disorder is probably more diverse. They argue that it is more likely that abnormalities may arise at different stages during the development of the mechanisms which support normal facial identity processing (Susilo & Duchaine, 2013). Indeed, in neurotypical individuals the development of these exact mechanisms appears to be protracted. For example, activation in face specific areas of the brain (these areas are discussed in detail later on in the chapter) of children when viewing face stimuli is typically not observed, as the areas are relatively immature in terms of both specialisation of function and volume (Scherf, Behrmann, Humphreys, & Luna, 2007). When activation is observed in face selective areas, it is often much less localised than it is in the adult brain (Passarotti et al., 2003). Passarotti et al. (2003) stated that the neural systems involved in face processing continue to experience changes in terms of their functional specialization into late

childhood. This account of increasing specialization in brain function is also in line with behavioural evidence supporting the idea that face recognition ability continues to improve into adulthood in the general population (Germine et al., 2011). As such, it is possible that abnormalities occurring at any stage during childhood development may give rise to a deficit in normal face processing ability which is then evident throughout the lifespan.

This may in part explain, or at least bear some relationship, to the heterogeneity of the disorder. For example, some cases of prosopagnosia are suggested to be apperceptive, rather than associative in nature (De Renzi, Faglioni, Grossi, & Nichelli, 1991). Apperceptive deficits may arise due to abnormal structural encoding of form (i.e. the invariant shape cues which individuate a face) information in upright faces (Németh, Zimmer, Schweinberger, Vakli, & Kovács, 2014; Towler, Gosling, Duchaine, & Eimer, 2012). While associative deficits may emerge at later downstream stages of processing and result from disconnections between visual face memory and semantic information (e.g. Anaki, Kaufman, Freedman, & Moscovitch, 2007; Towler & Eimer, 2012).

1.2.4. Changes in face recognition: the effect of ageing

As noted, face recognition performance does not remain stable throughout the lifespan. Rather, recognition appears to peak at just beyond the age of 30 years (Germine et al., 2011). Subsequently, performance appears to decline as a function of increasing age, with decline evident at around 50 years of age (Bowles et al., 2009). This decrease in face recognition performance appears to be specific to the recognition of faces in comparison to the recognition of other within-category objects (Boutet & Faubert, 2006). What is important to note is that although such changes are evident, they are dissociable from changes in general cognitive ability that occur with the ageing process (Hildebrandt, Wilhelm, Schmiedek, Herzmann, & Sommer, 2011). These deficits in face recognition performance have been suggested to be mediated by a number of factors including poorer sensory acuity typically associated with the ageing process (Barnes, De L'Aune, & Schuchard, 2011; Norton, McBain, & Chen, 2009; Owsley, Sekuler, & Boldt, 1981), changes in how faces are encoded (Chaby, Narme, & George, 2011; Firestone, Turk-Browne, & Ryan, 2007; Gao et al., 2009; Grady et al., 1995; Murray, Halberstadt, & Ruffman, 2010), and marked differences in the underlying neural activity associated with processing faces (Grady, 2002; Haxby et al., 1994; Lee, Grady, Habak, Wilson, & Moscovitch, 2011), including less lateralised activation in face processing areas (Lee et al., 2011). The brain areas engaged in face processing are discussed in detail further on in the chapter.

1.2.5. 'Face-space' and false alarms: changes in face representations in older age

Perhaps one of the most striking behavioural effects which demonstrate this age related decline, is the higher level of 'false alarm' memory responses that older adults report during face recognition tasks, compared to younger adults (Bartlett & Leslie, 1986; Bartlett, Leslie, Tubbs, & Fulton, 1989; Rhodes, Castel, & Jacoby, 2008). For example, older adults show a bias for stating that they find a novel face familiar even when they have not encountered the face before. It is possible that such findings may relate to the fact that older adults have been exposed to a larger number of faces across the lifespan than younger adults. This larger 'database' of faces in memory (or what Valentine (1991) referred to as 'face-space') may elicit larger false alarm responses. By increasing the number of typical faces in face memory, newly encountered faces may be likely to elicit stronger signals of resemblance to another face within this large face-space, thus leading to false recognition.

Specifically, Valentine's model suggests that each face we encounter is represented within a space comprised of multiple dimensions which are used to discriminate faces (Valentine, 1991). See Figure 1.2 for a schematic illustration of the 'face-space' model. At the centre of the space lies the average face. It is argued this average face may function as a perceptual 'norm' by which other faces are represented (review see Rhodes & Jaquet, 2011). Specifically, rather than representing each identity as an absolute value in face-space, the perceptual system may represent each face as a deviation from this central face. Typical faces are likely to be represented in the clustered centre of face-space. As such, the further a face is away from the centre the more distinctive the face is and as a consequence it may be more readily recognised (Bruce, Burton, & Dench, 1994; Newell, Chiroro, & Valentine, 1999; Valentine & Ferrara, 1991; Valentine, 1991; Wickham, Morris, & Fritz, 2000). A number of face adaptation studies have offered compelling evidence that faces are likely coded in this norm-based or 'central tendency' manner (Leopold et al., 2005; Leopold, Toole, Vetter, & Blanz, 2001).

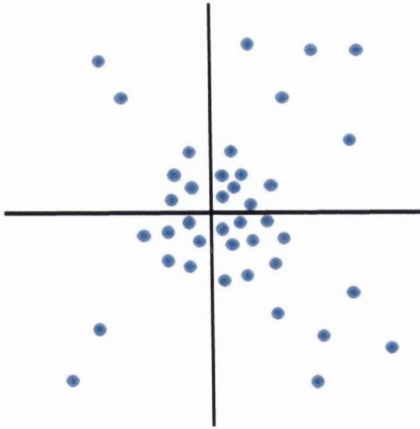


Figure 1.2. Illustration of norm-based encoding of identity, adapted from Valentine (1991). Each blue dot represents a facial identity. Dots which are closer together indicate faces which are similar in visual appearance. Note that identities which are further from the centre may be recognised more readily, as they do not compete with the identities represented in the clustered centre. Valentine stated that face space likely has n dimensions, although for the purpose of illustration Valentine offered a 2 dimensional representation.

However, the benefit on face recognition for distinctive faces is less evident in older, compared to younger adults (Carr, Castel, & Knowlton, 2014). As such, changes in how the face is represented may affect older adults' face recognition performance. Moreover, these effects are not only evident for face memory, since recent findings suggest that older adults are also poorer than younger adults at discriminating between and learning unfamiliar faces (Lee, Smith, Grady, Hoang, & Moscovitch, 2014). Lee et al. (2014) suggest that as we age face representations may become more broadly tuned, thus making it more difficult to represent a unique facial identity. Indeed, a functional magnetic imaging (fMRI) study by Lee, Grady, Habak, Wilson & Moscovitch (2011) supports this finding. In that study, Lee and colleagues demonstrated that older adults showed similar adaptation effects to repeated exposure of the same, relative to a different, facial identity. In general, adaptation occurs when a brain region treats the presented stimuli as containing the same underlying properties (e.g. identity or viewpoint) and this adaptation can have significant perceptual aftereffects on subsequently viewed stimuli (Köhler & Wallach, 1944). Failure of this region to adapt more strongly to repeated exposure to the same facial identity suggests that functional specificity in the face processing system may become less selective with the ageing process. These changes in underlying neural activity have also been regarded as a possible reason why older adults also find it difficult to represent facial identity across changes in visual appearance e.g. viewpoint, or expression, as the selective properties of neurons to a particular viewpoint of the face may also decline (Habak, Wilkinson, & Wilson, 2008; Lee et al., 2011). This age-related decline in representing the face across changes in visual appearance is

referred to again in Chapter 2 and is the focus of Chapter 3. In addition, there is also some evidence to suggest that face space may not be optimally represented in individuals with developmental prosopagnosia (e.g. Palermo, Riolta, Wilson, & Jeffery, 2011), which may contribute to this cohort's poor ability to recognise facial identity from structural form cues.

1.2.6. Face processing: underlying neural processes

1.2.6.1 *The fusiform face area*

Many researchers have argued for domain specificity in face processing, and that the cognitive processes involved in face recognition appear to be distinct and specific for processing faces in comparison with other stimuli (Duchaine, Dingle, Butterworth, & Nakayama, 2004; McKone, Kanwisher, & Duchaine, 2007). This remains controversial, however, because others have suggested that face processing reflects a particular type of perceptual expertise in discriminating between similar visual stimuli, as discussed below. Evidence for domain specificity largely originated from neuropsychological investigations that reported selective deficits to face recognition, i.e. acquired prosopagnosia, following damage to particular regions of the temporal lobe of the brain, which are involved in the processing of faces. Subsequent functional magnetic resonance imaging (fMRI) revealed a face module, known as the fusiform face area (FFA), which is located in visual areas of the ventral stream and which is thought to be engaged in the perception of facial identity (Kanwisher, McDermott, & Chun, 1997). In particular, this area is selectively activated when viewing face images over images of other non-face objects or scrambled face images (suggesting that the area is not simply responsive to visual qualities of the image e.g. luminance). Electrical stimulation of this region specifically distorts the perception of faces over other stimuli (Parvizi et al., 2012).

Although some have argued that this face sensitive activation may reflect 'domain specificity' (Kanwisher et al., 1997) for the processing of faces, it is important to note that the FFA has been suggested to also reflect an area recruited for within-category visual discrimination of other non-face stimuli, which require a similar level of visual 'expertise' for individual discrimination (Gauthier, Skudlarski, Gore, & Anderson, 2000). This 'expertise hypothesis' (Gauthier & Nelson, 2001; Gauthier, Skudlarski, et al., 2000) suggests that the observed cortical specialization for faces is a product of perceptual exposure and training in discriminating between similar visual stimuli. For example, it has been shown that individuals who are experts (or who are trained to become experts) in within-category discrimination of non-face stimuli, such as cars or birds, show similar underlying neural activity in the FFA when perceiving faces or other 'expertise'

stimuli (Gauthier et al., 2000; Rossion, Gauthier, Goffaux, Tarr, & Crommelinck, 2002). In addition, highly familiar non-face stimuli that require holistic processing due to their multipart visual organisation (Bilalic, Langner, Ulrich, & Grodd, 2011) elicit activation in face selective areas. However, this finding remains controversial and a number of studies have reported greater neural activation to faces in comparison with objects of expertise in the FFA (Grill-Spector, Knouf, & Kanwisher, 2004; McKone et al., 2007). It is also important to note that expertise in face recognition and the involvement of the FFA in face processing appear to develop implicitly across all neurotypical individuals. On the other hand, the involvement of the FFA in discriminating between other 'similar' stimuli often involves explicit perceptual discrimination training. Moreover, others have argued that normal face recognition does not rely on the development of normal object recognition, as might be assumed from the expertise hypothesis (Germine, Cashdollar, Düzel, & Duchaine, 2011). However, a bridging of these two viewpoints may be realised through the 'interactive specialization' approach (Johnson, 2011). Interactive specialization considers both the predisposition for cortical activation to e.g. face stimuli (intrinsic factors) and the level of exposure to faces (extrinsic factors), which modulates the degree of localization and specialization of this cortical activity. Thus, under this model areas of the brain may start out as having "broadly tuned functions" and with development they gradually become more refined for more specialized functions (Johnson, 2000, 2001), with direct consequences on localization and specialisation of neural activation in the brain.

1.2.6.2. The core face processing network

However, it is also important to note that a number of functional imaging studies in prosopagnosics have revealed normal activation in FFA when viewing face images (Fox, Iaria, Duchaine, & Barton, 2013; von Kriegstein, Kleinschmidt, & Giraud, 2006), suggesting that activation in FFA alone is not sufficient to support normal identity processing (Fox, Iaria, & Barton, 2008). Indeed, it is thought that the FFA is part of a larger neural network involved in face perception and recognition (Haxby et al. 2000). These areas include the occipital face area (OFA), which is argued to be involved in the early perception of facial structure (Gauthier, Tarr, et al., 2000; Haxby et al., 2000) and areas in the superior temporal sulcus (STS), which are thought to be concerned with processing the changeable aspects of the face, i.e. movement associated with expression or gaze direction (Hoffman & Haxby, 2000; Puce, Allison, Bentin, Gore, & McCarthy, 1998). However, the functional and hierarchical nature of the connections between these areas is debated. Specifically, Haxby and colleagues have argued that face processing begins in the OFA, with subsequent analysis of the face occurring downstream in FFA (for identity) or STS (for social signal analysis). However, others have contended that the changeable aspects of the face may be

processed in STS through an independent network which does not entail ventral based areas i.e. OFA or FFA (O'Toole, Roark, & Abdi, 2002; Pitcher, Duchaine, & Walsh, 2014). Moreover, it has been argued that the facial motion may also be used to support identity, in addition to social signal, processing (O'Toole et al., 2002; Roark, Barrett, Spence, Abdi, & O'Toole, 2003). The next section will turn to this in more detail.

1.3. Recognising Others: Beyond the Static Face

1.3.1. Dynamic faces

As demonstrated throughout this chapter much research has been dedicated to understanding both the perceptual and neurological processes which underlie human face recognition. Although this has brought us far in understanding how the brain processes facial identity, the contribution of such studies to our understanding of person recognition in the real world is limited by the fact that only information in the visual domain is considered and moreover that such evidence is largely based on the use of static images of faces as stimuli. Importantly, ageing effects and prosopagnosia (both acquired and developmental) have to date been studied almost exclusively through the use of static face images. Such investigations belie the true nature of the processes involved in the perception of faces we often encounter every day, as faces move and speak, and are thus inherently dynamic, multisensory stimuli.

In the real world we rarely experience a face in static form and most often see the face in motion, across a large range of viewpoints and expression changes. It would seem odd if we encountered someone who did not move their face. Indeed, static face images extracted from moving sequences are viewed as less appealing than their dynamic counterparts, a finding referred to as the 'frozen face' effect (Post, Haberman, Iwaki, & Whitney, 2012). In addition, it has been reported that most regions of interest (ROI) in the face processing system are more successfully located in neuroimaging studies through the use of dynamic (in comparison to static) face images in functional localizers (Fox, Iaria, & Barton, 2009). As we are constantly surrounded by moving stimuli, it is possible that this finding may be explained by the idea that there are more neurons responsive to moving, relative to static, face images in the human brain (Fox et al., 2009; Schultz & Pilz, 2009).

1.3.1.1. Accommodating motion for recognition in models of face processing

The bias in using static images in studies of face perception has been suggested to reflect an underlying assumption about face perception itself, that recognition is based on our ability to

recognise the face through invariant characteristics (likely processed in OFA and FFA as highlighted above) whilst ignoring changes which occur through movement (Bruce & Young, 2012). Moreover, seminal models of face perception, such as that proposed by Bruce and Young (1986) have argued that dynamic facial cues were directly related to the communication of social signals e.g. expression, speech movements, rather than identity (see Figure 1.3 for illustration of the Bruce and Young, 1986 model). Yet recent findings suggests that our ability to recognise a face as an individual may depend on our ability to process these dynamic changes and in turn create a more robust representation of the face in memory. For example, the way someone moves their forehead when in thought, may be peculiar to that individual and thus we may recognise the person not in spite of this movement, but perhaps in light of it. Thus, spatio-temporal information associated with an individual's face may act as a unique signature to facilitate the recognition of that person (Bülthoff, Cunningham, & Wallraven, 2011; Knappmeyer, Thornton, & Bülthoff, 2003).

Indeed, a number of studies have demonstrated that motion plays an integral role in our ability to recognise familiar others, under degraded visual conditions when available structural form cues are unreliable (Knight & Johnston, 1997; Lander & Bruce, 2000; Lander, Christie, & Bruce, 1999; Lander & Chuang, 2005; O'Toole et al., 2002). For example, Knight and Johnston (1997) demonstrated that when visual images of famous faces were degraded by presenting the images in their negative photographic form, participants were more accurate at identifying the target face when it was shown as a video clip rather than a static image. Indeed, Hill & Johnston (2001) demonstrated that observers could reliably discriminate the sex and identity of a face, when *only* motion cues were available. In that study, Hill & Johnston animated an average avatar face with different dynamic facial identities. These motion profiles had been motion captured by tracking the movement of markers positioned on each actor's face during a previous recording session. As such, even when no reliable shape cues to the face were available, observers could still extract relevant motion cues to support identity and sex classification (Hill & Johnston, 2001). Furthermore, the perceptual system's sensitivity to detecting identity and other socially relevant information is not only restricted to facial motion, as observers can reliably determine meaningful social and identity information from sparse dynamic visual displays which represent a body in motion (Cutting & Kozlowski, 1977; Johansson, 1973; Kozlowski & Cutting, 1977) (this will be addressed later in the Chapter). Recovering identity information from motion may be necessary for recognition under challenging viewing environments when facial form information may be compromised e.g. in dim lighting or at a distance (O'Toole et al., 2002).

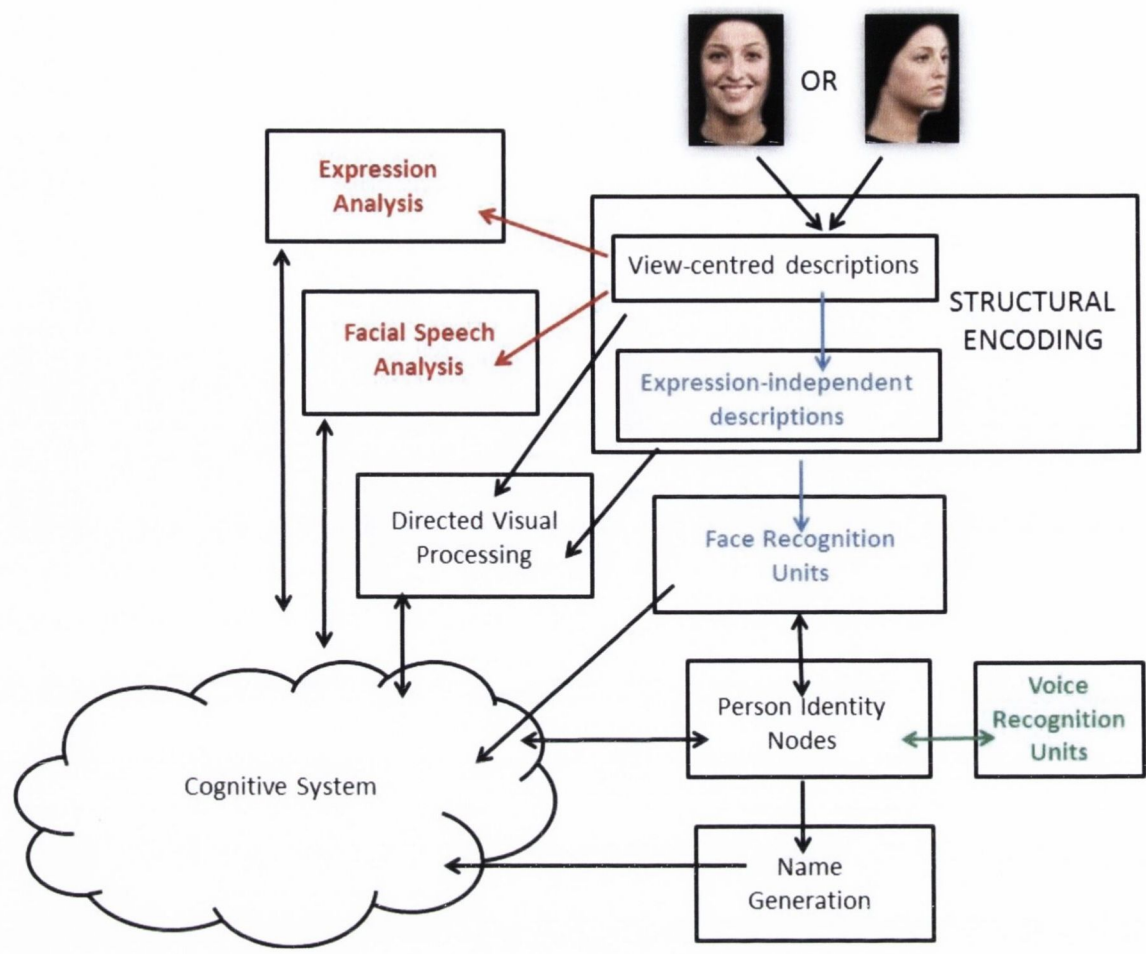


Figure 1.3. A schematic illustration of the Bruce and Young (1986) functional model of face recognition. The model proposes that static-based form cues (blue text) support face recognition, while dynamic facial cues are processed independently to support social signal analysis (red text). The revised model by Burton et al. (1990) assumes that facial familiarity decisions would occur in Person Identity Nodes (PIN), rather than in a 'face recognition unit'. Burton and colleagues argued that units in the PIN could support familiarity decisions across different modalities, such as the voice (green text). Note this assumes that face and voice cues can only interact to support recognition at supramodal, i.e. post perceptual, stages of processing. The illustration is adapted from Bruce and Young (1986).

Thus, motion may provide an additional route to recognition (argued to be processed in the dorsal stream in STS, see Figure 1.4) and assist in determining the familiarity of a face when facial form information is uncertain. However, the role for motion when learning new, i.e. unfamiliar, faces remains somewhat unclear (Christie & Bruce, 1998; Lander & Bruce, 2003; Pike, Kemp, Nicola, & Phillips, 1997). Specifically, there is conflicting evidence to suggest that motion may enhance face learning by creating a more robust structural representation of the face (Pike et al., 1997; Pilz, Thornton, & Bülthoff, 2006; Thornton & Kourtzi, 2002), suggesting interactions between the ventral and dorsal streams (see Figure 1.4). However, others have shown evidence that, in a similar vein to familiar face recognition, motion may support face learning by providing

idiosyncratic dynamic signatures which can be associated with an identity (Hill & Johnston, 2001; Knappmeyer et al., 2003; Longmore & Tree, 2013) (this is discussed in detail and is the focus of Chapter 3 and Chapter 4).

However, what is evident is that preserving *characteristic motion* cues affects performance when learning new faces (Wallis, 1998; Wallis & Bülthoff, 2001) and recognising familiar others (Lander et al., 1999). That is, motion presented in forward form is more beneficial for the perceptual system than disrupted characteristic motion (e.g. reversed, or scrambled). This can be understood in terms of the *temporal-association hypothesis* (Wallis & Bülthoff, 1999). Under this hypothesis the visual system learns to associate multiple views of, for example a face, when they are not only physically similar but also *temporally correlated* (Wallis & Bülthoff, 2001). Thus, if multiple views of a face are seen in close temporal proximity, it is most likely that they all correspond to the same face and are thus integrated into a robust single representation of the face in memory. Wallis and Bülthoff (2001) illustrated this sensitivity to change over time information with a rather extreme example. They observed that individuals learned to associate a view of a different identity with the same face when all images were presented in spatiotemporal coherence. They presented five views of a face in smooth rotation and switched the identity of an intermediate view and observed that people were likely to regard all views as belonging to the one person, having a direct effect on their memory for that face. Yet when the view sequence was 'jumbled', disrupting spatiotemporal association, this effect disappeared. Spatiotemporal association has been suggested to be more relevant for the recognition of faces over objects and may in part reflect the nature in which we encounter faces in movement in the natural environment (Liu, 2007).

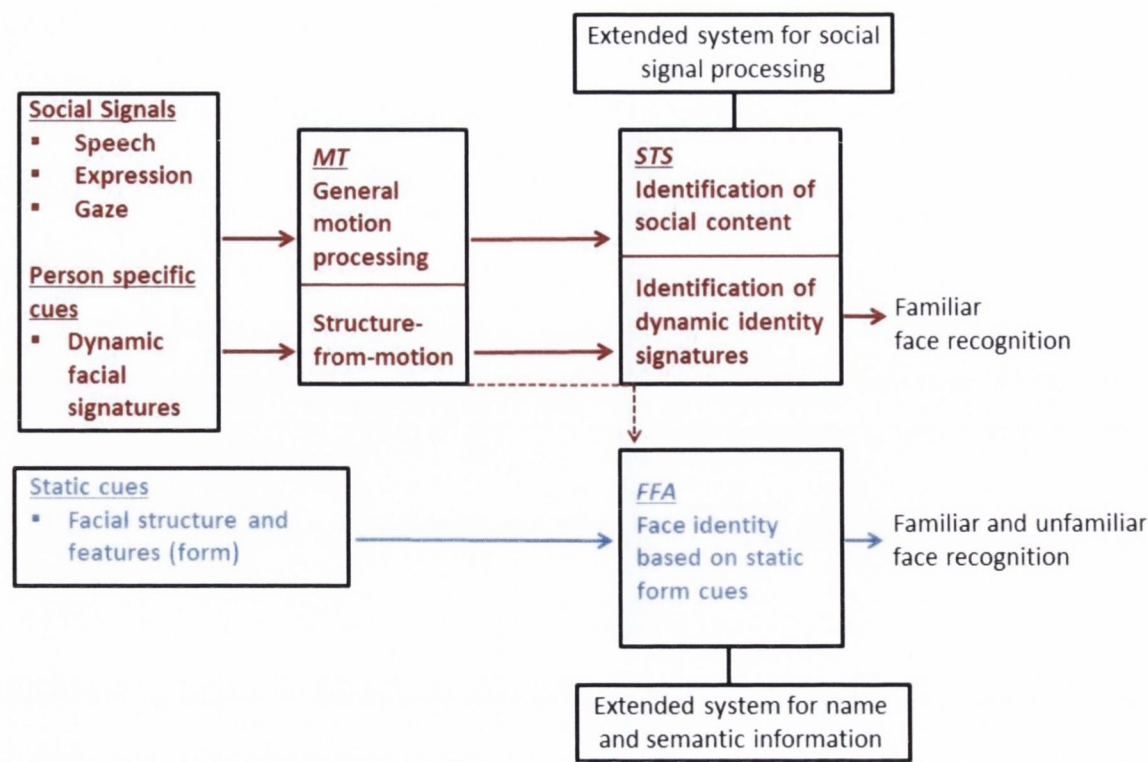


Figure 1.4. A schematic illustration of the working model for face recognition adapted from O’Toole et al. (2002). This model uses Haxby et al. (2000) extended neural face processing system as a framework for understanding how static and dynamic cues may support face recognition. Like Bruce and Young (1986), the model highlights that static based form cues (processed in FFA) are used to support recognition (blue text). However, unlike Bruce and Young (1986), this model suggests that dynamic cues (red text) are also relevant to face recognition. Motion cues may support recognition independently of structural form information (processed in STS), or they may enhance the structural representation of the face through interactions between STS and FFA (dashed red line).

1.3.2. Body and biological motion perception

As highlighted above, face processing is a fruitful and active area of research, offering unique insight into how we recognise others. However, it is important to consider that our ability to recognise individuals is not restricted to the face, rather there is strong evidence to suggest that individuals can recognise others from a range of sensory cues including, but not limited to, the body (Cutting & Kozlowski, 1977; Stevenage, Nixon, & Vince, 1999; Westhoff & Troje, 2007). Indeed, the body has many of the same characteristics that the face has to offer; the body conveys socially relevant information, with repeated exposure it becomes highly familiar, it is based on the same configural arrangement across individuals (Peelen & Downing, 2007) and importantly it is often seen in motion (O’Toole et al., 2014; O’Toole et al., 2011). Moreover, it appears that the body (Rhodes, Jeffery, Boeing, & Calder, 2013) and possibly motion cues displayed by the body (e.g. Hill & Pollick, 2000) may be represented based on the same norm-

based manner as previously described for face representations (Leopold et al., 2005; Rhodes & Jaquet, 2011; Valentine, 1991).

While it has been argued that the face may offer a more salient cue for recognition than the body (Bruce, Henderson, Newman, & Burton, 2001b; Burton et al., 1999), there is evidence that these two sources of visual information may interact when representing identity (O'Toole et al., 2011; Pilz, Vuong, Bülthoff, & Thornton, 2006; Pilz, Vuong, Bülthoff, & Thornton, 2011). For example, Pilz et al. (2006) noted that when participants were presented with a face and a moving body in which the facial identity was ambiguous, (i.e. a 50 percent morph between two different identities) that perception of facial identity was biased by the characteristic motion signatures which were displayed by the body. In addition, O'Toole et al. (2011) demonstrated that identity matching is best when observers are exposed simultaneously to the face and the body in motion, rather than to face and the body in isolation.

Moreover, a number of studies have demonstrated that motion cues displayed by the body are sufficient to support person perception, in the absence of the form cues from the face or the body. For example, Cutting & Kozlowski (1977) demonstrated that observers could reliably recognise both themselves and their friends when their movements were depicted by a small number of point of lights in motion. These 'point-light displays', which were first described and implemented through the pioneering work of Johansson (1973), are typically created by positioning markers or points of light on an actor's joints and reducing the luminance of the background, such that only these points of light are visible. When these displays are presented in their static form, very little information can be extracted from them. In contrast, when viewed in motion these visible points of light give rise to the perception of a person in motion (Johansson, 1973).

Other studies have demonstrated that these dynamic cues can be rapidly acquired for the purpose of learning new identities (Hill & Pollick, 2000; Stevenage et al., 1999; Troje, Westhoff, & Lavrov, 2005; Westhoff & Troje, 2007). Moreover, it has also been reported across a number of studies, that even when the observer is unfamiliar with the identity, they can still categorise the sex of the walker in the dynamic visual displays with performance being better than predicted by chance (for a review see Pollick, Kay, Heim, & Stringer, 2005). Thornton (2006) also reported that when the natural movement profiles of a moving person (i.e. presented by video and not a point-light display) were visually degraded, rendering form cues unreliable, that observers could still detect the sex of the individual from the movement. This finding parallels the face perception

literature discussed above and suggests that dynamic cues can be readily extracted when form cues are degraded. This remarkable ability will be discussed in more depth in Chapter 6.

1.3.2.1. Body and biological motion processing: underlying neural processes

A similar neural network has been proposed for the processing of bodies (Downing, Jiang, Shuman, & Kanwisher, 2001; Grossman & Blake, 2002; Peelen & Downing, 2005; Taylor, Wiggett, & Downing, 2007), to that which has been proposed for faces (Gauthier, Tarr, et al., 2000; Haxby, Hoffman, & Gobbini, 2002; Kanwisher et al., 1997). Specifically, as the OFA has been implicated in the processing of facial features, a region in the lateral occipitotemporal cortex, referred to as the extrastriate body area or EBA (Downing et al., 2001) has been implicated in the processing of images of the human body and body parts, showing an increased functional response profile to these stimuli than other object categories including that of faces.

Additionally, an area in the fusiform gyrus, labelled the fusiform body area (FBA) has also been reported to respond to visual images of the human body, in the absence of the face (Peelen & Downing, 2005). In parallel to activations in the FFA, which has been proposed to process the facial configuration, rather than individual features of the face, the FBA has been shown to exhibit enhanced functional activation in response to complete body images in comparison with body parts (Taylor et al., 2007). This suggests that the FBA may play a role in representing the configuration of the full body form (Taylor et al., 2007). Additionally, areas in posterior STS have been shown to be involved in the processing of dynamic biological motion cues from the body (Bonda, Petrides, Ostry, & Evans, 1996; Grossman & Blake, 2002; Grossman et al., 2000). Damage to this network has been shown to give rise to deficits in body and body motion recognition, similar to the deficits observed in acquired prosopagnosia (e.g. Moro et al., 2008).

1.3.3. Voice perception

Although the voice, like the face, is uniquely associated with a person's identity, most research has focused on how we process vocal information for the purpose of perceiving speech, and it is only in recent years that attention has turned to how we process paralinguistic information in speech to support speaker recognition. Yet in evolutionary terms speech is only a recent addition, and our ancestors have been deducing information from vocalizations to support recognition for millions of years (Belin et al., 2004). The fundamental nature of voice recognition is apparent not just from birth, with new-borns showing a preference for their mother's voice (DeCasper & Fifer, 1980), but there is evidence to suggest that this ability to discriminate voices may in fact emerge in utero (Fifer & Moon, 1994).

1.3.3.1 Representing voices in the brain: underlying neural processing

The voice, like the face, is a unique identity marker and conveys similar identity information including age, gender (this will be discussed in detail in Chapter 6), ethnicity and familiarity as the face and has thus been referred to as an 'auditory face' (Belin et al., 2004). Indeed, there is evidence that the brain may represent voice identity, also referred to as speaker identity, in the same way as the face (Latinus & Belin, 2011; Schweinberger et al., 2008; Zäske, Schweinberger, & Kawahara, 2010). Specifically, Latinus and Belin (2011) demonstrated that voice identity is also represented in terms of a norm-based coding, with each voice identity likely being represented in a similar manner to that described for multidimensional face-space mentioned earlier in the chapter (see Valentine, 1991). Indeed, the neocortex has developed a network that supports voice recognition that is structured along the same principles as the network that supports face recognition. As the FFA has been implicated in the recognition of facial identity (Grill-Spector et al., 2004; Kanwisher et al., 1997; Parvizi et al., 2012), recent findings have revealed voice selective regions, specifically temporal voice areas (TVA), which are located in mid and anterior regions of the superior temporal sulcus (mSTS and aSTS, respectively) (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Belin et al., 2004; Belin, Zatorre, & Ahad, 2002). These areas are sensitive to, among other things, identity information in speech. Similar to the prosopagnosic literature, insult to these areas, particularly when localised to the right hemisphere, can result in deficits in voice recognition known as phonagnosia (Van Lancker & Canter, 1982). In contrast, damage to left hemisphere areas often result in deficits in speech perception (aphasia) but with intact voice recognition (Van Lancker & Canter, 1982). Recent evidence also suggests the existence of *developmental* phonagnosia (Garrido et al., 2009). In parallel, to developmental prosopagnosia, this deficit appears to be largely restricted to the ability to recognise the *identity* of the individual, rather than the sex or the emotion which is depicted in the voice (Garrido et al., 2009).

1.3.4. Interactions between the face and voice in person perception

It appears that the brain has evolved to efficiently process face and voice information in a similar fashion. Recent evidence suggests that vocal and facial information is integrated and shared across modalities at earlier stages of processing (Ghazanfar & Schroeder, 2006; Ghazanfar, 2011; Schweinberger, Kloth, & Robertson, 2011) than were previously assumed (Burton, Bruce, & Johnston, 1990). Pioneering work by von Kriegstein and colleagues reported intriguing findings that the FFA is active during familiar voice recognition, even when there is no visual input (von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005). This has important implications for existing

models of person recognition, which have assumed that the face and the voice are processed independently and only combined at later stages of processing to support recognition in supramodal 'person identity nodes' (Bruce & Young, 1986; Burton et al., 1990; Ellis, Jones, & Mosdell, 1997), see Figure 1.3. Furthermore, it has been reported that even the haptic and the visual modalities may encode face information in a similar way, for the purpose of supporting face recognition (Casey & Newell, 2007). This adds to the growing evidence that activation in unisensory areas of the brain can be modulated by cross modal input (Alais, Newell, & Mamassian, 2010; Driver & Spence, 2000; Driver & Noesselt, 2008; Ghazanfar & Schroeder, 2006) and suggests that the ventral temporal lobe may represent such person identity information *independent* of the input modality.

In addition, von Kriegstein et al. (2005) observed that when the task of the participant was to focus on the identity of the voice, activation between the FFA and mSTS and aSTS, predominantly in the right hemisphere, was apparent. As mentioned such areas are involved in the processing of identity information from the voice (Belin et al., 2011). In contrast, areas in posterior STS are concerned with processing speech content. Such direct sharing of information is advantageous, as it allows for robust person recognition, through cross-modal compensation, under both suboptimal viewing and/or listening conditions. This may be evident with decline in unisensory processing, as occurs with the ageing process (e.g. Maguinness, Setti, Burke, Kenny, & Newell, 2011) or in a dimly lit or noisy room where it may be difficult to see a person's face or to hear the person's voice. However, to date research on ageing has largely focused on the role of dynamic facial information in speech, rather than speaker (i.e. identity), perception. Results from such studies have demonstrated that while speech perception declines with older age dynamic visual information can enhance the perception of the auditory signal (e.g. Maguinness et al., 2011; Sommers, Tye-Murray, & Spehar, 2005). Moreover, in general it appears that multisensory inputs may enhance perception to a greater extent in older adults, relative to younger adults (Laurienti, Burdette, Maldjian, & Wallace, 2006; Peiffer, Mozolic, Hugenschmidt, & Laurienti, 2007), suggesting that the brain may maintain robust perception by combining sensory inputs as unisensory processing declines.

Indeed, von Kriegstein et al.'s (2005) surprising finding of visual activation to an auditory only input, parallels many of the findings from the speech perception literature, which for a long time had been largely viewed as a purely auditory process. However, the results from McGurk & MacDonald (1976) seminal study on the influence of dynamic visual information on heard speech significantly disrupted this assumption. In that study, the authors demonstrated that when conflicting visual and auditory speech cues are presented, for example, when a visually articulated

syllable /ga/ is dubbed with an acoustic syllable /ba/, the observer perceives a novel illusory percept which is a fusion of the auditory and the visual signal e.g. /da/, this effect is commonly referred to in the literature as 'The McGurk Effect'. This new percept is a definitive example of how speech perception is a multisensory process, as what is perceived is neither the unisensory acoustic or visual signal. Other studies subsequently revealed that lip movements specific to speech, in the absence of sound, were sufficient to induce activation in the primary auditory cortex (Calvert et al., 1997).

It has been argued that the addition of visual speech information (viseme) likely decreases the processing load on the auditory cortex by possibly inducing activation in the same neurons in the auditory cortex which are responsible for processing phonetic information in speech (Besle, Fort, Delpuech, & Giard, 2004; Davis, Kislyuk, Kim, & Sams, 2008; Jääskeläinen et al., 2004). In addition, there is also evidence to suggest that gesture information provided by upper limb movements which occur naturally during the production of speech (Hubbard, Wilson, Callan, & Dapretto, 2009), may also help to resolve ambiguities and enhance the processing of the speech signal (Goldin-Meadow & Alibali, 2013; Hubbard et al., 2009). These speech movements from the face and the body have been shown to enhance activation in the auditory cortex, suggesting that they may facilitate speech perception in a similar manner (Calvert & Campbell, 2003; Calvert et al., 1997; Hubbard et al., 2009). This benefit of combining audio and visual signals may be mediated by the saliency of the auditory signal, such that the multisensory benefit is more pronounced when auditory performance is suboptimal (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby & Pollack, 1954).

Furthermore, research reported by von Kriegstein and Giraud (2006) has provided further evidence for the existence of an audio-visual person recognition network. They demonstrated, in an fMRI study that the learning of audio-visual face and voice pairings increased the functional coupling between face and voice areas in the cortex. Furthermore this functional coupling was observed only for natural audio-visual pairings that offer some level of cross-modal redundancy (e.g. a face and a voice versus a phone and a ring tone). Blank, Anwender and von Kriegstein (2011) conducted a follow up study using diffusion tensor imaging (DTI). DTI offers the possibility of assessing the connectivity of brain areas through probabilistic tractography. Blank et al. (2011) demonstrated direct structural connections between the FFA and the voice identity sensitive areas in aSTS and mSTS. Such direct sharing of information between regions may be used to optimize person recognition. Thus, at least for the naturalistic coupling of face and voice stimuli, the multisensory effects may result from direct interactions between voice and face recognitions

areas in the brain, as opposed to top-down modulation from supra-modal areas, as was once thought.

In line with the findings reported by von Kriegstein and colleagues, Schweinberger, Kloth and Robertson (2011) revealed evidence for face and voice interactions occurring at strikingly early stages of processing. They demonstrated through electroencephalogram (EEG) recordings, that event related potential (ERP) responses occur in the frontocentral areas at around 50-80 milliseconds (ms) following audio-visual (AV) presentation. These ERP responses occurred earlier and were more pronounced (super-additive) for AV presentations, compared to visual or auditory unimodal presentations. Schweinberger et al. (2011) argued that such early audio-visual processing is mandatory; however it may not reflect audio-visual integration for the purpose of recognition. For example, detecting AV incongruencies in person recognition emerges at a later stage in the processing stream. AV mismatches in face and voice identity elicited differences in ERP responses at around 250-600 ms (Schweinberger et al., 2011). Nevertheless, results from a familiar voice recognition study by Schweinberger, Robertson, and Kaufmann (2007) demonstrated that individuals found it difficult (slower reaction times) to detect the familiarity of a familiar voice when it was presented with a dynamic face of a non-corresponding identity, suggesting that the neurological effects reported by Schweinberger et al. (2011) can have direct behavioural consequences. However, similar to the face processing literature, where the benefit for motion on unfamiliar face processing is unclear, there is also debate as to whether these dynamic cross-modal interactions are more relevant for *established* familiar identity representations, rather than for establishing new face and voice representations (e.g. Schweinberger, Robertson, & Kaufmann, 2007). This will be returned to in detail and be the focus of Chapter 5.

These audio-visual effects appear to parallel the findings observed by Wallis and Bülthoff (2001) reported earlier in the Chapter. Specifically, Wallis and Bülthoff observed that participants tended to integrate images of different identities into a single face representation when the images were shown in spatiotemporal alignment (i.e. fluid motion). These 'illusory effects' tell us a lot about the perceptual system's sensitivity to change over time information. They also demonstrate the optimal nature in which person recognition has evolved, when we consider that it is unlikely that we would ever encounter a face and a voice articulating speech in spatial and temporal alignment that did not belong to the same identity or encounter a change in facial identity as the a face moved from side to side.

1.4. Outline and Scope of the Thesis

Taken together, the findings discussed above demonstrate that information from the face, voice and body are likely represented and possibly processed in a similar manner in the brain. Indeed, when we consider a typical communication scenario our sensory system is simultaneously bombarded with multiple sensory inputs, for example when 'listening' to someone speak, the voice is accompanied by visual movements of the face and often with movement from the body. This common processing likely assists in sharing information across the senses (Belin et al., 2004; Schweinberger, Kloth, et al., 2011). However, to date these sensory interactions have largely been studied in neurotypical young adult populations only. While these studies have been fruitful for understanding how the brain integrates different sensory cues to identity, they have also left outstanding questions, particularly pertaining to how these cues may interact when learning new identities, rather than accessing established representations of familiar persons, an issue which has been highlighted throughout this chapter.

Furthermore, many of the findings which have challenged the assumption that different sensory inputs (e.g. face and voice) are processed independently which have been held by traditional models of person perception, have largely arisen with recent years. Thus further support or refinement of these emerging theoretical models is warranted. The aim of this thesis is to address these concerns by examining how multiple sensory cues, e.g. face, voice and body, may interact to support person perception in both younger and older adults and in cases of developmental prosopagnosia. Importantly, compared to neurotypical young adults, both cohorts exhibit a decline (i.e. with normal ageing) or a distinct deficit (i.e. developmental prosopagnosia) in their ability to represent identity from static facial cues alone. However, to date, it is poorly understood how impaired face processing, as reported in cases of DP, may impact on the processing of other sensory cues to identity, e.g. the voice or facial motion, or whether person perception in older age may benefit from sensory cue combination, as unisensory processing declines. Such findings would clarify how the interactions between the senses may shape our representation of a person in memory. These questions were tested empirically in the chapters that follow, through a series of behavioural experiments.

Specifically, Chapter 2 examined static face processing in older adults and age-matched developmental prosopagnosics, in order to characterise the nature of the static face processing deficits in these cohorts. Chapter 3 moved on to establish if motion could enhance unfamiliar face processing in older, relative to younger adults, with the rationale that interactions between form and motion cues may be more readily observed in this cohort as static face processing declines.

Chapter 4 then explored how learning a face in motion may affect the representation of the face in two cases of developmental prosopagnosia, a disorder which to date has been examined largely in terms of static face processing ability. Chapter 5 examined how facial motion may affect unfamiliar voice processing in older and younger adults. This chapter proposed that if voice representations are weaker in older adults then the benefit of sensory cue combinations may be more readily observed in this cohort. Chapter 5 also investigated if poor face processing, as evidenced in cases of developmental prosopagnosia, may affect the ability to represent speaker identity by this group. Chapter 6 focused on how information from the voice may modulate the perception of the gender-ambiguous biological motion displays, in younger and older adults and developmental prosopagnosia. Here it was predicted that the sharing of sensory cues to the sex of an individual may assist in the classification of the sex of the stimulus when visual cues alone are insufficient. The theoretical implications of the results observed from the series of behavioural experiments described in the thesis are discussed in detail in Chapter 7. Possible avenues for future research are also explored.

Chapter 2

Characterising Face Processing Abilities in
Neurotypical Older Adults and Age-Matched
Developmental Prosopagnosics

2 Characterising Face Processing Abilities in Neurotypical Older Adults and Age-Matched Developmental Prosopagnosics

2.1. Introduction

The processing of facial information from static images has been studied extensively in neurotypical younger adult populations. These observations have given rise to our understanding of how unfamiliar and familiar faces are perceived, processed and represented in the human brain (Bruce & Young, 1986; Haxby et al., 2002; Valentine, 1991). In parallel, the study of face processing from images in persons with prosopagnosia has also assisted in developing and refining our understanding of the normal mechanisms underlying face processing (e.g. Fox, Iaria, Duchaine, & Barton, 2013; Fox, Iaria, & Barton, 2008). These two populations have also been extensively studied alongside one another. Specifically, many of the standardised tests of face perception and face recognition which have been used to diagnose cases of prosopagnosia (both acquired and developmental), classify abnormal task performance based on normative data collated from younger adult participants (e.g. Duchaine & Nakayama, 2006). However, this approach is somewhat problematic owing to the fact that face processing performance does not remain stable across the lifespan.

As highlighted in Chapter 1, face recognition performance appears to decline as a function of increasing age, with decline evident at around 50 years of age (Bowles et al., 2009). Failure to account for this deterioration can have a significant impact on the ability to accurately and reliably diagnose disorders of face processing. For example, Bowles et al. (2009) demonstrated that comparing performance on a standardised test of face recognition, The Cambridge Face Memory Test (Duchaine & Nakayama, 2006), by people aged 70 years or over to the norm for a 21 year old produced a diagnosis rate of 22.6 % for prosopagnosia in the population tested. Even more worryingly, a comparison of performance across the same populations on the Cambridge Face Perception Test (Duchaine, Germine, & Nakayama, 2007), a standardised test for face perception, produced a diagnosis rate of 60% for prosopagnosia in this older age group. However, when older individuals' scores were compared to their own age group norms, the diagnosis rate for prosopagnosia dropped dramatically to 3.2 % and 8.0 % for the Cambridge Face Memory Test and the Cambridge Face Perception Test, respectively.

The focus of this chapter is two-fold. Through the implementation of a number of standardised face processing tests which examine both face memory and face perception, this chapter hopes to gain insight into what aspects of face processing appear vulnerable to the ageing

process. Specifically, a number of studies have demonstrated that face memory is poorer in older adults (Bartlett et al., 1989; Bartlett & Leslie, 1986; Memon, Bartlett, Rose, & Gray, 2003). Moreover, recent studies have suggested that face perception, specifically matching faces for identity, is poorer in older age, particularly when the task is to match the face across changes in visual appearance (Habak et al., 2008; Lee et al., 2011). Thus it is possible that poor structural encoding of face images may account for some of the effects observed on face memory, with deficits at earlier stage processing affecting later stages of recall. In turn, these investigations will also be valuable when attempting to diagnose for prosopagnosia in older age. This chapter explicitly examines this issue by comparing the performance of two older adults, UM and PL, who presented with suspected developmental prosopagnosia, to the data collected from a sample of older adult participants on an extensive battery of tests designed to assess face processing abilities.

The face processing battery can be divided into two parts. *Part I* of the battery was used to assess age-related changes in the processing of unfamiliar faces in older relative to younger adults. In addition, both prosopagnosics' performance was contrasted with older adults' performance on this part of the battery. *Part II* of the battery was used solely to characterise face processing in prosopagnosia in more depth and it included measures of familiar face recognition. As such older adults' performance on this section was used to assess face processing in prosopagnosia in older age, rather than the data being compared with younger adults' performance. While this thesis is not primarily concerned with object processing, a series of object processing tests are included in this chapter in order to characterise how face specific the processing deficits are in those who present with prosopagnosia in older age. Thus, this chapter does not focus on comparing younger and older adults' object perception or recognition performance, but rather it contrasts the suspected developmental prosopagnosics' object processing results with that of age-matched controls. A full description of the tests which comprised the face and object processing battery are described in detail below.

2.2. Face Processing Battery *Part I*

The Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006) was used to measure age-related changes in face recognition performance. This test uses unfamiliar face images to assess face recognition across learned and novel images of the face (design, stimuli and procedure are described in detail later in the chapter). The use of novel test images is an important factor in any test of face processing. Specifically, face recognition can be said to occur when identity can still be matched despite changes in e.g. viewpoint or expression, as this type of

matching relies on the ability to represent the essential structure of the face, rather than pictorial aspects of the original studied image (Bruce, 1982). The CFMT has been shown to be sensitive to the perceptual processes which are thought to underlie face processing ability. Specifically, performance declines substantially when the face images are shown in an inverted, relative to upright, orientation. Inversion is thought to disrupt the holistic or configural processing of the face image (Maurer et al., 2002; Piepers & Robbins, 2012; Yin, 1969). Furthermore, performance on the CFMT has been shown to be largely dissociable from measures of general object recognition (Bowles et al., 2009) and it has been shown to reliably diagnose face recognition impairments in prosopagnosia (McKone et al., 2007). As a measurement tool it exhibits high internal (Bowles et al., 2009; Wilmer et al., 2010) and external test- retest reliability (Wilmer et al., 2010).

The revised Australian version of the Cambridge Face Memory Test was also administered in the current study (CFMT-Aus; McKone et al., 2011). The CFMT-Aus is identical in format to the CFMT but uses different face stimuli. Both versions were administered as performance on the CFMT has been shown to be affected by the ageing process (i.e. Bowles et al., 2009), however the face images used to develop the CFMT-Aus may more reliably match the ethnicity of the local population. Specifically, the CFMT-Aus was developed using images from the Glasgow Face Database (Burton, White, & McNeill, 2010) (Caucasian faces with Northern European/British Isles ethnicity), while the original CFMT was developed using face images from the Harvard Face Database (Caucasian faces with South European/Middle Eastern ethnicity). Thus in the current context the CFMT-Aus may provide a more reliable measure of ageing effects and has yet to be examined in an older adult cohort. In addition, this version of the test may more accurately characterise a face recognition impairment in prosopagnosics from Northern European/British Isles, as the face stimuli more reliably match the faces they are exposed to on a regular basis (McKone et al., 2011).

The Cambridge Face Perception Test (CFPT; Duchaine et al., 2007) was used to assess the effect of older age on face perception. The test involves sorting a series of novel face images by the degree to which they match a target face. The test contains both upright and inverted trials. As all test and target images are shown simultaneously, this test relies on the ability to discriminate between facial identities, rather than memory for previously learned faces. However, performance on upright trials in this test has been previously shown to relate to face memory performance, suggesting that in neurotypical adults both face perception and face memory share overlapping abilities (Bowles et al., 2009). Although developmental prosopagnosics

often perform poorly on the CFPT, it is important to note that there are also cases where CFPT performance is normal in this group, in spite of a face recognition impairment (Bowles et al., 2009; Bradley Duchaine, Murray, Turner, White, & Garrido, 2009; Lúcia Garrido et al., 2008; Lee et al., 2010). This suggests that in some cases of prosopagnosia abnormalities may also arise at later stages of face processing such as when accessing semantic information about a stored facial representation (Anaki et al., 2007; Towler & Eimer, 2012).

2.3. Face Processing Battery *Part II*

The Famous Faces Test was also administered to the older adult participants and both prosopagnosics. This test was developed onsite and followed a similar format to that previously reported in the literature (e.g. FFT; Duchaine et al., 2007, see Methods section below for further details). The FFT has been argued to be a theoretically robust test as it strongly mimics familiar face recognition in everyday circumstances. Specifically, people have had sufficient exposure to the (famous) facial identity to establish a reliable representation of the face, such that recognition can occur even when presented with a novel i.e. never before seen face image of that particular identity (Bowles et al., 2009).

The Benton Face Recognition Test (BFRT; Benton, Hamsher, Varney, & Spreen, 1994) was administered to both prosopagnosics and the older adult group. This test was largely administered in order to comply with the standard test battery which has been used to establish face processing impairments in prosopagnosia. The test is argued to be a measure of face recognition ability, however it has been shown to have a rather poor success rate in diagnosing cases of prosopagnosia (Duchaine & Nakayama, 2006; Duchaine & Nakayama, 2004; Longmore & Tree, 2013). For example, Duchaine & Nakayama (2006) reported that the BFRT had a 25% success rate in correctly diagnosing prosopagnosia, compared to a 75% success rate for the CFMT. The reason that prosopagnosics can perform so well on this task is due to the fact that both the target identity and the test images (which are to be matched to the target) are shown simultaneously, under free viewing conditions and for an unlimited time. Furthermore, due to the nature of the image set viewers can easily match the identities by matching particular features such as the hairlines or eyebrows of the face images. Duchaine & Weidenfeld (2003) observed that younger neurotypical adults can do the task when only these cues are available; as such the BFRT is arguably a test of image matching rather than face recognition.

2.4. Object Processing Battery

The Cambridge Car Memory Test (CCMT; Dennett et al., 2012) was used to investigate object recognition in prosopagnosics, compared to their older adult, age-matched controls. This test examines the ability to recognise car images, across learned and novel images and follows an identical format to the CFMT. Car stimuli were chosen as they share similar perceptual properties to faces (e.g. the windows, door and wheels occur in a set configuration). Furthermore, cars are three dimensional real world objects which can be shown from different viewpoints (Dennett et al., 2012). As such, the CCMT largely matches the CFMT in terms of cognitive and perceptual demands. Performance on the CCMT is however dissociated from performance on the CFMT, suggesting that each test may harness relatively independent perceptual abilities (Dennett et al., 2012; McKone et al., 2011).

The Birmingham Object Recognition Battery (BORB; Riddoch & Humphreys, 1993) was used to assess object recognition and object knowledge in both prosopagnosic individuals in comparison to controls. From the battery The Minimal Feature View Task (a task designed to assess object matching across novel extrapolated viewpoints); The Picture Naming Task (designed to assess object recognition); The Object Decision Task (designed to assess object recognition through detecting if objects are real or unreal entities); and The Association Match Task (which examines the ability to detect if objects are semantically related) were administered. Although normative data from older adults are available for the BORB tasks, the age range of participants on which the norms are based is rather large. For example, for the Minimal Feature View Task norms are based on data acquired from neurotypical individuals with an age range of 50-80 years. These norms suggest a mean performance score of 23.3/25, with a standard deviation of 2. Based on these data, impairment in performance is only likely to be detected if a score of 19 or below is obtained on the task. As will be seen later in this chapter, performance on this task for older adults with a narrower age range 60-80 years (mean age=69 years, SD=5 years) was rarely below 25, with a standard deviation of 0.4. Thus using the available BORB norms may in fact decrease the likelihood of detecting a co-occurring object processing impairment in prosopagnosia when, in comparison to age-matched controls, an impairment may exist.

2.5. General Method

2.5.1 Developmental prosopagnosia case reports

UM is a 76 year old female, who has a self-reported difficulty in recognising facial identity, in the absence of any neurological damage. This deficit has been consistent throughout her life and she has relied on compensatory strategies for recognition, including environmental context and external facial feature matching, such as hairstyle. UM describes that when viewing faces she can clearly see the eyes, nose and lips and that the features do not appear jumbled, but rather when features are perceived together they do not make one person appear reliably different from the next. She reports that her mother and five of her siblings, as well as her niece, also have this deficit, providing further evidence for the hereditary nature of the disorder (see also Duchaine, Germine, & Nakayama, 2007; Grueter et al., 2007; Grüter, Grüter, & Carbon, 2008; Kennerknecht, Grueter, Welling, & Wentzek, 2006; Lee, Duchaine, Wilson, & Nakayama, 2010). Interestingly, UM became aware of the prevalence of the disorder in her family only when she explicitly asked each member: prior to this UM had assumed that she was the only one affected. UM reports that she can, however, readily detect the expression, attractiveness or sex of a face and that she has no deficit in her ability to recognise objects or animals. However, UM does report spatial disorientation, particularly while driving or navigating through unfamiliar environments. This spatial disorientation appears to be a common feature accompanying developmental prosopagnosia (see Grueter et al., 2007).

PL is a 73 year old female, who also reports a distinct deficit in the ability to recognise facial identity, in the absence of any history of brain trauma or insult. However, unlike UM, PL is unaware of any existing relatives who may also share her condition. She reports both complete failure to recognise and 'false alarm' recognition (i.e. perceiving an unfamiliar other as familiar), however when prompted for identity she has no difficulty in recalling semantic information pertaining to the individual (e.g. relationship to, occupation etc.). Again like UM she does not report a deficit in the ability to identify objects or animals and reports that she can readily recognise the expression and sex of a face. Like UM, PL heavily relies on context for face recognition.

2.5.2. Older adult participants (age-matched control group)

Thirty-eight, community dwelling older adults (age range 60-84 years) took part in this battery of tests assessing face and object perceptual abilities. Of these thirty-eight participants,

twenty-two (9 female, 21 right handed, mean age= 72.1 years, SD= 7.3 years) completed the original Cambridge Face Memory Test (Duchaine & Nakayama, 2006) only. The remaining sixteen older adults (14 female, all right handed, age range=60-80 years, mean age=69.4 years; SD=5.6 years) took part in the larger battery of face and object testing (note they did not complete the original CFMT). Due to time constraints two of these individuals (1 male, 1 female) did not return to complete the entire object testing battery and thus scores for the Birmingham Object Recognition Battery (Riddoch & Humphreys, 1993) are based on fourteen older adults' performance (see Table 1).

2.5.3. Tests for general sensory acuity and cognitive impairment

Importantly, both UM and PL scored within the normal range ≥ 24 of 30 on the Montreal Cognitive Assessment (MoCA) and Mini Mental State Exam (MMSE), both reliable indicators of global cognitive function status (Folstein, Folstein, & McHugh, 1975; Luis, Keegan, & Mullan, 2009; Nasreddine et al., 2005). Both tests assess measures of verbal fluency, delayed recall, executive function and abstraction. Visual acuity, was acceptable for their age as measured at 40cm (mean binocular LogMar score; UM= 0.06, PL=0.2), as was contrast sensitivity (score UM, PL=1.95) measured using the Pelli-Robson Contrast Sensitivity Test.

For the 38 older adults in the age-matched control group, cognitive function was assessed using either the MoCA (Nasreddine et al., 2005) or the Mini Mental State Exam (Folstein, Folstein, & McHugh, 1975) and all participants scored within the normal range (i.e. ≥ 24 of 30) suggesting that there was no evidence of cognitive impairment in these individuals. Of these, the twenty-two older adults who completed the CFMT, had normal visual acuity for their age as measured at 4m (mean binocular LogMar score=0.1, SD=0.1) and normal contrast sensitivity (mean score=1.9, SD=0.1). The sixteen participants who completed all tests had normal visual acuity for their age as measured at 40cm (mean binocular LogMar score 0.13, SD=0.1), and normal contrast sensitivity (mean score=1.94, SD=0.03) measured using the Pelli-Robson Contrast Sensitivity Test.

The experiments, and all experiments reported hereafter in the thesis, were approved by The School of Psychology Research Ethics Committee, Trinity College Dublin (see Appendix A for letter of approval) and conformed to the Declaration of Helsinki. All participants gave their informed, written consent to participate.

2.5.4. Comparison analysis

Performance data from all older adults on the Face Processing Battery *Part I* (see below for tests administered) was compared to available published younger adult norm data (see Table 1.1) using a series of independent *t*-tests. As both the sample size and the variance in performance across both age groups was unequal (see Table 1.1 for sample size and standard deviations from the mean), the group means were compared using independent *t*-tests with Welch's correction. This test takes into account the difference in standard deviations between the two groups as a consequence of the sample size and results in a corrected *t* value and corrected degrees of freedom. The scores for older adults on all administered tests are summarised and compared with available (published) normative data from younger adults in Table 1.1 (see page 42), with age effects shown in bold print.

UM and PL's scores on each administered test in the Face Processing Battery *Part I and II* and the Object Processing Battery were compared to the current older adult age-matched controls using a modified *t*-test described by Crawford, Garthwaite, & Porter (2010). The test is designed to compare whether a single case score is significantly different from a control or normative sample. It is an appropriate test even with a modest control sample size (Crawford et al., 2010; Crawford & Garthwaite, 2002; Crawford & Howell, 1998) and has been used consistently in the prosopagnosia literature (e.g. Duchaine et al., 2007; Lee et al., 2010; Longmore & Tree, 2013; Németh, Zimmer, Schweinberger, Vakli, & Kovács, 2014; Steede, Tree, & Hole, 2007). Along with a *t* score, the test also affords a point estimate of the effect size for the difference between the case and controls (*z*-cc) and an accompanying 95% confidence interval. This analysis was carried out using the revised statistical package (SINGLIMS) (revised see Crawford et al., 2010; original see Crawford & Howell, 1998). The scores for UM and PL and age-matched older adults on all administered tests are summarised in Table 1.2 (see page 54), with impaired performance indicated in bold print.

2.6. Face Processing Battery *Part I*

2.6.1. Cambridge Face Memory Test (Test 1)

2.6.1.1. *Method*

Stimuli and Apparatus

All face stimuli for the CFMT were provided by Duchaine & Nakayama (2006) and consisted of greyscale images of 52 unfamiliar Caucasian (with South European/Middle Eastern ethnicity) male identities, of which 6 were used as target faces and the remainder as distractor faces. Faces images for the CFMT-Aus were provided by the authors of McKone et al. (2011) and consisted of greyscale images of 52 unfamiliar Caucasian (with Northern European/British Isles ethnicity) male identities. As with the CFMT 6 faces were used as target identities and the remainder as distractor identities. All face images used in the CFMT and CFMT-Aus displayed a neutral expression and were edited to include a black mask which occluded non-facial features, e.g. hair. Both tests were identical in design and followed the same experimental procedure. Both tests consisted of 3 sequential testing stages: *Learn Stage*, *Novel Stage* and *Noise Stage* (described in detail in Procedure section below). The CFMT was presented on a computer screen using an executable Java® programme, i.e. '.jar' format, from the authors of Duchaine & Nakayama (2006), this both displayed the experiment and recorded participants responses. However, the CFMT-Aus was presented using a custom-built programme, according to the experimental protocol outlined by McKone et al. (2011), using Presentation® software. Both tests were presented on a Dell OptiPlex 755 PC and displayed on a HP L1710 17" LCD colour monitor. The screen resolution was set to 1024 x 768 pixels. Images subtended a visual angle of approximately 5.4° vertically onscreen, at a viewing distance of 55cm.

Procedure

Each test was administered in 3 consecutive stages; *Learn Stage*, *Novel Stage* and *Noise Stage* (see Figure 2.1 for an illustration of each stage).

Learn Stage

In the *Learn Stage* participants were presented with 3 static images of the same target face which varied by viewpoint, each image was presented for 3,000ms and separated by an inter-stimulus interval (ISI) of 500ms in which a blank screen was presented. Images were always

shown in a set order (30° left viewpoint, 0°front, 30° right viewpoint) and were presented in the centre of the screen. The participant was instructed to memorize each image. After all images were displayed participants were presented with 3 separate, three-alternative forced choice (3AFC) test arrays. In each array, 3 face images were shown side by side, and one image was of the target face while the other two were distractor faces. Each array contained one of the 3 learned viewpoints of the face, and target and distractor faces were always presented from the same viewpoint. In the first, second and third arrays all images were shown from a 30° left view, 0°front view and a 30° right view, respectively. The position of the target face in the array (image left, centre or right of the array) was pseudo-randomised across arrays. For each test array the participant was instructed to choose which of the three faces matched the identity of the previously learned target face. Responses were made by pressing the '1', '2', or '3' button on the keyboard, each of which corresponded to the position of a face (left, centre or right). The 3AFC array remained onscreen until a button response was made. This button response triggered the offset of the current trial and the onset of the next trial. After the three separate arrays were presented a message appeared onscreen indicating that a new identity was to be learned and to press the spacebar to memorize this new face. This procedure was repeated for six separate identities, giving rise to a total of 18 trials. The order of all trials was the same across all participants. At the end of the 18 trials participants were presented with all 6 learned identities again, shown from the same 0°, frontal view as in the Learn Stage. Participants were instructed to memorize all faces; the images remained onscreen for 20 seconds.

Novel Stage

After the 20 second review a series of 30 3AFC arrays were presented. Participants were instructed that in each array one face would always match the identity of one of the 6 target faces previously viewed. The arrays were similar to that reported in the *Learn Stage* with the exception that all test images were novel images of the face which varied in viewpoint or lighting. Each target face appeared 5 times under different viewing conditions; however the same target face did not appear on any 2 consecutive trials. Again, like in the learn stage, a response button press of 1, 2, or 3 initiated the offset of the current trial and onset of the next trial. Each trial was preceded by a 500ms ISI. All trials were presented in a fixed pseudo-randomised order. At the end of the 30 trials participants were again presented with the same 6 identities shown from the 0°, frontal view available in the Learn Stage. Participants were instructed to again memorize all faces and the images remained onscreen for 20 seconds.

Noise Stage

This stage was as described in the *Novel Stage* with the exception that all novel face images contained additional visual noise. The noise was a Gaussian blur which was set at 30% strength (see lower panel of Figure 2.1 for illustration). Each target face appeared 4 times under different viewing conditions, giving rise to a total of 24 trials. All trials were presented in a fixed pseudo-randomised order.

Scoring

The dependent variable was accuracy in recognising the target faces, i.e. correctly identifying the target face from the 3 available images (note by default misses or false alarm responses are recognised as errors in this task). The maximum overall score was 72 correctly identified trials. Performance was also calculated separately for each of the testing stages, *Learn Stage* (max score 18), *Novel Stage* (max score 30) and *Noise Stage* (max score 24).

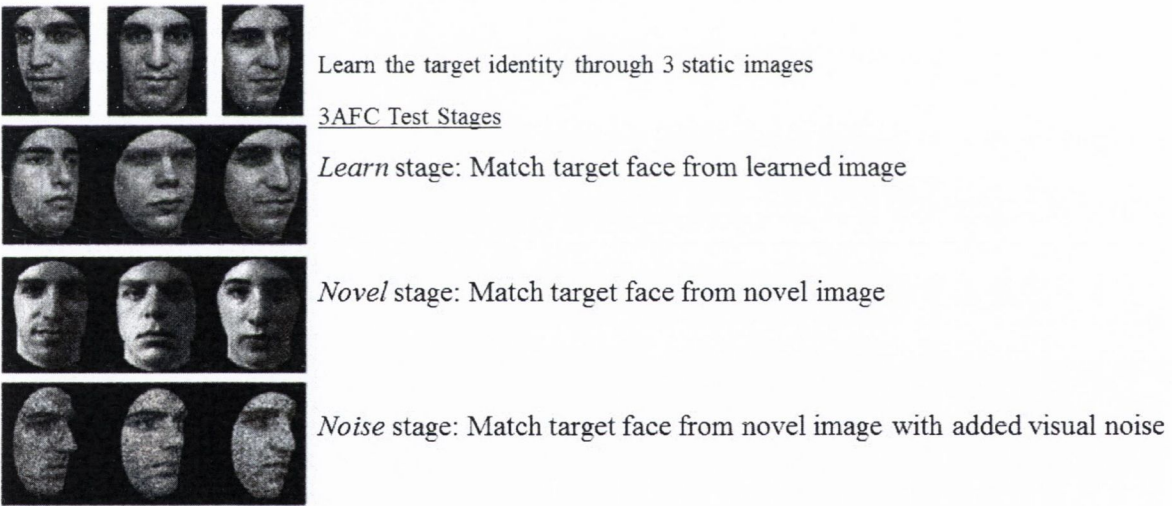


Figure 2.1. Illustration of each test stage of the CFMT. Images adapted from Duchaine and Nakayama (2006).

2.6.1.2. CFMT Results

Performance in younger and older adults

Analysis of the accuracy responses revealed that for older adults their Total score on the CFMT was poorer than that of younger adults [$t(24)=3.55$, $p<0.001$, $r^2=0.5$]. Furthermore, older adults had poorer performance for each stage of the CFMT than younger adults: *Learn* stage [$t(24)=3.55$, $p=0.002$, $r^2=0.34$]; *Novel* stage [$t(55)=5.55$, $p<0.001$, $r^2=0.36$]; *Noise* stage [$t(54)=4.16$,

$p < 0.001$, $r^2 = 0.24$]. These results suggest that older age is associated with a decline in face memory performance. The accuracy scores can be viewed in Table 1.1, page 42.

Performance in prosopagnosics and age-matched older adults

Analysis of the accuracy responses revealed that PL's Total score on the CFMT was worse than that of the older adult age-matched group [$t(21) = -1.68$, $p = 0.05$, $Z_{CC} = -1.72$, 95% CI = -2.38, -1.05]¹. However, UM's overall CFMT performance did not reliably differ from that of the older adult controls [$t(21) = -1.07$, $p = 0.15$, $Z_{CC} = -1.72$, 95% CI = -1.62, -0.55]. Nevertheless, both UM and PL showed impairment on at least one stage of the CFMT, compared to the older adult group. For UM impairment was evident at the *Learn* stage [$t(21) = -3.28$, $p = 0.002$, $Z_{CC} = -3.34$, 95% CI = -4.47, -2.23]; the *Novel* stage [$t(21) = -1.90$, $p = 0.04$, $Z_{CC} = -1.94$, 95% CI = -2.67, -1.20]; but not for the *Noise* stage [$t(21) = 1.34$, $p = 0.1$]. PL showed poor performance in the *Learn* stage [$t(21) = -2.58$, $p = 0.009$, $Z_{CC} = -2.64$, 95% CI = -3.55, -1.71]; while her performance was not significantly different from controls for the *Novel* [$t(21) = -1.68$, $p = 0.16$, $Z_{CC} = -1.03$, 95% CI = -1.06, -0.51] or *Noise* stage [$t(21) = -1.10$, $p = 0.14$, $Z_{CC} = -1.13$, 95% CI = -1.67, -0.56]. The accuracy scores can be viewed in Table 1.2, page 54.

2.6.1.3. CFMT Discussion

Overall performance was worse for the older than younger adults on the CFMT, an effect which was evident when matching the face across learned and novel face images (see Table 1.1). This confirms previous findings that older age is associated with poor face memory (e.g. Bartlett et al., 1989; Bartlett & Leslie, 1986; Bowles et al., 2009; Memon et al., 2003). Compared to the older adult, age-matched controls, PL also demonstrated a significant impairment in face recognition. However, UM's CFMT total score did not reliably differ from that of the older adult control group. It is possible that UM's relatively normal performance on the CFMT original may relate to the stimuli used in this test. It has been suggested that some prosopagnosics may use unusual matching strategies for this task (e.g. eyebrow matching), which can result in relatively normal scores, but with a significant increase in time taken perform the task (Bowles et al., 2009). Further analysis of reaction time² performance revealed that the average time spent by UM viewing each test array before a response was made ($M = 10958\text{ms}$) was significantly different than controls ($M = 5752\text{ms}$, $SD = 1728\text{ms}$) [$t(21) = 2.95$, $p = 0.004$, $Z_{CC} = 3.01$, 95% CI = 2.01, 4.00]. However, PL's reaction time performance ($M = 6713\text{ ms}$) was similar to that of controls

¹ Z_{CC} = Effect Size for difference between case and controls; CI = Confidence Interval

² Note that it is common for prosopagnosic studies to compare response times in DP participants to the mean response time for the control population (e.g. Duchaine, Yovel, Butterworth, & Nakayama, 2006).

[$t(21)=0.54$, $p=0.30$, $Z_{cc}=0.56$, 95% $CI=0.10$, 1.00]. As such, UM may have been using an unusual, non-face specific strategy to produce relatively normal scores on this task, in spite of abnormal face processing.

The CFMT-Aus may more reliably characterise UM's self-report deficit in face recognition as the face images used in the CFMT-Aus more reliably match the faces which she is surrounded by on a regular basis. In parallel, owing to this better ethnicity match it is possible that older adults' performance may differ on the CFMT-Aus relative to the CFMT (McKone et al., 2011; Valentine, 1991).

2.6.1.4. CFMT-Aus Results

Younger and older adults

Analysis of the accuracy scores revealed that the Total score by the older adults on the CFMT was worse than that of the younger adults [$t(15)=3.55$, $p=0.01$, $r^2=0.30$], suggesting a decline in face recognition in older age. However, this decline was evident only when recognising novel images of the face: *Novel* stage [$t(15)=2.10$, $p=0.05$, $r^2=0.20$]; *Noise* stage [$t(15)=2.81$, $p=0.01$, $r^2=0.27$]. Matching identity across identical images of the face in the *Learn* stage was unaffected by the ageing process [$t(15)=1.3$, $p=0.20$, $r^2=0.09$]. See Table 1.1, page 42, for accuracy scores.

Comparison between CFMT and CFMT-Aus in older adults

Older adults' overall Total performance on the CFMT and the CFMT-Aus was compared using an independent Student t test. This analysis demonstrated that performance was significantly better for the CFMT-Aus than the CFMT [$t(36)=-3.88$, $p<0.001$]. However, the variance in performance was similar for both tests. Levene's test for equality of variances was [$F(1,36)<1$].

Developmental prosopagnosics and age-matched older adults

Analysis of Total CFMT-Aus scores, revealed that, relative to age-matched older adults, UM was significantly impaired on this task [$t(15)=-3.47$, $p=0.002$, $Z_{cc}=-3.57$, 95% $CI=-4.92$, -2.21]. When matching the identity to a learned face image UM showed similar performance to the older adult group: *Learn* stage [$t(15)=-1.36$, $p=0.1$, $Z_{cc}=-1.40$, 95% $CI=-2.09$, -0.69]. However, UM was less accurate when matching facial identity across changes in visual appearance: *Novel*

stage [$t(15)=-3.31, p=0.002, Z_{cc}=-3.41, 95\% CI=-4.70, -2.10$] and *Noise* stage [$t(15)=-2.35, p=0.02, Z_{cc}=-2.43, 95\% CI=-3.40$ to -1.43]. See Table 1.2, page 54, for accuracy scores.

2.6.1.5. CFMT-Aus Discussion

Older adults' overall face recognition performance improved for the CFMT-Aus, relative to the CFMT. This effect is consistent with findings in the literature for younger adults (McKone et al., 2011) and it is thought to relate to a better ethnic match between the participant and the face stimuli used. Previous studies suggest that ethnic familiarity improves unfamiliar face perception (Chiroro, Tredoux, Radaelli, & Meissner, 2008; Chiroro & Valentine, 1995). However, older adults' performance for recognising novel images of the face was consistently poor across both the CFMT and the CFMT-Aus. This indicates that age effects are more robustly observed when facial identity is to be matched across changes in visual appearance. The CFMT-Aus also clarified UM's face recognition deficit, as her performance was very poor relative to that of the age-matched older adult group. This parallels the findings of McKone et al. (2011) who reported three cases of prosopagnosia where the participants performed 'normally' on the CFMT but poorly on the CFMT-Aus. It is argued that prosopagnosics may engage abnormal matching strategies for the CFMT, matching distinctive local features for faces from a different, relative to same, ethnicity as the viewer.

2.6.2. Cambridge Face Perception Test (Test 2)

2.6.2.1. Method

Stimuli and Apparatus

Face perception was also assessed using the Cambridge Face Perception Test (Duchaine et al., 2007). All images were made available by the authors of Duchaine et al. (2007) and the test was provided in an executable '.jar' format which both ran the test and recorded participants' responses. Target images consisted of 8 male faces presented from a $\frac{3}{4}$ profile view. Test images consisted of frontal view images which varied in visual similarity to the target image. For each target face 6 images were created by morphing the target face with a different identity, and images were morphed to contain 88%, 76%, 64%, 52%, 40%, and 28% of the target face, giving rise to a total number of 64 test images. See Figure 2.2 for an example of a morphed sequence of images.

Procedure

In each test trial the participant was presented with the target face (at the top of the screen) while 6 test images were presented concurrently below the target image. The participant’s task was to sort the test images from ‘most like the target face’ to ‘least like the target face’ (see Figure 2.2 for an illustration of a sample trial). In the standard CFPT this ranking is made by clicking on an image and using the mouse to place the image in the preferred order e.g. moving the image to the top of the 6 image line up if it looks most like the target face. However, many older participants were not familiar with the function of a mouse, therefore, for the purposes of the present study (and for consistency) all participants reported aloud their preferred ranking of the test images and this ranking was manually executed using the mouse by the experimenter. All trials terminated after 1 minute, after which the next trial was initiated. The test consisted of 16 trials in total. In 8 trials of the total trials the target and test images were shown in an upright orientation, in the remainder test and target images were inverted. The order of the upright and inverted trials was intermixed, and presented in a fixed pseudo-randomised order across participants.

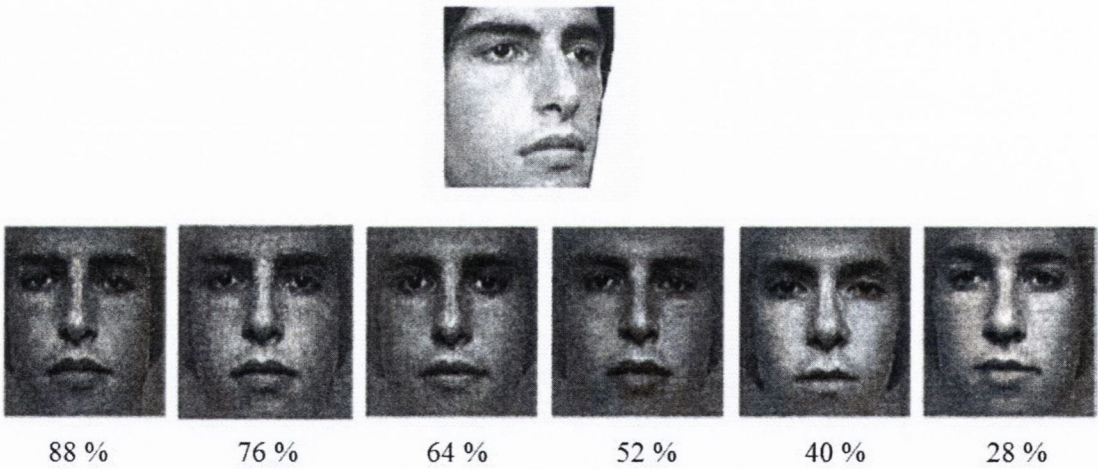


Figure 2.2. Illustration of a correctly ordered sample trial for the CFPT. Images on the lower panel are morphed to contain varying percentages of the target face image, which must be arranged in order of similarity to the target identity (upper panel). Images adapted from Duchaine et al. (2007).

Scoring

Scores for all test trials were calculated by summing the number of deviations in position that a test image was from its correct rank ordering. For example, if a test image which shared the most similarity (i.e. 88%) to the target face was one position from its correct rank order, i.e. placed in 2nd place instead of 1st place, this was counted as one deviation error. If this item was

placed in 3rd place this was counted as 2 deviation errors and so forth. Error scores were calculated for both upright and inverted trials. Note that higher scores indicate poorer performance on the task. A score of 93.3 (for upright or inverted trials) signifies chance performance as it reflects the original pseudo-randomised ordering (i.e. deviation from its correct rank order) of the stimuli. Scores were calculated separately for upright and inverted trials.

2.6.2.2. CFPT Results

Younger and older adults

Analysis revealed that older adults' Total score on the CFPT was poorer than that of younger adults for both upright [$t(25)=5.61, p<0.001, r^2=0.56$] and inverted [$t(26)=3.67, p<0.001, r^2=0.34$] trials (see Table 1.1, page 42). Interestingly, the size of the inversion effect (increase in error scores for inverted relative to upright performance) was similar across both younger ($M=28.4, SD=13.3$) and older ($M=25.38, SD=14.1$) groups [$t(43)=0.85, p=0.39, r^2=0.02$].

Prosopagnosics and age-matched older adults

Compared to the performance of the age-matched older adult group, both prosopagnosics were impaired (to the same degree) on upright trials in this task [$t(15)=1.87, p=0.04, Z_{cc}=1.93, 95\% CI=1.08, 2.76$]. Thus, UM and PL found it difficult to detect subtle differences across face images. However, when images were inverted there was no observable difference between the performance of older controls and that of UM's [$t(15)=0.37, p=0.36, Z_{cc}=0.38, 95\% CI=-0.13, 0.88$] or PL's [$t(15)=0.004, p=0.49, Z_{cc}=0.004, 95\% CI=-0.49, 0.49$] performance (see Table 1.2, page 54).

2.6.2.3. CFPT Discussion

The performance of older adults was worse than that of younger adults on the CFPT, for both upright and inverted trials. This result suggests that older age is also associated with a decline in the ability to discriminate subtle differences between facial identities, even when retrieval demands on memory are low. However, both younger and older adults showed a similar cost in the ability to detect these subtle differences when the face was shown in an inverted, relative to an upright, orientation. This similar cost implies that faces are processed in a holistic or configural manner (for a review see Maurer et al., 2002) even into older age (see also Boutet & Faubert, 2006; Bowles et al., 2009). In terms of prosopagnosics' performance on this task, both UM and PL were poorer than their age-matched controls at discriminating between the face

images for upright, but not inverted trials. Therefore, the deficit observed in the performance of each prosopagnosic appears to be specific to processing which underpins face perception rather than a general deficit in visual perception.

Table 1.1. Accuracy scores* for younger adults and older adults (with standard deviations) on the face processing battery (part I). Scores in bold indicate a significant effect of age.

	Older Adults	N	Younger Adults	N
<i>Face Processing</i>				
<u>Test 1</u> Cambridge Face Memory Test ^a				
CFMT <i>Learn</i>	16.7/18 (1.4)	22	17.8/18 (0.5)	35
CFMT <i>Novel</i>	15.6/30 (3.4)	22	22.0/30 (5.3)	35
CFMT <i>Noise</i>	10.6/24 (3.2)	22	14.9/24 (4.6)	35
<u>CFMT <i>Total</i></u>	43.0/72 (6.4)	22	58.4/72 (9.5)	35
<u>Test 1</u> Cambridge Face Memory Test- Australian ^b				
CFMT-Aus <i>Learn</i>	17.1/18 (1.5)	16	17.6/18 (0.7)	75
CFMT-Aus <i>Novel</i>	23.0/30 (4.4)	16	25.3/30 (4.0)	75
CFMT-Aus <i>Noise</i>	11.7/24 (4.0)	16	14.8/24 (4.0)	75
<u>CFMT-Aus <i>Total</i></u>	51.8/72 (7.5)	16	57.7/72 (7.3)	75
<u>Test 2</u> Cambridge Face Perception Test ^{a*}				
CFPT <i>Upright</i>	50.0/93.3 (11.4)	16	31.4/93.3 (9.1)	29
CFPT <i>Inverted</i>	76.9/93.3 (15.9)	16	59.8/93.3 (13.1)	29

*With the exception of the CFPT where performance is calculated as an error score. 93.3 is the reference point for chance performance on the CFPT, higher scores indicate poorer performance. ^a Younger adult norm data (18-35 years) for the original CFMT and CFPT from Bowles et al. (2009); ^b Younger adult norm data (18-32 years) for the CFMT-Aus from McKone et al. (2011).

2.7. Face Processing Battery *Part II*

2.7.1. Famous Faces Test (Test 3)

2.7.1.1. *Method*

Stimuli and Apparatus

Face recognition was also assessed using a version of the Famous Faces Test (FFT), which was developed onsite. Stimuli consisted of a set of 40 unique static face images of local and international actors, politicians and television personalities downloaded from online sources. Each image was edited in Adobe Photoshop® to include a black mask which occluded any external features of the face, e.g. hair. All images were of a frontal view of the face and were roughly equated for size and subtended a visual angle of approximately 7° horizontally and 9° vertically on the screen at a viewing distance of 60cm. The test was programmed in and run using Presentation® software and the hardware used to drive and display the experiment is as described in the CFMT (see prior procedure section).

Procedure

Each image was presented in the centre of the screen for 5 seconds. After presentation a response screen appeared which prompted the participant to report aloud the name of the individual previously viewed. If participants could not retrieve the name, but were familiar with the identity, they were encouraged to report any unique identifying information about the individual e.g. name of character or television show the actor appeared in. The verbal responses were recorded by the experimenter who was present in the room and responses were later assessed for accuracy. The response screen prompted the participant to press the spacebar to view a new identity, once a response had been uttered. This procedure continued until all 40 face images had been viewed. Images were presented in a fixed pseudo-randomised order across participants.

Finally, when all images had been viewed participants were read aloud a list of names of all the identities presented during the testing phase. Participants were asked to indicate if they were familiar (i.e. had sufficient exposure to the facial identity) or unfamiliar with each identity. Again, all verbal responses were recorded by the experimenter.

Scoring

Two scores were calculated for the FFT (see Table 1.2, page 54). The first score represented the number of identities which were successfully recognised when a list of the names was read aloud to the participants at the end of testing phase ("*familiar with*"). The second score represented the number of faces correctly identified when the identity was presented as a face image during the testing phase ("*correctly identified*"). Items were scored as correct if the participant gave the correct celebrity name or if they provided unique information which could be directly associated with that identity. For example, for actor Daniel Day Lewis a response such as "he starred in the movie *My Left Foot*" was accepted, however generic responses such as 'Actor' were not.

2.7.1.2. FFT Results

Prosopagnosics and age-matched older adults

Both UM and PL were familiar with (i.e. had sufficient experience viewing the facial identity) a similar number of famous identities as the older adult group [$t(15)=-0.17$, $p=0.43$] see Table 1.2, page 54. However, when presented with images of the facial identities, PL was significantly impaired in recognising the face images (4/40 correctly identified) [$t(15)=-2.77$, $p=0.007$, $Z_{cc}=-2.85$, 95% CI=-3.97, -1.72]. UM showed a similar trend for poorer recognition from the face image alone (16/40 correctly identified), relative to the older adult group (mean score= 27.7/40 correctly identified). However, this trend failed to reach statistical significance [$t(15)=-1.37$, $p=0.09$].

2.7.1.3. FFT Discussion

PL found it very difficult to recognise the identity of the famous faces presented in the testing phase of this task, while UM showed a trend for poor recognition from the face alone. Importantly, both prosopagnosics demonstrated the same degree of familiarity with the famous identities used in the current version of the FFT. Thus their deficit in recognition cannot be explained through a lack of exposure to the faces presented in the study, but rather it reflects a poor ability to recognise the identity of an individual in a static face image.

2.7.2. Benton Face Recognition Test (Test 4)

2.7.2.1. Method

Stimuli and Apparatus

Face perception was assessed using the Long Version of the Benton Facial Recognition Test (Benton et al., 1994). The test stimuli consisted of 22 unfamiliar target face images (11 male and 11 female) and 108 test images, which were either identical to the target face (6 target test images) or matched the identity of the target face but varied by viewpoint (24 target test images) or by lighting (24 target test images). The remainder of the face images were used as distractor identities. All images were presented in black and white printed format from the original spiral bound test image booklet (Benton et al., 1994) and each contained a black mask which largely occluded non-facial features such as hair style, but left other features i.e. the ears and hairline visible. All 22 test items were administered in line with the guidelines in the test manual and this procedure is described in detail below.

Procedure

The test consisted of three sections. In the initial section participants were presented with 6 test trials. In each trial a target face was presented from a frontal view and positioned in the centre of the booklet page. Six test images were also displayed concurrently on a separate page which directly faced the page containing the target face. All test images were presented from the same frontal viewpoint (*'matching of identical front-view photographs'* condition). The booklet was held at roughly arm's length from the participant. In each trial participants were asked to point to the test image which matched the identity of the target face (5 were distractor identities). The remaining 18 trials were similar in format with the exception that in each of the trials 3 of the test images matched the identity of the target face (always shown from a frontal view) and either varied by viewpoint (*'matching of front-view with three-quarter-view photographs'* condition) or by lighting condition (*'matching of front-view photographs under different lighting conditions'* condition). Participants were always informed as to whether the target was present in one image (i.e. point to one image) or three images (i.e. point to three images) in the test array. There was no time restriction placed on participant responses. Responses were recorded by the experimenter who administered the test and each response initiated the onset of the next test trial.

Scoring

An overall accuracy score was calculated, with a maximum total of 54 correctly identified test images. Benton et al. (1994) also indicate that in the case of older participants the score must be adjusted so as to account for age and years of education (note that this adjustment is applied to the total score in order to provide a score for clinical diagnosis of perceptual/cognitive impairment i.e. in this case prosopagnosia). For age category 55-64 years, an adjustment of +3 points is made for those with 6-11 years of education and +1 for 12+ years of education. For ages 65-74 years, +4 points is allowed for those with 6-11 years of education and +2 years for 12+ years of education. No adjustment is made for younger age categories (16-54 years) as Benton et al. (1994) found no evidence of age-related performance differences in this age group.

2.7.2.2. BFRT Results

Prosopagnosics and age-matched-older adults

UM's performance on the BFRT was worse compared to that of the older adult controls [$t(15)=-3.12$, $p=0.004$, $Z_{cc}=-3.21$, 95% CI=-4.44, -1.96]. In contrast, PL's performance was not significantly different from that of the older adult group [$t(15)=0.03$, $p=0.49$, $Z_{cc}=0.03$, 95% CI=-0.46, 0.52]. See Table 1.2, page 54, for accuracy scores.

2.7.2.3. BFRT Discussion

In line with previous reports, the BFRT is not sufficiently sensitive at diagnosing low level face processing impairments in prosopagnosia: a deficit in performance was found for UM only. Compared to that of the control group, PL's performance appeared to be within the normal range. It is important to note the PL's performance was almost identical to that of the control group (47/54, compared to a group mean of 46.9/54, $SD=3.7$), and as such her performance did not even show a trend for poorer face matching relative to controls. These conflicting results likely arise as observers can apply unusual strategies to complete the task e.g. feature matching hairlines.

2.8. OBJECT PROCESSING BATTERY

2.8.1. Cambridge Car Memory Test (Test 5)

2.8.1.1. Method

Stimuli and Apparatus

Object recognition was examined using the Cambridge Car Memory Test (CCMT; Dennett et al., 2012). The format and procedure of the test was identical to that described for the CFMT (see Face Processing Battery Part I Methods section above), with the exception that all stimuli were computer generated images of 52 cars, created in 3D Studio Max and made available by the authors of Dennett et al. (2012). As with the CFMT, 6 of the stimuli were used as targets and the remainder as distractors. All car images were the same grey colour and contained no identifying branding or logos. The CCMT was provided in an executable Java® programme, i.e. '.jar' format, from the authors of Dennett et al. (2012): this programme both displayed the experiment and recorded participants' responses. All other details remain as described for the CFMT.

Procedure

As described in the CFMT.

Scoring

As described in the CFMT.

2.8.1.2. CCMT Results

Prosopagnosics and age-matched older adults

Both UM [$t(15)=-0.04, p=0.49$] and PL's [$t(15)=0.11, p=0.46$] Total performance on the CCMT was similar to that of the age-matched older adult control group (see Table 1.2, page 54). This similar performance was evident across all stages of the CCMT. *Learn* stage UM and PL [$t(15)=0.33, p=0.37$]; *Novel* stage UM and PL [$t(15)=0.07, p=0.47$]; *Noise* stage UM [$t(15)=-0.62, p=0.27$] and PL [$t(15)=-0.17, p=0.43$].

2.8.1.3. CCMT Discussion

Both of the prosopagnosic participants showed similar object recognition to their age-matched older group, as assessed by the CCMT. This suggests that, for both prosopagnosics, their deficit in visual recognition is largely face specific. In the neurotypical population performance on the CFMT or CFMT-Aus and the CCMT appear to be largely dissociated, as the correlation in performance across both tasks is poor. For example, based on scores from 67 younger adults, McKone et al. (2011) reported a weak correlation between the CFMT-Aus and the CCMT [$r=0.21$, $p=0.09$]. In terms of the current older adult data, correlations between performance on the CFMT-Aus and CCMT for 16 older adults were rather weak across all stages of the CFMT-Aus and the CCMT. *Learn* stage [$r=0.39$, $p=0.15$]; *Novel* stage [$r=0.33$, $p=0.23$]; *Noise* stage [$r=-0.01$, $p=0.97$]. These results suggest that even into older age both face and object recognition may reflect relatively independent visual abilities.

2.8.2. Birmingham Object Recognition Battery (Tests 6A-6D)

Object recognition and object knowledge was evaluated using the Picture Naming Task (Long Version), the Object Decision Task (Hard 'Version A') and the Association Match Task from the Birmingham Object Recognition Battery (BORB; Riddoch & Humphreys, 1993). Object matching was assessed using the Minimal Feature View Task from the BORB. Note all BORB tasks were manually administered by presenting each participant with the printed images from the original published assessment battery booklet (Riddoch & Humphreys, 1993). The booklet was held at roughly arm's length, under free viewing conditions. There was no time limit for participants to respond. The testing procedure for each task (described in detail below) conforms to the task instructions outlined in the booklet. Note the results for all administered BORB tasks are reported together in the BORB results section towards the end of the description of all BORB tasks.

2.8.3. Picture Naming Task (Test 6A)

2.8.3.1. Method

Stimuli and Apparatus

The Picture Naming Task (Long Version) is designed to assess object recognition. The test consists of 76 images (line drawings or sketches) of objects of which half are classified as animate (e.g. animals) or inanimate (e.g. tools). These two categories are included as animate objects are

argued to contain exemplars which share similar perceptual structures (i.e. animals which share a similar number of parts), while inanimate objects can contain dissimilar perceptual structures (e.g. tools and jewellery may have a dissimilar number of parts). This is an important feature of the task, if it is to be contrasted with face recognition, where all face images share similar perceptual structure across identities. Furthermore, in each of the two categories there was an equal number of words which are from low and high frequency of use categories in written word.

Procedure

Participants were presented with each line drawing of an object, one at a time, and asked to state aloud the name of the presented object. Participants' responses were recorded manually on paper by the experimenter who administered the assessment.

Scoring

Accuracy scores were calculated as outlined in the BORB instruction manual. Specifically, only the object names listed in the manual were accepted as correct responses for each item. The maximum score was 76 correctly identified test images.

2.8.4. Object Decision Task (Test 6B)

2.8.4.1. Method

Stimuli and Apparatus

The Object Decision Task is designed to assess object knowledge. In the test the participant must rely on established object representations to decide whether an encountered image matches (or does not match) a stored representation. Specifically, the test consists of 64 images (line drawings) of objects (animals and tools), of which half were real and half were unreal. Unreal images of objects were constructed by combining one feature of an object with a feature from a different object of the same stimulus class i.e. animal or tool. Half of the unreal items were easy to identify as being an unreal construct (Easy 'Version B') as they contained features from objects which were visually dissimilar e.g. the head of a bird with the body of a donkey. The remaining images were harder to identify as an unreal construct (Hard 'Version A') as each image consisted of combined features which were visually similar to their original state e.g. a snake's head paired with a tortoise's body. Note that Hard Version 'A' was administered in line with recommendations from the original manual. Specifically, if the participant can readily

complete the Hard Version 'A' with minimal error, then it is not necessary to follow up with the easier version of the test, as it can be assumed that this would be readily completed (see also Germine, Cashdollar, Düzel, & Duchaine, 2011). All our older adult participants reliably completed version 'A'. However, if a follow up test was needed administering Easy Version 'B' overcomes any problems associated with image repeats, as Version A and B contain independent image sets.

Procedure

Participants were presented with each image one at a time and asked to state aloud if they thought the item was real or unreal. Participants' responses were recorded manually on paper by the experimenter who administered the assessment.

Scoring

Correct matches were calculated based on the scoring outlined in the BORB instruction manual, which highlighted each correct response. The maximum possible score was 32 for Hard Version A.

2.8.5. Association Match Task (Test 6C)

2.8.5.1. *Method*

Stimuli and Apparatus

The Association Match task is designed to assess object knowledge, specifically stored semantic relationships between objects and their uses. The task consists of 30 test trials. On each trial three images (line drawings) of objects were presented, one at the top of the page (target image) and two test images, shown side by side, at the bottom of the page (90 images are presented in total). The targets images were images of both animate (e.g. humans, animals) and inanimate (e.g. tools, vehicles) objects. One of the two test images was always strongly associated with the target image e.g. a car and a road.

Procedure

Participants were presented with each test trial individually. They were instructed that one of the two images at the bottom of the page was often associated with the image at the top and that their task was to indicate aloud which image was most likely associated with the target. The response was manually recorded by the experimenter who was present in the room.

Scoring

Responses were coded as correct if they matched the correct option which was outlined in the BORB manual. The maximum score was 30 correctly identified test items.

2.8.6. Minimal Feature View Task (Test 6D)

2.8.6.1. Method

Stimuli and Apparatus

The Minimal Feature View task is designed to assess object matching. The test consists of 25 test trials. In each trial the target image (a sketch of an object shown from a standard viewpoint) is displayed at the top of the sheet. Below, two test images are also concurrently displayed, one of which always matches the target object but is shown from a novel viewpoint. The novel viewpoint does not contain the main identifying feature of the object which was displayed in the target image, but the overall shape of the object remains constant. The distractor images were paired by visual similarity to the test target and were matched by viewpoint.

Procedure

Participants were presented with each test trial one at a time. On each trial they were told that two of the images on the page were different views of the same object. Participants were asked to indicate which of the two test images matched the target object. There was no time restriction placed on responses. Once a response was made the next trial was presented. All responses were recorded by the experimenter, who was always present in the testing room.

Scoring

Correct matches were calculated based on the scoring outlined in the BORB instruction manual, which highlighted each correct response. The maximum possible score was 25.

2.8.7. BORB Results

Prosopagnosics and age-matched older adults

The performance of both prosopagnosic participants was comparable to that of their age-matched counterparts' controls across a range of the BORB tasks. Specifically, UM [$t(13)=-0.99$, $p=0.17$] and PL's [$t(13)=-1.6$, $p=0.07$] performance on the Picture Naming Task was similar to

the control group. Their performance on the Object Decision Task was also similar to controls, (UM [$t(13)=1.25, p=0.17$] and PL [$t(13)=-1.43, p=0.1$]), as was their performance on the Minimal Feature View Task [$t(13)=0.55, p=0.30$], where both UM and PL scored a maximum of 25/25. However, while performance for the Association Match Task was normal for UM [$t(13)=-1.17, p=0.30$], PL's results showed evidence of impaired performance on this task [$t(13)=-2.73, p=0.008, Z_{cc}=-2.83, 95\% CI=-4.02, -1.63$]. All scores can be viewed in Table 1.2, page 54

2.8.7.1. BORB Discussion

Both UM and PL's object recognition appeared relatively comparable to the older adult group. However, PL was poorer than her age-matched counterparts at identifying the relationship between objects and their function. As such, using age-matched norms may provide a more reliable indicator of co-occurring deficits in prosopagnosia (however note that the decrease in performance was rather small i.e. 28/30, unlike her performance on a number of face processing tasks). Indeed, PL's ability to match or name objects appears relatively intact. In contrast, her ability to match and name facial identities is significantly compromised (see Face Processing Battery Part I and II). Thus, although it is possible that overlapping deficits may exist in prosopagnosia, they appear to be exaggerated for face compared to object processing.

2.9. Relationship between measures of face processing in older adults

Previous studies have demonstrated that aspects of face processing, such as face perception and face memory, may share overlapping abilities (Bowles et al., 2009; Russell et al., 2009). In terms of the current data, a series of correlational analyses on participants' performance to both the upright and the inverted CFPT trials found evidence to support this assumption. Specifically, performance on the CFPT Upright task was strongly associated with overall performance outcomes on the CFMT-Aus in the older adult group ($r=-0.7, p=0.01$), see Figure 2.3. In contrast, performance on the inverted trials showed no relationship with overall performance on the CFMT-Aus ($r=-0.3, p=0.3$), indicating that it is face-specific processing that explains the relationship in performance across these tasks. Furthermore, it was also observed that scores on each stage of the CFMT-Aus bore no relationship with performance on the corresponding sub-stages for the CCMT (all $ps>0.2$). There was also no evidence that the CFPT (upright or inverted) was related to measures (each test stage and total scores) on the CCMT (all $ps>0.1$). Taken together these results suggest that performance in one visual domain (e.g. object recognition) did not predict performance outcome in another (e.g. face recognition). In contrast, face perception

appears to be strongly related to face memory performance suggesting they share common underlying processing mechanisms.

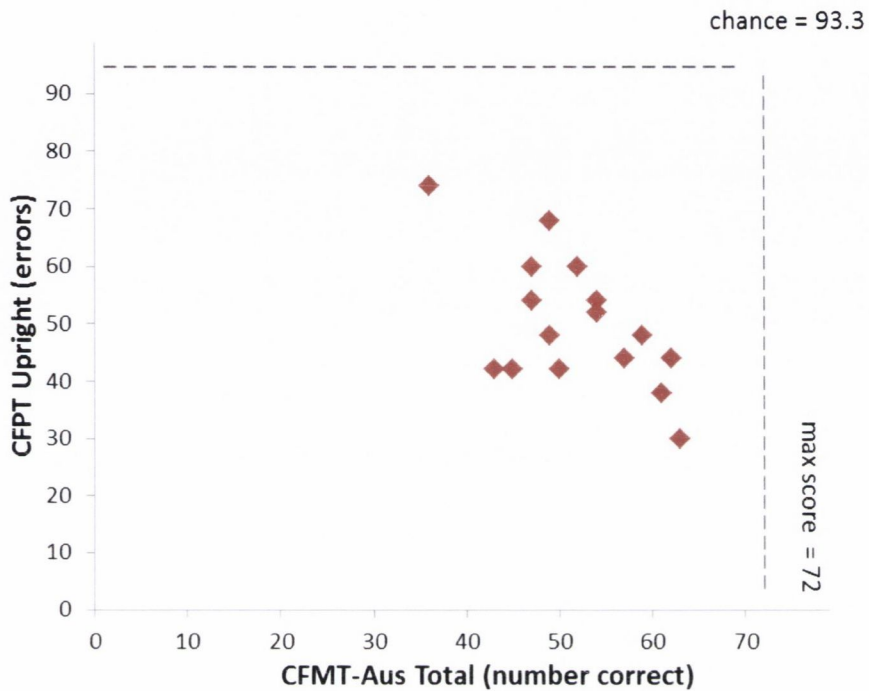


Figure 2.3. Plot showing the correlation between performance on the CFPT and the CFMT-Aus, for the older adult group (N=16). Note that better performance on the CFPT (lower errors) was associated with increased accuracy on the CFMT (higher scores).

Table 1.2. Accuracy* scores for developmental prosopagnosics UM and PL and age-matched older controls (with standard deviations) on each of the face and object tests in the complete test battery. Scores in bold indicate impaired performance.

	UM	PL	Age-Matched Older Adults	N
<i>Face Processing</i>				
<u>Test 1</u> Cambridge Face Memory Test				
CFMT <i>Learn</i>	12/18	13/18	16.7/18 (1.4)	22
CFMT <i>Novel</i>	9/30	12/30	15.6/30 (3.4)	22
CFMT <i>Noise</i>	15/24	7/24	10.6/24 (3.2)	22
<u>CFMT Total</u>	36/72	32/72	43.0/72 (6.4)	22
<u>Test 1</u> Cambridge Face Memory Test- Australian				
CFMT-Aus <i>Learn</i>	15/18		17.1/18 (1.5)	16
CFMT-Aus <i>Novel</i>	8/30		23.0/30 (4.4)	16
CFMT-Aus <i>Noise</i>	2/24		11.7/24 (4.0)	16
<u>CFMT Total</u>	25/72		51.8/72 (7.5)	16
<u>2</u> Cambridge Face Perception Test*				
CFPT <i>Upright</i>	72/93.3	72/93.3	50.0/93.3 (11.4)	16
CFPT <i>Inverted</i>	86/93.3	80/93.3	79.9/93.3 (15.9)	16
<u>3</u> Famous Faces Test				
FFT (<i>familiar with</i>)	38/40	38/40	38.4/40 (2.2)	16
FFT (<i>correctly identified</i>)	16/40	4/40	27.7/40 (8.4)	16
<u>4</u> Benton Facial Recognition Test				
BFRT (<i>age adjusted scores</i>)	35/54	47/54	46.9/54 (3.7)	16
<i>Object Processing</i>				
<u>5</u> Cambridge Car Memory Test				
CCMT <i>Learn</i>	15/18	15/18	14.2/18 (2.4)	16
CCMT <i>Novel</i>	16/30	16/30	15.7/30 (4.3)	16
CCMT <i>Noise</i>	14/24	15/24	15.4/24 (2.2)	16
<u>CCMT Total</u>	45/72	46/72	45.3/72 (6.9)	16
<u>6</u> Birmingham Object Recognition Battery				
6A BORB <i>Picture Naming</i>	67/76	65/76	70.3/76 (3.2)	14
6B BORB <i>Object Decision (Hard A)</i>	22/32	29/32	25.7/32 (2.5)	14
6C BORB <i>Association Match</i>	29/30	28/30	29.7/30 (0.6)	14
6D BORB <i>Minimal Feature View</i>	25/25	25/25	24.8/25 (0.4)	14

*With the exception of the CFPT where performance is calculated as an error score. 93.3 is the reference point for chance performance on the CFPT, higher scores indicate poorer performance.

2.10. Explaining older adult variance in face processing performance

Although there was a larger degree of variance in performance across the administered standardised face processing tests in the older adult group, relative to the available younger adult norms (see Table 1.1 for standard deviations in performance), we found no evidence that the individual age of the participant contributed to this variance. Specifically, across all tasks (total scores and scores for each sub stage) there was no correlation in the older adult group between age in years and test performance (all $ps > 0.1$). This result highlights the fact that the age range used in the current thesis is appropriate for establishing between group effects (e.g. younger versus older adults). However, at an individual level task performance cannot be predicted by the age of the participant in the older adult group. Indeed, other studies have demonstrated that other factors, such as underlying neural activity in the face processing network, may be more reliable indicators of face recognition performance in this population (e.g. Lee, Grady, Habak, Wilson, & Moscovitch, 2011).

In addition, there was no evidence that measures of global cognitive function, as assessed by the MoCA (Nasreddine et al., 2005) or MMSE (Folstein et al., 1975), bore any relationship to any of the attained scores (or scores for each sub stage) on the face processing battery (all $ps > 0.3$). This result is in line with previous studies which have assessed face recognition and cognitive ability across the lifespan (Bowles et al., 2009; Hildebrandt et al., 2011). This suggests that age-related changes in face processing are unlikely to be mediated, and are likely to be dissociated (Hildebrandt et al., 2011), from general changes in cognitive function which accompany older age. In terms of visual function there was no evidence that contrast sensitivity was related to face processing performance across the range of tasks (all $ps > 0.3$). It was, however, observed that measures of near visual acuity were related to performance outcomes on the overall CFMT-Aus ($r = -0.6$, $p = 0.01$), such that a decline in visual acuity (higher LogMar scores) was negatively correlated with performance on the face memory task (lower accuracy). However, this effect was modulated by performance on one stage of the CFMT-Aus, the *Noise* stage, ($r = -0.54$, $p = 0.04$). There was no evidence of a relationship between visual acuity and performance scores in the *Learn* ($p = 0.6$) or the *Novel* ($p = 0.1$) stages of the task. This effect can thus be attributed to the addition of visual noise in this condition. Although presenting a face in visual noise is argued to isolate the processing mechanisms which are thought to underlie face perception (Mc Kone, Martini, & Nakayama, 2001), it is likely that visual acuity mediated the ability to reliably discriminate the face images in this condition. This highlights the importance of controlling for sensory acuity when using degraded visual stimuli in this population, as effects attributed to age

may also be mediated by poor sensory acuity. Importantly these effects were not evident when the face stimuli were super-threshold (see Learn and Novel stage), as such these stimuli may be effective in establishing changes in face processing with age.

2.11. Summary

The results from this chapter confirm that older age is associated with a decline in face recognition. The effect of age appears to be most reliable when the face image is to be recognised across changes in visual appearance. As such, older age is associated with a decline in the ability to represent the essential structure of the face, rather than pictorial aspects of the original studied image (Bruce, 1982). Moreover, older age was also associated with a decline in face perception. Compared to younger adults, older adults found it more difficult to detect subtle differences between facial identities. In this age group, face perception and face memory performance also appeared to be related, such that decreased sensitivity to detecting differences between facial identities was associated with a decrease in recognition accuracy at later stages of recall. Therefore, it is possible that poor structural encoding of facial identity at early stages of processing may affect the representation of the face in memory (see also Lee, Smith, Grady, Hoang, & Moscovitch, 2014). However, the results also emphasised that the way in which faces are perceived, i.e. holistic processing, remains relatively stable across the lifespan and as such age effects appear to reflect changes in face processing, rather than a decline in general visual function.

The results from this chapter also demonstrated that when controlling for this age related decline it is still possible to accurately detect and characterise face processing impairments in older adults who present with suspected developmental prosopagnosia. Indeed, compared to their older adult counterparts, both UM and PL demonstrated significant impairment in the ability to match and also to recognise facial identity. As both perception and recognition of facial identity was impaired, it is likely that both prosopagnosics have deficits in their ability to encode the structural representation of the face, affecting later stages of recall. This suggests that in this case both prosopagnosics have perceptual deficits which can be characterised as apperceptive (Németh et al., 2014; Towler & Eimer, 2012; Towler et al., 2012). Other studies have demonstrated that some prosopagnosics can, however, accurately perceive and match facial identity, but still exhibit significant impairments in face memory (e.g. McKone et al., 2011). This emphasises the heterogeneous nature of the deficits expressed in those with developmental prosopagnosia and the importance of attempting to characterise the nature of these deficits. It has been suggested that the heterogeneity of the disorder may result from dysfunction at

different stages along the face processing network (Dinkelacker et al., 2011; Fox et al., 2008; Grüter et al., 2008; Le Grand et al., 2006; Schmalzl, Palermo, & Coltheart, 2008; Towler & Eimer, 2012).

Chapter 3

The Effect of Motion on Face Perception in Younger and Older Adults

3 The Effect of Motion on Face Perception in Younger and Older Adults

3.1. Abstract

Faces are inherently dynamic stimuli. However, face perception in younger adults appears to be mediated by the ability to extract structural cues from static images and a benefit of motion is inconsistent. In contrast, static face processing is poorer and more image-dependent in older adults. This chapter therefore compared the role of facial motion in younger and older adults to assess whether motion can enhance perception when static cues are insufficient. In a series of studies, older and younger adults learned faces presented in motion or in a sequence of static images, containing rigid (viewpoint) or non-rigid (expression) changes. Immediately following learning, participants matched a static test image to the learned face which varied by viewpoint (Experiment 1) or expression (Experiment 2) and was either learned or novel. First, an age effect was observed with better face matching performance in younger than in older adults. However, it was found that face matching performance improved in the older adult group, across changes in viewpoint and expression, when faces were learned in motion relative to static presentation. There was no benefit for facial (non-rigid) motion when the task involved matching inverted faces (Experiment 3), suggesting that the ability to use dynamic face information for the purpose of recognition reflects motion encoding which is specific to upright faces. These results suggest that ageing may offer a unique insight into how dynamic cues support face processing, which may not be readily observed in younger adults' performance.

3.2. Introduction

All faces share the same basic featural arrangement across individuals, yet we have the ability to recognise each face as unique. Furthermore, we can recognise the face of a familiar individual across large variations in visual appearance (Bruce, Henderson, Newman, & Burton, 2001; Burton, Wilson, Cowan, & Bruce, 1999). In contrast, our ability to recognise unfamiliar faces (i.e. faces previously viewed once or twice) is more prone to error (Bruce et al., 2001b) particularly when a change in visual appearance, such as expression or viewpoint, occurs (Bruce, 1982; Burton, Wilson, Cowan, & Bruce, 1999; Hancock, Bruce, & Burton, 2000; Newell, Chiroro, & Valentine, 1999). Such effects suggest that unfamiliar face recognition is more image-dependent than familiar face recognition (Bindemann & Sandford, 2011; Bruce, 1982; Hancock et al., 2000; Longmore, Liu, & Young, 2008).

However, it is important to note that many studies which have observed image-dependency in unfamiliar face processing have employed static face images as stimuli (e.g. Longmore, Liu, & Young, 2008). Yet, in the real world faces are most likely seen in motion across a range of viewpoint and expression changes. As mentioned in Chapter 1, the use of static images may in part be driven by models of face processing which have argued that facial form (i.e. the unchangeable aspects of the face) and facial motion provide two distinct sources of information, with the former representing facial identity and the latter representing emotion and communicative signals such as speech (Bruce & Young, 1986; see Calder & Young, 2005 for more recent review on how identity and expression recognition may interact). Yet, emerging evidence from studies based on younger adults suggests that motion cues may be relevant to, and enhance the perception and recognition of, facial identity (Lander & Bruce, 2000, 2003; Lander, Christie, & Bruce, 1999; Lander & Chuang, 2005; Pilz, Thornton, & Bülthoff, 2006; Pilz, Bülthoff, & Vuong, 2009; Thornton & Kourtzi, 2002; Bennetts, Burke, Brooks, & Robbins, 2010).

Several studies on younger adults have reported that both rigid and non-rigid motion (O'Toole et al., 2002) of the face can improve recognition relative to static images. Rigid motion refers to movement of the whole head, as when the head rotates from left to right, whereas, non-rigid motion refers to movements made through internal deformations of the face, such as those associated with changes in expression or speech production (Bülthoff, Cunningham, & Wallraven, 2011; Knappmeyer, Thornton, & Bülthoff, 2003; O'Toole et al., 2002). O'Toole and her colleagues proposed that motion may facilitate familiar and unfamiliar face recognition in a number of ways (O'Toole et al., 2002; see Roark, Barrett, Spence, Abdi, & O'Toole, 2003 for review).

For example, rigid motion may help create a robust representation of the face in memory by providing more information about its three-dimensional structure across multiple viewpoints (Farivar, Blanke, & Chaudhuri, 2009), enhancing unfamiliar face processing (representation enhancement hypothesis). Indeed, Pike, Kemp, Nicola, & Phillips (1997) provided evidence in support of this by reporting that learning unfamiliar faces moving in a rigid manner resulted in better recognition of those faces than faces learned through either a single static image or a sequence of multiple static images. However, although Lander & Bruce (2003) also reported an advantage on recognition accuracy for faces previously learned moving in a rigid manner compared to a single static image, they failed to observe an advantage for learning dynamic over static faces when the static condition contained multiple images. Thus, unlike Pike et al. (1997), Lander and Bruce proposed that the provision of information about different views of a face during learning, from either dynamic or multiple static images, may be sufficient to observe a subsequent benefit on recognition. Nevertheless, earlier findings reported by Lander & Bruce (2000), found better recognition performance when the spatiotemporal alignment of the facial motion was consistent relative to when it was disrupted. This suggests that face perception appears to be sensitive to the change in the facial information over time and not only the quantity of static information exposed during the motion sequence (see e.g. Schultz, Brockhaus, Bülthoff, & Pilz, 2013 for a discussion of the neurological underpinnings of such effects).

In contrast, non-rigid motion may be more relevant for familiar face recognition. Specifically, O'Toole et al. (2002) suggested that non-rigid motion may benefit face recognition through the extraction of idiosyncratic motion cues which are built up over time (i.e. repeated exposure to a face), and are processed independently from the visual form of the face (known as the supplemental information hypothesis). Therefore, non-rigid motion may provide a unique dynamic signature of a person's identity such that this motion can support face recognition independently from the structural information of the face. For example, the results of several studies suggest that motion benefits the recognition of familiar faces under degraded visual conditions, such as when form cues have been blurred or pixelated (Knight & Johnston, 1997; Lander et al., 1999; Lander & Chuang, 2005).

However, there is some, albeit conflicting evidence, to suggest that non-rigid motion may also be relevant to the processing of unfamiliar faces, even when face images are not degraded (e.g. Lander & Bruce, 2003). For example, Thornton & Kourtzi (2002) reported faster reaction times to match identity across changes in expression when the target face was presented with a dynamic facial expression (non-rigid motion) relative to a single static image depicting the expression apex. Interestingly, they reported no advantage for non-rigid facial motion when the

task was changed to an expression matching task, which led the authors to conclude that a general arousal state created by the motion does not provide a simple explanation of the observed effects. Pilz et al. (2006) extended these findings by providing evidence that the presence of non-rigid motion in the learned face facilitated the matching of the test image even when a change in both viewpoint and expression occurred between learning and test. Schultz & Pilz (2009) later demonstrated that, compared to a static presentation, both form (e.g. fusiform gyrus) and motion (i.e. Superior Temporal Sulcus or STS) processing areas of the face network showed enhanced cortical activation when viewing a face moving in a non-rigid manner. Taken together such findings suggest that motion may increase the efficiency by which faces are perceived, possibly by allowing for faces to be represented in memory in a manner which is invariant to incidental changes in image properties (i.e. *face based* rather than *image based* encoding). However, it is important to note that other studies have failed to find a benefit for non-rigid motion on unfamiliar face recognition (Christie & Bruce, 1998; Hill & Johnston, 2001). For example, Christie & Bruce (1998) observed that learning an unfamiliar face moving in a non-rigid manner (expressive movements) did not enhance subsequent recognition relative to when the face was learned from a number of static images. In fact, they reported a trend for poorer recognition performance for novel expressions of the face, when the face had been learned in motion.

One possibility for such conflicting results may relate to the fact that in general younger adults perform well when matching and representing facial identity (Otsuka et al., 2009). In this age group, where perceptual abilities have matured and are optimal, learning a face from static images may provide enough structural information to create a robust, high quality facial representation, and any advantage for dynamic information may be obscured (Otsuka et al., 2009). Therefore, it is possible that motion is relevant for the processing of unfamiliar faces, but the advantage for motion may be more readily observed in a less developed face processing system. This idea is supported by recent evidence which suggests that motion may enhance face recognition in infancy (Otsuka et al., 2009). Specifically, Otsuka and colleagues observed that 3 and 4 month old infants showed enhanced unfamiliar face recognition when a face was learned moving in a non-rigid manner. This dynamic advantage was observed particularly when recognising novel face images and performance remained superior even when the static learning condition was equated for structural information (i.e. multiple images).

In parallel, motion effects may also be more readily observed during the ageing process. Specifically, older age has been shown to be associated with a performance decline in recognising unfamiliar faces (Bartlett, Leslie, Tubbs, & Fulton, 1989; Bartlett & Leslie, 1986; Boutet & Faubert,

2006; Bowles et al., 2009; Grady, 2002). For example, in comparison to younger adults, face memory in older adults is marked by a reduction in accuracy and an increase in false alarm responses (Bartlett et al., 1989; Bartlett & Leslie, 1986; Edmonds, Glisky, Bartlett, & Rapcsak, 2012; see also Lee, Smith, Grady, Hoang, & Moscovitch, 2014). Ageing effects are also evident in face matching tasks over less extended intervals between learning and test where retrieval demands on memory are low (Boutet & Faubert, 2006; Cronin-Golomb, Gilmore, Neargarder, Morrison, & Laudate, 2007; Grady, 2002; Grady et al., 1998, 2000; Konar, Bennett, & Sekuler, 2013). For example, Boutet & Faubert (2006) observed age-related deficits in performance in both a face matching task, where participants matched a test image to a target face shown approximately two seconds before, and in a facial recognition task involving more long-term memory. Moreover, Grady et al. (2000) demonstrated that older adults were less accurate than younger adults in matching facial identity, even when the target and novel test images were presented concurrently.

Ageing effects are particularly pronounced when matching for identity across face images which vary in visual appearance (Habak, Wilkinson, & Wilson, 2008; Lee, Grady, Habak, Wilson, & Moscovitch, 2011). For example, Habak et al. (2008) observed that older adults' performance was worse than that of younger adults, in a task which involved matching identity across face images which differed by viewpoint (e.g. frontal to 20° side view). Furthermore, this age-related cost was evident even when the exposure time to the target face image was increased. Habak et al. (2008) suggested that such effects may be mediated by age-related changes in the neural activity in face processing regions which link facial identity across views. Indeed, many studies have demonstrated marked changes in activation in face processing regions in the older, relative to the younger adult brain, when viewing unfamiliar faces (Burianová, Lee, Grady, & Moscovitch, 2013; Grady et al., 1994; Grady, 2002; Grady, McIntosh, Horwitz, & Rapoport, 2000; Lee, Grady, Habak, Wilson, & Moscovitch, 2011). For example, Lee et al. (2011) used functional magnetic imaging (fMRI) to demonstrate that, unlike younger adults, older adults failed to show neural adaptation to repeated facial images of the same identity, shown from the same viewpoint, in the fusiform face area (FFA), an area in the ventral stream which is known to be involved in the processing of facial form information (Grill-Spector et al., 2004; Kanwisher et al., 1997; Parvizi et al., 2012). This finding suggests that the selective properties of neurons in the fusiform face area (FFA) (e.g. to a specific viewpoint of a face) may diminish with the ageing process. In addition, the authors reported less lateralization in the fusiform gyrus (see also Daniel & Bentin, 2012) and recruitment of additional frontal areas of the brain, with older age. The recruitment of such additional areas may help compensate for age-related decline in the face processing network. Interestingly, Lee et

al. (2011) noted that 'high performing' older adults (i.e. based on faster reaction time performance in a face matching task conducted outside the scanner) recruited a similar face network as younger adults. However, activation in this network was not sufficient to support the matching of identity across views, as older adults who engaged this network did not reliably exhibit faster response times in this condition. Consistent with the conclusions drawn by Habak et al. (2008), the results observed by Lee et al. suggest that the ability to represent facial identity across different views diminishes with age and that this performance decrease likely emerges due to functional changes in the face processing network.

The assertion that ageing is associated with more view-dependent face recognition and face perception (e.g. matching) is, however, largely based on studies which have used static face images as stimuli (e.g. Bartlett et al., 1989; Boutet & Faubert, 2006; Bowles et al., 2009; Habak et al., 2008; Lee et al., 2011). It is possible that motion may represent a particular challenge for face perception in older adults, as more general motion processing abilities e.g. detecting and discriminating the speed of moving stimuli, have been shown to be poorer in older than younger adults (Roudaia, Bennett, Sekuler, & Pilz, 2010; Snowden & Kavanagh, 2006). On the other hand, as ageing is associated with reduced performance in the perception of faces from static images (e.g. Boutet & Faubert, 2006; Habak et al., 2008), it remains possible that the additional information available in moving relative to static face images may benefit face perception in older adults.

3.2.1. Outline of the Experiments

The series of three experiments reported below investigated whether unfamiliar face matching performance in older adults benefits from viewing moving over static images of faces during learning. In addition, in separate experiments the role of rigid (Experiment 1A, 1B) and non-rigid (Experiment 2) motion on face matching performance was examined using an identical face matching paradigm for both types of motion. It was then investigated whether the ability to effectively use dynamic facial information for the purpose of matching identity may reflect motion encoding which is specific to upright faces (Experiment 3).

3.3. EXPERIMENT 1A

In Experiment 1A both younger and older adults learned unfamiliar faces presented as moving (rigid) or static image sequences and were required to match the learned identity by choosing from two available static test images which were either of a learned or novel viewpoint

of the face. An immediate matching paradigm was chosen as it has been previously shown to be sensitive to the effects of motion on unfamiliar face processing. It is argued that more explicit 'old/new' recognition paradigms may in fact bias the observer to adopt a strategy for memorising pictorial, i.e. image specific, content (Pilz et al., 2006). As the current studies assessed the matching of unfamiliar faces over very brief intervals, these tasks tap into the mechanisms involved in face perception more than face recognition (which is largely examined over longer intervals). Other sequential face matching studies (e.g. Thornton & Kourtzi, 2002) have however typically employed paradigms where the participant is required to indicate if a single test and target face are the same or a different facial identity. However, as highlighted above and in Chapter 1, a number of face recognition studies have indicated that older age is associated with a significant performance bias to state that a *novel* facial identity matches a previously learned facial identity (i.e. high false alarm responses for same versus different or old versus new, e.g. see Edmonds et al., 2012) a two-alternative choice (2AFC) paradigm was employed in the series of face matching studies reported below in an attempt to overcome this bias.

Specifically, these experiments compared face matching performance for faces learned either dynamically (i.e. a video clip) or from a sequence of three static images which were extracted from the video sequence. The current studies attempted to control for the role that additional static information in motion may play in the dynamic enhancement previously reported in face matching tasks by including multiple static images, rather than just a single static image, during the learning phase. Previous matching studies have largely presented one face image in the static condition, which is typically presented for the same duration as the video clip (Pilz et al., 2006; Thornton & Kourtzi, 2002). Thus it remains plausible that it may be the additional structural information available in the motion, e.g. additional viewpoints of the face, and not the motion itself, which facilitates face perception (e.g. Lander & Bruce, 2003). It was expected that if the combination of motion and form information leads to a more robust representation of facial identity, then face matching performance in the older adult group may improve when target faces are presented as dynamic rather than static sequences. Moreover, this benefit may be particularly evident when the task is to match a novel view of a face to its learned identity. However, it is possible that younger adults' performance may be similar across both static and dynamic sequences, as both conditions contain multiple views of the face. The diversity in facial information available in the static condition may be sufficient to reduce viewpoint-dependency effects in the matching performance of the younger adult group.

3.3.1. Method

Participants

Fifteen³ younger adults, recruited from the student population at Trinity College Dublin (14 female, 11 right handed, mean age= 22.9 years, SD= 7.2 years) participated in this experiment and were given course credits in return for their time. All younger adults reported normal or corrected to normal vision. Sixteen, community-dwelling, older adults were also recruited for this task. All participants were naive as to the purpose of the experiment. Data collected from two of the older adults was subsequently removed from analysis for the following reasons; one individual had cataracts at the time of testing, while the other individual showed evidence of cognitive decline and scored 20 of 30⁴ on the Montreal Cognitive Assessment (i.e. MoCA, Nasreddine et al., 2005). Data from the other 14 older adult participants were included in subsequent analyses (5 female, 13 right handed, mean age= 72.9 years; SD= 7.9 years).

For the older adults, cognitive function was assessed using either the MoCA or the Mini Mental State Exam (Folstein et al., 1975) and all participants scored within the normal range (i.e. ≥ 24 of 30) suggesting there was no evidence of cognitive impairment in these individuals. All older participants had normal visual acuity (mean binocular LogMar score = 0.06, SD= 0.1) and contrast sensitivity (mean score= 1.88, SD= 0.1, assessed using the Pelli-Robson Contrast Sensitivity Test) for their age. All participants wore appropriate corrective lenses, if needed, at the time of testing.

Stimuli and Apparatus

The stimulus set consisted of 20 videos of faces shown moving in a rigid manner, and were taken from the Amsterdam Dynamic Facial Expression Set (van der Schalk, Hawk, Fischer, & Doosje, 2011). Each video contained an individual actor's face (9 female, 11 male) moving between a frontal and 45° view. Video sequences began either with the actor directly facing the front (0°) and then turning their head to their right (45°), or with the actor facing towards their left (45°) and turning to face the front (0°). For convenience, these conditions are referred to as 'turn to right' and 'turn from left' respectively. In all rigid motion sequences, the actor always displayed a neutral expression and the actor's hair was visible during filming. All videos were edited in VirtualDub open source video editing software (www.virtualdub.org) to be 29 frames in

³ The number of younger and older adult participants recruited for each experiment, and all experiments hereafter, was based on previously reported *N* numbers for similar behavioural experiments e.g. Lee et al. (2011) and Pilz et al. (2006).

⁴ It is argued that MoCA scores of ≤ 23 are indicative of cognitive decline (Luis, Keegan, & Mullan, 2009).

length and were displayed at a rate of 25 frames per second. All video clips were 1,125ms in duration. Three static stimuli were extracted from the 20 edited videos. The static stimuli consisted of the first frame, a mid-sequence frame (on average frame 15) and the last frame of the video clip (see Pitcher, Dilks, Saxe, Triantafyllou, & Kanwisher, 2011 for similar procedure). The display duration of each static image was 375ms. The three static images were shown in sequence (i.e. first, mid, last frame), and an inter stimulus interval (ISI) of 150ms was inserted between image presentations to avoid potential effects from implied motion. Thus for each facial identity a dynamic and a static sequence was created (40 target stimuli). The videos and the static images were presented for an equal amount of time (1,125ms) during the learning exposure.

The static images used as the test stimuli were extracted from the first and final frame of either the *turn to right* or *turn from left* video and were either contained within the static learning condition (learned) or not (novel). As each video sequence always contained the same frontal view, learned images in the test condition consisted of a frontal viewpoint, while novel images were that of an unseen 45° view (either left or right-facing). All test images were edited in Adobe Photoshop® to include a black mask which occluded non-facial features such as hairstyle. The experiment was programmed in and run using Presentation® software and was presented on a Dell OptiPlex 755 PC and displayed on a HP L1710 17" LCD colour monitor. The screen resolution was set to 1024 x 768 pixels and the video clips and static images were presented at a resolution of 540 x 432 pixels. Participants were positioned in front of the screen at a distance of approximately 60 cm. As such, the videos and static images subtended a visual angle of approximately 19.1° horizontally and 15.3° vertically on the screen. The experiment was conducted on-site, in a testing laboratory.

Design

The experiment was based on a mixed, factorial design with Age (younger or older adults) as a between-subjects factor and both Target face (dynamic or static) and Test view (learned or novel) as within-subject factors. The experimental protocol was based on a delayed matched to sample, two alternative forced choice paradigm, in which a target face was first presented in the centre of the screen and was followed by a pair of face images shown simultaneously, side by side. The participant was required to learn the target face and then to choose which of the two test faces matched the identity of the previous target face. The dependent variables were response accuracy and response time taken to match the identity of the target and test face.

The experiment consisted of 80 trials in total. In half of the trials the target face was presented dynamically (i.e. a video), while in the other half of the trials 3 sequential static images of the target face were shown. The two test face images were always presented from the same view, and one always matched the target face. In half of the trials the test images were presented from a frontal, or 0° view, which was a view always shown during the presentation of the target face (i.e. learned). In the other half of trials, test images were of a 45° view of the face which was a view not previously presented in the target image (i.e. novel). During the presentation of the target face, half of the participants were shown the *turn to right* sequence in the learning phase and then presented with the 45° left view of the actor's face as the novel viewpoint in the test, while the other half were shown the reverse i.e. the *turn from left* sequence as the target face and then the 45° right view of the actor's face as the novel viewpoint.

All 80 trials were presented in a single test block and the order of trials was randomized across participants. One of the test faces on the left and right side of the screen was always a correct match, while the other was a distractor face (see Figure 3.1 for an illustration of a sample trial). The distractor face was always matched to the sex of the target face and faces were paired by facial similarity. The position of the correct matched face (left or right on the screen) was counterbalanced across trials. Each face identity appeared twice as a target face and twice as a distractor face in both the dynamic and static conditions. The number of exposures to each facial identity was therefore counterbalanced across learning conditions.

Procedure

Participants were instructed to learn the face presented in the centre of the screen and, following a short interval (500ms), to indicate which of two face images matched the identity of the face they had learned. Responses were made by pressing either the "z" key to indicate the face on the left or "m" key for the face on the right on the keyboard. Participants were instructed to respond as quickly and as accurately as possible. The test images remained onscreen until a response was made. A response triggered the offset of the current trial and the onset of the next trial. Each trial was preceded by an inter-trial interval of 500ms in which a fixation cross was displayed.

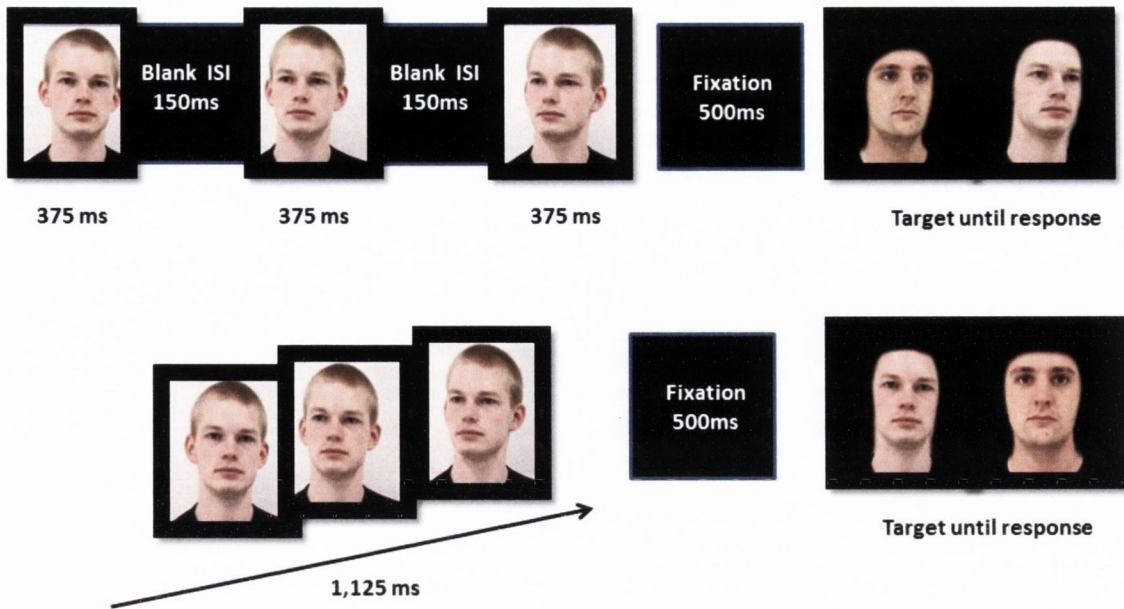


Figure 3.1. An example of a trial which illustrates the face images shown in the static (upper panel) and rigid motion (lower panel) training conditions used in Experiment 1A. The pair of images shown on the right depicts the test phase and includes an image of the target and the distractor face. The upper panel depicts an example of a novel test view of the target face, whereas the lower panel depicts a learned face image.

3.3.2. Results

Accuracy

Accuracy was calculated as the percentage of correctly matched faces in each condition for each participant. The mean percentage of correct responses across participants was subjected to a $2 \times 2 \times 2$ mixed design analysis of variance (ANOVA), with Age (younger or older) as a between-subject factor and both Target face (dynamic or static) and Test view (learned or novel) as within-subject factors. A significant main effect of Age was observed [$F(1,27)=20.12$, $p<0.001$, $\eta_p^2=0.43$], with lower face matching accuracy in the older ($M=85.9\%$) than younger ($M=97.0\%$) adult groups. A main effect of Test view was also observed [$F(1,27)=7.63$, $p=0.01$, $\eta_p^2=0.22$], with better accuracy for matching across learned than novel viewpoints of the face. There was a significant interaction between Test View and Age [$F(1,27)=7.63$, $p=0.01$, $\eta_p^2=0.22$]. A *post hoc* Tukey, Unequal N, HSD analysis was used to provide further details of this interaction. This analysis revealed no difference in younger adults' face matching performance between learned ($M=97.0\%$) and novel ($M=97.0\%$) views. However, accuracy rates in older adults were lower when matching novel ($M=83.0\%$) compared to learned ($M=88.6\%$) views of the face ($p=0.003$). There

was no effect of the Target face [$F(1,27) < 1$] and no two-way interactions between the Target face and either Age [$F(1,27) < 1$] or Test view [$F(1,27) < 1$]. However, a significant three-way interaction between Test view, Target face and Age was found [$F(1,27)=4.43$, $p=0.04$, $\eta_p^2=0.14$]. A *post hoc* Tukey, Unequal N, HSD analysis revealed better performance for matching learned than novel views in the older adult group, but only when the Target face was presented as a static sequence ($p=0.007$). In contrast, this viewpoint cost found when older adults matched static face views was not evident for the dynamic Target face condition. In this dynamic condition matching performance was the same across both novel and learned views in the older adults ($p=0.91$). No viewpoint cost was evident in younger adults' face matching performance across either target training conditions ($p=.99$). The mean accuracy for each condition is reported in Table 2.1.

Table 2.1. The mean percent accuracy performance (with standard deviations) in matching learned and novel views (Experiment 1A) and expressions (Experiment 2 and 3) of the learned target face, for each of the younger and older adult groups.

	Older Adults		Younger Adults	
	Static	Dynamic	Static	Dynamic
<i>Experiment 1A: Rigid Motion</i>				
Learned View	89.6 (10.5)	87.5 (10.0)	97.0 (6.2)	97.0 (3.2)
Novel View	81.4 (12.0)	85.0 (11.9)	98.3 (3.1)	95.7 (4.6)
<i>Experiment 2: Non-Rigid Motion</i>				
Learned Expression	87.5 (8.3)	84.3 (10.4)	96.9 (6.6)	95.0 (6.6)
Novel Expression	74.6 (9.7)	80.7 (10.2)	93.4 (6.0)	93.8 (5.0)
<i>Experiment 3: Non-Rigid Motion Inverted</i>				
Learned Expression	77.5 (12.4)	76.8 (11.5)	84.7 (8.7)	85.3 (7.9)
Novel Expression	65.0 (14.0)	66.4 (10.6)	77.2 (7.7)	76.6 (11.5)

Viewpoint Change Cost

It was noted that the variability in performance across individuals in the older adult group was large (i.e. 65-100% learned). With this in mind, and based on previous studies (e.g. Russell, Duchaine, & Nakayama, 2009), the *relative* difference in percent accuracy performance was calculated across learned and novel viewpoints, normalised to each individual's own 'baseline' accuracy performance for learned viewpoints $[(\text{learned}-\text{novel})]/\text{learned} \times 100$. This was calculated for each individual in the older age group across trials in both the static and dynamic target training conditions. This calculation more accurately reflects the *relative cost* in performance for matching novel compared to learned views of the face as it takes into account the cost in performance with changes in view as a consequence of their overall performance when no view change occurred. This score is referred to hereafter as the *viewpoint change cost*. For consistency, the same analysis was also applied to the data acquired from the younger adults. However, it is also important to note that accuracy performance was at ceiling for the younger adults across conditions therefore making these effects difficult to compare across age.

Based on the *viewpoint change cost* for each individual, a 2×2 mixed design analysis of variance (ANOVA) was conducted, with Age (younger or older adults) as the between-subject factor and Target face (dynamic or static) as the within-subject factors. A significant main effect of Age was observed [$F(1,27)=6.57$, $p=0.02$, $\eta_p^2=0.20$], with a larger cost in accuracy for matching faces across changes in viewpoint in the older (M cost=5.84 %) than younger (M cost=-0.2 %) adult group. There was no effect of Target face [$F(1,27) < 1$]. However, a significant interaction between Target face and Age [$F(1,27)=4.12$, $p=0.05$, $\eta_p^2=0.13$] was observed, and is shown in Figure 3.2. To provide further insight into the basis of this interaction, simple main effects of Age were analysed separately for the static and dynamic face Targets. These analyses revealed that compared to younger adults, there was a large viewpoint change cost in older adults when the Target face was static [$F(1,27)=8.76$, $p=0.006$, $\eta^2=0.25$]. However, this age-related cost with viewpoint changes was not evident when the Target face was dynamic, as there was no significant difference in face matching performance between younger and older adults [$F(1,27) < 1$].

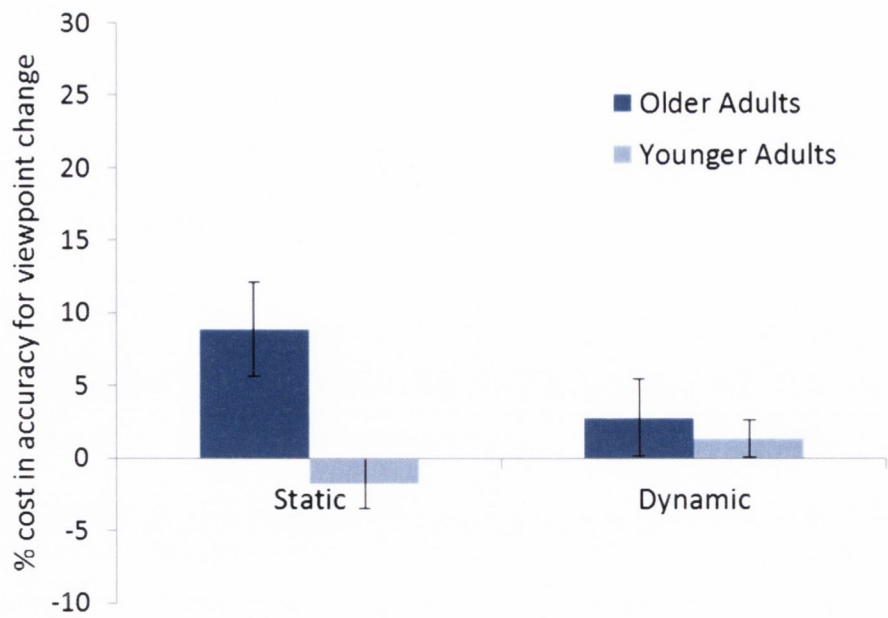


Figure 3.2. Plot showing the mean cost in accuracy for matching a novel relative to a learned viewpoint of the target face, for the older and younger adults tested in Experiment 1A. The results are shown for each of the static or dynamic (rigid motion) target face learning conditions. Error bars show ± 1 standard error of the mean (SEM).

Reaction Times

For the purpose of this analysis, RTs (in milliseconds) to correct matches only were included. Furthermore, RTs which were greater or less than 2.5 deviations from the mean RT for each condition, for each group, were excluded (approximately 8 % of the data were removed). A 2x2x2 mixed design analysis of variance (ANOVA) was then conducted on the time to match faces across both learned and novel viewpoints and across the Target face conditions in younger and older adults. A main effect of Age was observed [$F(1,27)=46.42, p<0.001, \eta_p^2=0.63$], with older adults taking more time to match faces ($M=1552\text{ms}$) than the younger ($M = 850\text{ms}$) adults. There was a main effect of Test view [$F(1,27)=4.60, p=0.04, \eta_p^2=0.15$], with slower reaction times to match facial identity across novel ($M=1204\text{ms}$) compared to learned ($M=1174\text{ms}$) viewpoints of the face. There was no main effect of Target face [$F(1,27)<1$] found. There was no evidence of an interaction between Test view and Age [$F(1,27)<1$], or between Target face and Age [$F(1,27) <1$] or Target face and Test view [$F(1,27) <1$]. The three way interaction between Test view, Target face and Age [$F(1,27)=1.00, p=0.33, \eta_p^2=0.03$] was also not significant. All mean reaction times across conditions are reported in Table 2.2.

Table 2.2. The mean reaction time taken in milliseconds (with standard deviations) to match learned and novel views (Experiment 1A) and expressions (Experiment 2 and 3) of the learned target face, for each of the younger and older adult groups.

	Older Adults		Younger Adults	
	Static	Dynamic	Static	Dynamic
<i>Experiment 1A: Rigid Motion</i>				
Learned View	1543 (405)	1516 (337)	835 (113)	848 (94)
Novel View	1566 (388)	1581 (442)	860 (120)	859 (90)
<i>Experiment 2: Non-Rigid Motion</i>				
Learned Expression	1700 (398)	1790 (499)	944 (129)	942 (199)
Novel Expression	2000 (453)	2018 (532)	1164 (242)	1149 (236)
<i>Experiment 3: Non-Rigid Motion Inverted</i>				
Learned Expression	1880 (536)	1888 (523)	1229 (260)	1373 (326)
Novel Expression	2362 (870)	2263 (640)	1629 (434)	1513 (344)

3.3.3. Discussion

Face matching performance was more accurate in younger than older adults, particularly when the test image was a novel viewpoint of the face. As such, the results from Experiment 1A confirm previously reported findings that face matching performance declines with age (e.g. Grady et al., 2000) and that this age-related cost is particularly apparent when matching facial identity across novel views of the face (Habak, Wilkinson, & Wilson, 2008). However, although older adults were overall worse at matching identity it was noted that the cost on identity matching across views, which has previously been reported to be exacerbated by the ageing process, was evident only when the target face was presented as a sequence of 3 static images containing rigid changes (i.e. head rotation). In contrast, when the face was presented dynamically (i.e. video) during learning there was no evidence for an age-related deficit in face matching accuracy across changes in viewpoint. Older adults were, however, slower in matching facial identity than younger adults. However, reaction times increased for both age groups when matching the face across changes in viewpoint.

Importantly, a dynamic advantage was not observed in the younger adult group, as face matching accuracy was similar across both the static and dynamic conditions. This suggests that the additional view information in the multiple static sequences was sufficient for younger adults to perform the task. In contrast to previous studies (e.g. (Pilz et al., 2006) the current study failed

to find a benefit on reaction time performance for matching moving relative to statically learned faces, in the younger (or older) adult group. This may relate to the nature of the task, as test images remained onscreen until a response was made. Other studies which have observed faster matching times for faces learned in motion have typically limited the exposure time of the test image (Pilz et al., 2006; Thornton & Kourtzi, 2002). However, it is also possible that the similarity between younger adults' reaction time performance for static and dynamic conditions may relate to the preserved spatial alignment of the static images. This was examined in more detail in Experiment 1B.

3.4. EXPERIMENT 1B

Experiment 1B was designed to investigate whether the preserved spatial alignment of the static images presented during target learning in Experiment 1A may have effected face matching performance, in the younger adult group. Specifically, using the same matching paradigm reported in Experiment 1A, this study examined whether learning a face through a sequence of static images in which the spatial alignment of images was preserved, relative to when it was disrupted, would affect the representation of the face in memory. The static-learning condition only was presented here. Prior findings from younger adult studies have demonstrated that spatiotemporal alignment is an important factor in associating multiple views of the same face in memory (Wallis & Bühlhoff, 2001). It was predicted that if spatiotemporal alignment assisted in integrating all images contained within the static sequence, then matching performance should be better when the images are viewed in ordinal sequence (e.g. 45° left view, 22.5° left view and full-face), relative to when the order of the sequence is disturbed (e.g. 45° left view, full face and 22.5° left view).

3.4.1. Method

Participants

Eighteen younger adults, recruited from the student population at Trinity College Dublin (7 female, 17 right handed, mean age= 24.4 years, SD= 4.3 years) participated in this experiment and were given course credits in return for their time. None of the participants had participated in Experiment 1A and all reported normal or corrected to normal vision.

Stimuli and Apparatus

The static stimuli were as described in Experiment 1A, i.e. they were extracted from the first frame, a mid-sequence frame (on average frame 15) and the last frame of each of the videos sequences of 'turn to right' and 'turn from left'. However, in the current Experiment the sequence order of the static stimuli was manipulated. In half of the trials the static learning stimuli were presented in the ordinal sequence described in Experiment 1A i.e. the first frame, a mid-sequence frame and the last frame of the video clip. This is referred to as a 'static fluid' sequence. However, for the other half of the trials the ordinal sequence of the static learning stimuli was disrupted as follows: 1) a mid-sequence frame, the first frame and the last frame; 2) or the last frame, the first frame and a mid-sequence frame; 3) or the first frame, the last frame and a mid-sequence frame was presented. This is referred to as a 'static jumbled' sequence. The 3 types of jumbled sequences were presented randomly across trials. Note that the following order of last frame, a mid-sequence frame and the first frame was never presented as a learning sequence in this condition as this is still considered a fluid sequence, albeit in reverse. A static fluid and static jumbled sequence was created for each of the 20 facial identities, giving rise to a total of 40 learning stimuli. As in Experiment 1A, the display duration of each static image presented during the learning stage was 375ms. Each static sequence (fluid or jumbled) contained an inter stimulus interval (ISI) of 150ms between each of the 3 images and as such each static fluid and static jumbled sequence was presented for an equal duration (1,125ms) during the learning exposure. As described in Experiment 1A, the static test images were extracted from the first and final frame of either the *turn to right* or *turn from left* video and as such were learned or novel viewpoints of the face.

Design

The learning phases were as described in Experiment 1A, with the exception that the dynamic trials were not included (40 trials) but were replaced with static jumbled sequences. The remaining 40 trials were presented in the same static sequence as described in Experiment 1A. As in Experiment 1A, half of the participants were shown the *turn to right* sequence in the learning phase (static fluid or jumbled) and then presented with the 45° left view of the actor's face as the novel viewpoint in the test, while the other half were shown the reverse i.e. the *turn from left* sequence (static fluid or jumbled) as the target face and then the 45° right view of the actor's face as the novel viewpoint. All other details remain as described in Experiment 1A.

Procedure

The procedure is as described in Experiment 1A.

3.4.2. Results

Accuracy

The mean percentage of correct responses across participants was subjected to a 2×2 repeated measures analysis of variance (ANOVA). Both Target face (static fluid or static jumbled) and Test view (learned or novel) were repeated factors. Analysis revealed no significant main effect of Target face [$F(1,17)=2.69$, $p=0.12$] or Test view [$F(1,17)=1.75$, $p=0.20$]. There was also no evidence of a significant interaction between Target face and Test view [$F(1,17)=1.99$, $p=0.18$].

There was no viewpoint change cost calculated for percent correct performance on this task, since a comparison of performance across younger and older adults was not required. Furthermore, there was no evidence for an effect of learning condition or test view on accuracy scores in this younger adult group, as performance ranged from 95 to 97 percent correct across all conditions.

Reaction Times

As in Experiment 1A, RTs (in milliseconds) to correct matches only were included in the analysis. In addition, RTs which were greater or less than 2.5 deviations from the mean RT for each condition, were excluded from the analysis (approximately 7 % of the data were removed). The reaction time data were analysed using a 2×2 repeated measures analysis of variance (ANOVA), with Target face (static fluid or static jumbled) and Test view (learned or novel) as repeated factors. No main effect of Target Face was found [$F(1,17) < 1$]. However, a main effect of Test view was observed [$F(1,17)=15.57$, $p<0.001$, $\eta_p^2=0.48$], with slower reaction times to match novel ($M=823\text{ms}$) relative to learned ($M=771\text{ms}$) views of the face. Furthermore, there was a significant interaction between Test view and Target face [$F(1,17)=9.53$, $p=0.007$, $\eta_p^2=0.36$]. A post hoc Tukey analysis revealed that the time taken to match the face across novel, relative to learned, views of the face was significantly increased when the face had been learned through a jumbled rather than a fluid static sequence ($M= 838$ ms; 754 ms respectively). See Figure 3.3 for an illustration of this interaction.

3.4.3. Discussion

The results from Experiment 1B revealed that the order in which a sequence of static views of a face are learned can affect the time to subsequently recognise that face. In particular, reaction time performance was slower to match facial identity across changes in visual appearance when the face had been learned through a series of static images in which the sequence order of the images was disrupted, relative to when it was preserved. Thus, in line with previous research (e.g. Wallis & Bülthoff, 2001), this result suggests that spatiotemporal alignment is an important factor for integrating all views of a face into a robust single representation of that facial identity. In turn, this result suggests that the *preserved* sequential order of face images in Experiment 1A may have reduced the possibility for any dynamic advantage on performance in the younger adult group, as all images may have been readily integrated into a single representation by this particular age group.

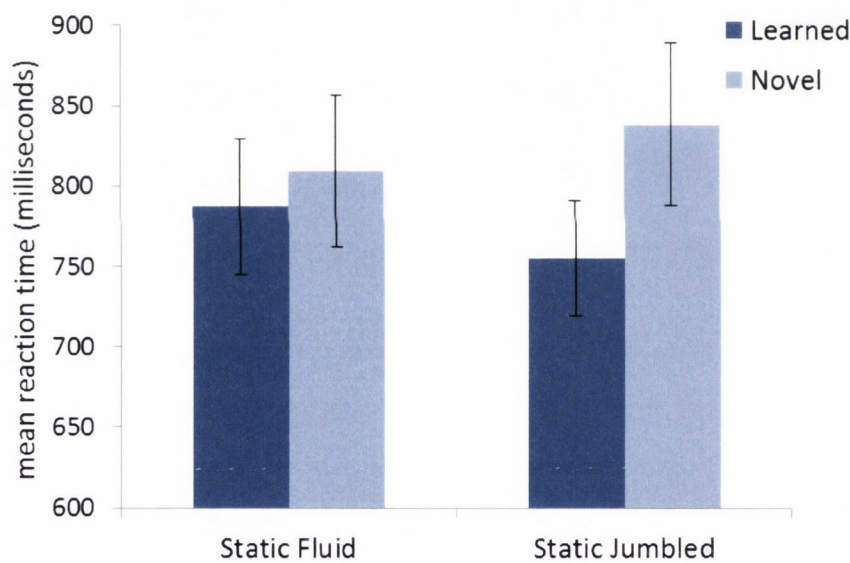


Figure 3.3. Plot showing the mean reaction times for younger adults to match a learned, relative to novel, view of the face across static fluid and static jumbled learning conditions. Error bars show ± 1 SEM.

3.5. EXPERIMENT 2

The following experiment investigated whether the dynamic enhancement observed in Experiment 1A for the older adult group would extend to dynamic faces when the motion was non-rigid (i.e. internal deformations of the face). Specifically, the current study assessed whether viewing a face with expressive movements during learning would increase accuracy in matching the identity of the face across novel expressions, relative to the presentation of static images of a face during learning. To that end, the same matching paradigm as described in Experiment 1A was

employed, but here the task involved matching faces across changes in expression. Again, performance on this task was compared between older and younger adults. As Experiment 1A, it was predicted that if dynamic information assists in forming a representation of the face which is more robust to incidental changes in image properties, then it should be possible to observe improved face matching performance in older adults when face targets are presented as dynamic rather than static sequences, particularly when matching to novel (unlearned) expressions of the face. However, as found in Experiment 1A, it is possible that the younger adult group may perform similarly across both static and dynamic learning conditions due to the diversity of information available in the multiple images of each face presented across both the dynamic and static conditions.

3.5.1. Method

Participants

Sixteen naive undergraduate students from Trinity College Dublin (14 female, 15 right handed, mean age= 18.9 years, SD= 1.5 years) participated in this experiment for course credits. All younger adults reported normal or corrected to normal vision. Eighteen naive, community dwelling, older adults also participated in this experiment, however data from four older adults were excluded from analyses as it was subsequently found that these individuals had mild cataracts at the time of testing. Data from the other 14 older adult participants were included in the analysis (11 female, 13 right handed, mean age= 72.9 years; SD= 6.4 years). All older adults scored in the normal range on the MoCA test (i.e. ≥ 24 of 30) suggesting there was no evidence of cognitive impairment in these individuals. All older participants had normal visual acuity (mean binocular LogMar score= 0.0, SD= 0.1 and LogMar score= 0.1, SD= 0.1 at 4m and 40 cm respectively) and contrast sensitivity (as assessed using the Pelli-Robson Contrast Sensitivity Test; mean score= 2.0, SD= 0.1) for their age. All participants wore appropriate corrective lenses, if needed, at the time of testing. None of the participants in this experiment had taken part in Experiment 1A, 1B.

Stimuli and Apparatus

The stimulus set is as described in Experiment 1A with the exception that only frontal views (i.e. 0°) of the actors' faces were used and the facial movement involved a change from a neutral to a happy or angry facial expression. The static stimuli consisted of the first frame, a mid-sequence frame (usually frame 15) and the final frame of the video. All video clips were 1,125ms

in duration. The display duration of each static image was 375ms, with an inter stimulus interval (ISI) of 150ms included between image presentation. The video and the static target images were therefore presented for an equal amount of time (1,125ms). The static images used in the test condition were always extracted from the final frame of the videos, which was the apex of the expression. As in Experiment 1A, the test images were edited in Adobe Photoshop® to include a black mask which occluded non-facial features, i.e. hairstyle. The apparatus used for testing is as described in Experiment 1A. The experiment was conducted on-site, in a testing laboratory.

Design

The experiment was based on a mixed, 3-way design with Age (younger or older) as the between-subjects factor, and both Target face (dynamic or static) and Test expression (learned or novel) as within-subject factors. All other details of the design are as reported in Experiment 1A, with the exception that target and test images varied by expression and not viewpoint. In half of the trials, one of the test images was of the apex of the expression shown during the target face presentation. In the other half of the trials one of the test images was of the target face presented from a novel expression which was not viewed during the target face presentation. In all trials the expression of the distractor face matched that of the target test face. During the learning of the target face, half of the participants were shown the 'happy' expression as the target face and were then presented with an angry expression as the novel expression during the test phase. The other half of the participants were shown the 'angry' expression as the target face and the happy expression as the novel expression (see Figure 3.4 for an illustration of a trial).

Performance was measured in response times and accuracy rates. However, as in Experiment 1A, the relative difference in percent correct for matching faces across learned and novel expressions, normalised to each individual's own 'baseline' accuracy performance for learned expressions of the face was also calculated i.e. $[(\text{learned} - \text{novel})] / \text{learned} * 100$. As in Experiment 1A, this *expression change cost* was calculated separately for both static and dynamic trials.

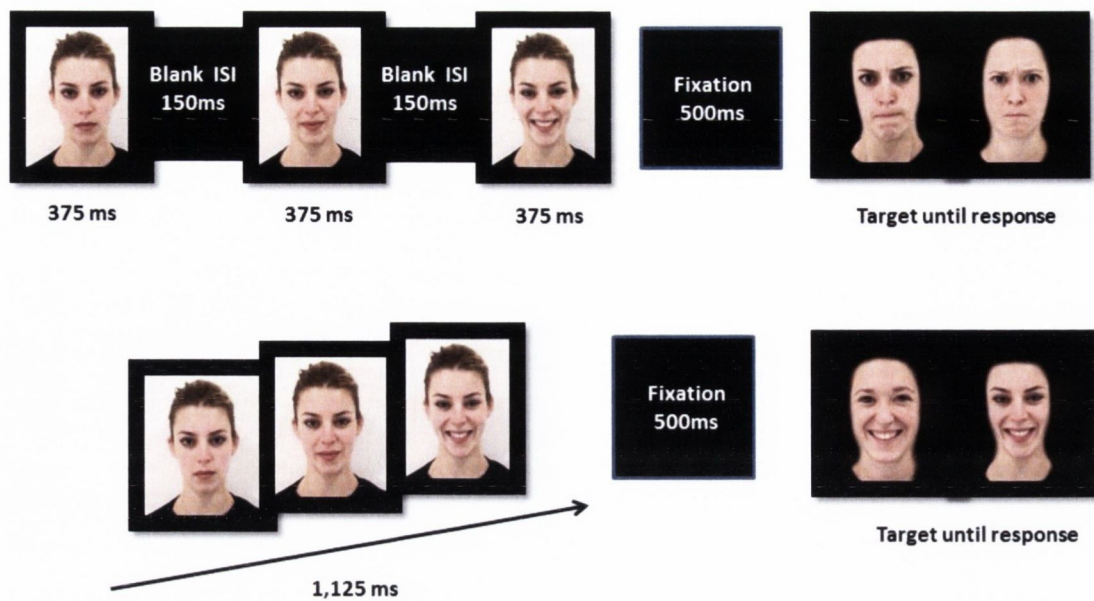


Figure 3.4. An example of a trial which illustrates the face images shown in the static (upper panel) and non-rigid motion (lower panel) training conditions used in Experiment 2. The pair of images shown on the right depicts the test phase and includes an image of the target and the distractor face. The upper panel depicts an example of a novel test expression of the target face, whereas the lower panel depicts a learned face image.

Procedure

The procedure was as described in Experiment 1A.

3.5.2. Results

Accuracy

All accuracy means are reported in Table 1. The mean accuracy rates across participants was subjected to a 2×2×2 mixed design ANOVA with Age (younger or older) as the between subjects factor and Target face (dynamic or static) and Test expression (learned or novel) as repeated measures. A main effect of Age was observed [$F(1,28)=35.34$, $p<0.001$, $\eta_p^2=0.56$], with lower accuracy in face matching performance in the older ($M=81.8\%$) compared to the younger ($M=94.8\%$) adult group. A main effect of Test expression was also observed [$F(1,28)=22.49$, $p<0.001$, $\eta_p^2=0.45$], with lower accuracy for matching the face across novel ($M=86.2\%$) compared to learned ($M=91.3\%$) expressions. There was no effect of Target face [$F(1,28) < 1$]. A significant interaction between Test expression and Age was found [$F(1,28)=6.95$, $p=0.01$, $\eta_p^2=0.20$]. A *post hoc* Tukey, unequal N, HSD analysis revealed no difference in young adults matching performance across learned and novel expressions of the face but performance in older adults was significantly

worse when matching novel compared to learned expressions of the face ($p < 0.001$). There was no interaction found between Target face and Age [$F(1,28) < 1$]. There was, however, a significant interaction between Test expression and Target face [$F(1,28) = 8.90$, $p = 0.005$, $\eta_p^2 = 0.24$]. A *post hoc* Tukey, unequal N, HSD revealed that accuracy for matching learned expressions of the face was similar across both static and dynamic training conditions. When static target faces were learned, however, matching novel expressions was less accurate than matching learned expressions ($p < 0.001$). In contrast, when dynamic faces were learned there was no difference in matching performance across learned and novel expressions of the face ($p = 0.33$). The three way interaction between Test Expression, Target face and Age [$F(1,28) = 3.40$, $p = 0.07$, $\eta_p^2 = 0.11$] approached significance. As this three way interaction approached significance and because there was a large variation in face matching performance across the older adults, these age-related effects were addressed through analysing the associated cost in performance with a change in expression.

Expression Change Cost

Performance calculated as the *expression change cost* (described in the Design section) was subjected to a 2×2 mixed design ANOVA with Age (younger or older) as the between-subject factor and Target face (dynamic or static) as the within-subject factor. A significant main effect of Age was observed [$F(1,28) = 8.25$, $p = 0.007$, $\eta_p^2 = 0.23$] with worse performance in the older adult group to face matching across changes in expression ($M \text{ cost} = 9.1\%$) than younger adults ($M \text{ cost} = 2.3\%$). There was a significant effect of Target face [$F(1,28) = 9.39$, $p = 0.004$, $\eta_p^2 = 0.25$], with better matching performance across changes in expression when the target faces were dynamic ($M \text{ cost} = 2.5\%$) than static ($M \text{ cost} = 8.5\%$). A significant interaction between Target face and Age group [$F(1,28) = 4.02$, $p = 0.05$, $\eta_p^2 = 0.13$] was observed, as shown in Figure 3. To analyse this interaction, tests of simple main effects comparing the Age groups were performed separately for the Static and Dynamic target face conditions. These analyses revealed a significant effect of Age when the target face was learned as a series of static images [$F(1,28) = 11.95$, $p < 0.001$, $\eta_p^2 = 0.30$], with a larger 'expression change cost' for the older compared to the younger adult group. However, there was no difference in the expression change cost for younger and older adults face matching performance when the target face was learned as a dynamic face [$F(1,28) < 1$]. In addition, a simple main effects analysis revealed that an effect of the learned target face was evident in the older adult group [$F(1,28) = 10.12$, $p = 0.007$, $\eta_p^2 = 0.44$], with poorer performance when the face was learned as a sequence of static images ($M \text{ cost} = 14.3\%$) compared to dynamic presentation ($M \text{ cost} = 3.8\%$). This difference was not evident in the younger adult group [$F(1,28) < 1$] as shown in Figure 3.5.

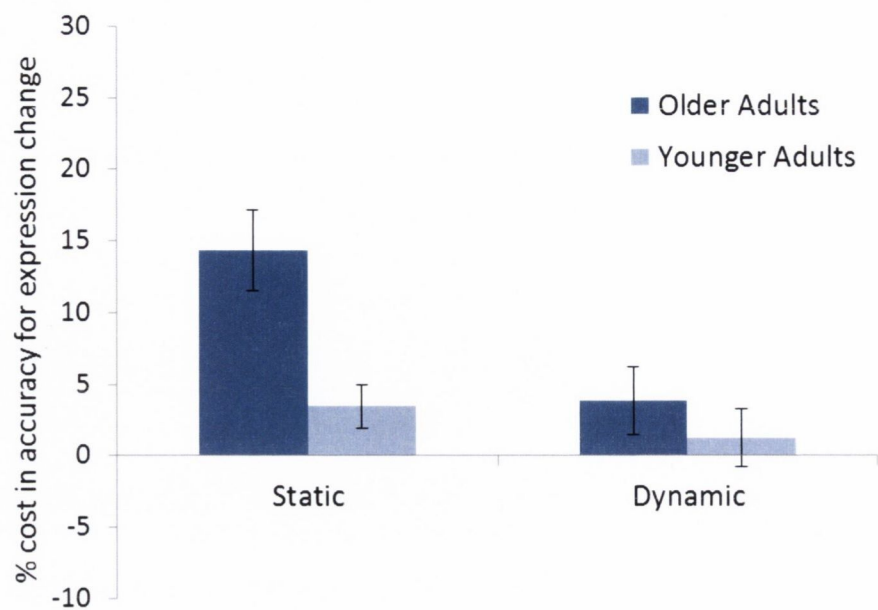


Figure 3.5. Plot showing the mean cost in accuracy for matching a novel relative to a learned expression of the target face, for the older and younger adults tested in Experiment 2. The data are presented for each of the static or dynamic (non-rigid motion) target face learning conditions. Error bars show ± 1 SEM.

Reaction Times

RTs to correct matches only were included. Furthermore, RTs which were 2.5 deviations away from the mean RT for each condition, for each group were excluded (approximately 7 % of the data were removed). The mean reaction time performance to correctly match faces were then subjected to a 2x2x2 mixed ANOVA with group Age (younger or older) as the between subjects factor and both Target face (dynamic or static) and Test expression (learned or novel) as repeated measures. A main effect of Age was observed $F(1,28)=44.30, p<0.001, \eta_p^2=0.61$], with older adults taking longer to match faces ($M=1877\text{ms}$) than younger adults ($M=1050\text{ms}$). A main effect of Test expression was also evident [$F(1,28)=80.50, p<0.001, \eta_p^2=0.74$], with faster reaction times when matching learned ($M=1317\text{ms}$) than novel ($M=1554\text{ms}$) expressions of the face. No main effect of Target face was found [$F(1,28)<1$]. There were no significant interactions between Test expression and Age [$F(1,28)<1$], between Target face and Age [$F(1,28)=1.82, p=0.19, \eta_p^2=0.06$] or between Test expression and Target face [$F(1,28)=1.31, p=0.26, \eta_p^2=0.04$]. The three way interaction between Test expression, Target face and Age [$F(1,28)<1$] was also not significant. All mean reaction times across conditions are reported in Table 2.2.

3.5.3. Discussion

In terms of accuracy, older adults were poorer at matching facial identity than younger adults, particularly when the image was of a novel expression of the face. However, it was noted that the *expression change cost* (i.e. the cost associated with matching a novel relative to learned view of the face) was evident in the static condition only, when the target face was presented as a series of 3 images extracted from the video sequence containing non-rigid changes (i.e. movement associated with expression). In contrast, when the face was learned in motion there was no evidence for an age-related cost in face matching performance across changes in expression. Moreover, for the older adults, face matching across novel expressions was better when the target face was presented moving rather than through a sequence of static images during learning. Indeed, the expression change cost dropped from approximately 14% when the target face was static, to approximately 4% when the target face was dynamic. Thus the results from Experiment 2 both replicate and extend those reported in Experiment 1A, and suggest that a dynamic enhancement in face matching performance in older adults is evident for both rigid and non-rigid facial changes. Furthermore, consistent with the results of Experiment 1A there was no observed benefit of motion, over static presentations of faces, on either accuracy or response time performance in the younger adult group. This may be because the multiple view information in both the static and dynamic sequences may have been sufficient to create a robust, less image-dependent, representation of the face in younger adults. Furthermore, the preserved spatial alignment of the static images may have benefitted this performance (see Discussion for Experiment 1A and 1B). Had the task been more difficult, for example comparing dynamic performance to a single static image, it is possible that the effect of motion on face matching may have emerged. However the reaction time results are consistent with those found in Experiment 1A in that older adults took longer to match faces for identity than younger adults. In addition, the reaction time cost associated with matching the face across changes in expression was similar for younger and older adults.

The results from Experiments 1A and 2 suggest that the availability of motion during face learning facilitates perception in older adults. The results may also suggest that dynamic information facilitates perception in general in older adults, although this would be counter to previous studies which reported relatively poorer motion perception in older compared to younger adults (e.g. Snowden & Kavanagh, 2006). To reconcile these findings, it may be the case that the benefit of motion may be specific to face processing. To that end, the purpose of Experiment 3 was to investigate whether the dynamic enhancement observed in the older adult

group was specific to face perception by testing whether the benefit for learning dynamic faces generalised to inverted faces, or was specific to faces viewed in an upright orientation.

3.6. EXPERIMENT 3

Previous studies have reported that face recognition is disrupted when the face is viewed in an inverted, relative to upright, orientation. Face inversion is thought to interfere with the holistic or configural processing of the face, which is argued to be specific to face perception (Yin, 1969; Young, Hellawell, & Hay, 1987; Maurer, Le Grand, & Mondloch, 2002). Interestingly, Thornton, Mullins, & Banahan (2011) noted that unlike other studies which have demonstrated the facilitative effect of motion for face recognition under canonical viewing conditions (e.g. Pilz et al., 2006; Thornton & Kourtzi, 2002), facial motion interfered with face processing when the image was inverted. In the Thornton et al. (2011) study, compared to static presentation of faces, reaction times taken to identify the sex of an inverted moving face typically increased by an average of 100 milliseconds compared to performance to the upright faces. The authors argued that such interference effects reflect how the encoding of facial motion cues is likely to be mandatory, as they cannot be suppressed in favour of the available form cues. Under inverted conditions motion may interfere with the observer's ability to track specific features of the face for identification purposes, thus affecting recognition performance. Similarly, facial motion appears to facilitate the recognition of familiar faces under degraded visual conditions, and this facilitation seems to be strongest when viewing upright faces (Knight & Johnston, 1997; but see Lander et al., 1999). Such results suggest that the ability to use motion cues for the purpose of recognition likely engages a motion encoding system which is tuned to upright, but possibly not, inverted faces (Hill & Johnston, 2001; see also Steede, Tree, & Hole, 2007a, 2007b).

As dynamic facial features may be difficult to encode under inverted viewing conditions, it was predicted that compared to a static presentation, motion may interfere with, or at least not be of benefit to, face perception. This may be particularly evident when matching identity across images of the face which differ in visual appearance. Moreover, the effect of motion on matching inverted faces may be more pronounced in the older adult group as upright motion cues appear important in forming a more robust, less image-dependent, face representation with older age (see results from Experiment 2). To address these questions, the following experiment replicated Experiment 2 with the exception that all target and test images and videos were rotated 180° degrees in the picture plane.

3.6.1. Method

Participants

Sixteen naive younger adults (11 female, 14 right handed, mean age= 20.4 years, SD= 3.6 years), recruited from the undergraduate population at Trinity College Dublin, participated in this experiment for research credits. All participants had self-reported normal or corrected to normal vision. Fourteen naive community-dwelling older adults also participated in this experiment (11 female, all right handed, mean age= 69.9 years; SD= 5.5 years). All older adults scored in the normal range on the MoCA (i.e. ≥ 24 of 30) and all had normal visual acuity for their age, as measured at 4 meters (mean binocular LogMar score= 0.04, SD= 0.1) and 40cm (mean binocular LogMar score= 0.09, SD= 0.1). Contrast sensitivity was also normal for age and was assessed using the Pelli-Robson Contrast Sensitivity Test (mean score= 1.97, SD= 0.08). All participants wore appropriate corrective lenses, if needed, at the time of testing. With the exception of 4 older adults who had also participated in Experiment 2 at least 2 months prior to this experiment, no participant had taken part in either of the preceding experiments.

Stimuli and Apparatus

The stimuli used were as described in Experiment 2, with the exception that all videos and images were inverted. The inverted stimuli were created by rotating the video or static image 180° in the picture plane. VirtualDub open source video software (www.virtualdub.org) was used to edit the video clips and Adobe Photoshop® software was used to manipulate the static images.

Design

The design of this experiment, and the approach to the analyses of the data, was as described in Experiment 2.

Procedure

The procedure was as described in Experiment 2.

3.6.2. Results

Accuracy

A 2×2×2 mixed ANOVA was conducted on the participants' mean accuracy results with group Age (younger or older) as the between subjects factor and Target face (dynamic or static) and Test expression (learned or novel) as repeated measures. A main effect of Age was observed [$F(1,28)=13.29$, $p<0.001$, $\eta_p^2=0.32$], with lower face matching accuracy in the older ($M=71.4\%$) than younger ($M=80.9\%$) adult group. A main effect of Test expression was also observed [$F(1,28)=25.97$, $p<0.001$, $\eta_p^2=0.48$], with lower accuracy for matching the face across novel ($M=71.7\%$) compared to learned ($M=81.3\%$) expressions. No interaction between Test Expression and Age was found [$F(1,28)<1$]. There was no effect of Target face [$F(1,28)<1$]. There were no significant two-way interactions between Target face and group Age [$F(1,28)<1$] or Test expression and Target face [$F(1,28)<1$]. There was no evidence for a three way interaction between Test expression, Target face and Age [$F(1,28)<1$]. The mean accuracy rates across age groups and conditions are reported in Table 2.1.

Expression Change Cost

The *expression change cost* was calculated i.e. the cost in percent correct for matching faces across novel compared to learned expressions for each participant and subjected these data to a 2×2 mixed ANOVA with the Age of the group (younger or older) as the between-subject factor and Target face (dynamic or static) as the within-subject factors. There was no effect of Age observed [$F(1,28)<1$] nor of Target face [$F(1,28)<1$]. There was no interaction between Target face and Age [$F(1,28)<1$].

Reaction Times

Reaction times which were less than or greater than 2.5 standard deviations from the mean RT, to the correct trials for each condition, for each group were removed. This amounted to a removal of approximately 6 % of the data. The participants' mean reaction time performance to the correct trials only was analysed using a 2×2×2 mixed design ANOVA with Age of group (younger or older) as the between-subject factor and Target face (dynamic or static) and Test expression (learned or novel) as repeated measures. A main effect of Age was observed [$F(1,28)=14.14$, $p<0.001$, $\eta_p^2=0.34$], with older adults taking longer to match faces ($M=2098\text{ms}$) than younger adults ($M=1436\text{ms}$). A main effect of Test expression was also evident [$F(1,28)=59.40$, $p<0.001$, $\eta_p^2=0.68$], with faster reaction times when matching learned

($M=1573\text{ms}$) than novel ($M=1917\text{ms}$) expressions of the face. No effect of Target face [$F(1,28)<1$] was evident. The two-way interactions between Test expression and Age [$F(1,28)=3.07$, $p=0.09$, $\eta_p^2=0.09$] or between Target face and Age [$F(1,28)<1$] were not significant. However, a significant two-way interaction between Test expression and Target face [$F(1,28)=6.72$, $p=0.01$, $\eta_p^2=0.19$] was observed. A *post hoc* Tukey, Unequal N, HSD analysis revealed results which reiterated the main effects of Test expression, with slower RTs for matching novel compared to learned expressions of the face, both across and within the static and dynamic conditions ($p<0.001$). Specifically, RTs were slower when matching novel expressions, when the face was learned as a series of static images ($M=1971\text{ms}$), compared to matching learned expressions when the target face was presented in motion ($M=1613\text{ms}$). In turn, RTs were also slower for matching novel expressions when the face was learned in motion ($M=1863\text{ms}$), compared to matching learned expressions in the static learning condition ($M=1533\text{ms}$) (see Figure 3.6). The three way interaction between Test Expression, Target Training Condition and Age [$F(1,28)=1.17$, $p=0.29$, $\eta_p^2=0.04$] was not significant. All mean reaction times across conditions are reported in Table 2.2.

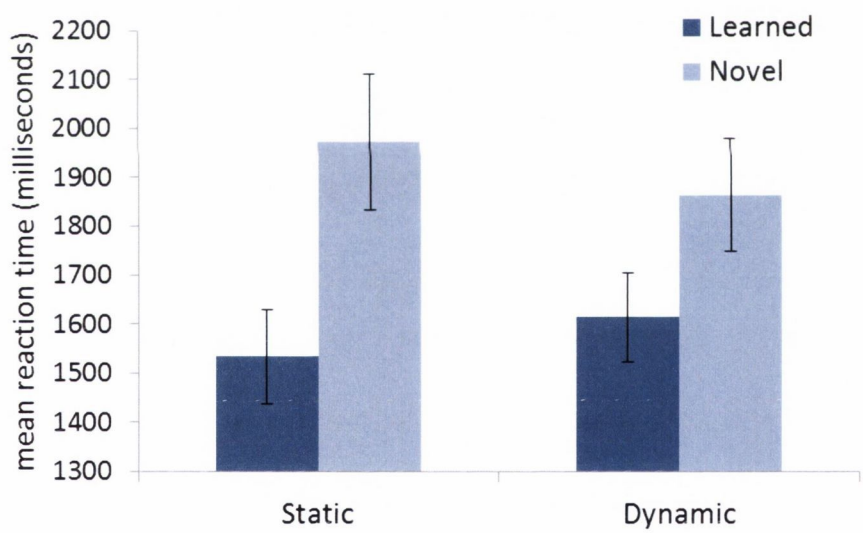


Figure 3.6. Plot showing the interaction between the test expression and learned target face on the mean response times in Experiment 3, in which all face images were inverted. Error bars show ± 1 SEM.

3.6.3. Discussion

In terms of overall accuracy, younger adults were better than older adults at matching facial identity in inverted images. However, unlike the results from Experiment 2 the cost in accuracy associated with matching inverted faces across changes in expression was similar across age groups and was found in both the accuracy and associated *expression change cost*

performance measures. This similar cost implies that faces are processed in a holistic or configural manner (for a review see Maurer et al., 2002) irrespective of age (see also Boutet & Faubert, 2006). Moreover, there was no observed effect of the motion of the target face on the expression change cost i.e. learning inverted dynamic faces was not associated with a reduction in the error rate in matching identity across facial expressions. However, although no beneficial role for motion on matching performance was found, unlike previous reports by Thornton et al. (2011) it was not observed that motion interfered with task performance in either the younger or older adult groups relative to static presentation, e.g. by increasing reaction times. This maybe because both static and dynamic learning conditions contained multiple image frames. Therefore, in both conditions there was a change in facial form information over time and strategies for identity matching under the inverted condition e.g. feature matching, may have been equally difficult to implement in both learning conditions. Consistent with the results of the previous experiments older adults were slower at matching faces for identity than younger adults. However, both younger and older adults showed a similar increase in the time taken to match the inverted face across changes in expression. Taken together, the results from Experiment 3 suggest that the ability to use dynamic information for the purpose of face perception likely relates to motion encoding which is tuned to upright faces only. Furthermore, any effects due to a general arousal state induced by motion cannot simply explain the results observed from Experiment 1A and Experiment 2. Rather motion confers an advantage which is related to the representation of facial identity.

3.7. General Discussion

The aim of the present study was to investigate the role of facial motion cues on the perception of unfamiliar faces. Previous studies showed inconsistent evidence for a benefit of facial motion on perception in younger adults. Here, the role of facial motion in face matching across changes in viewpoint and expression was examined, in younger and older adults. During learning, participants were presented with either a short video or 3 static face images of a target face. When presented as a video this target face was learned in either a rigid (Experiment 1A) or a non-rigid (Experiment 2) manner. Immediately following presentation, participants were asked to match the identity of the face image from two available images, one of which was a distractor face. All test images were static and contained either a learned or novel view (Experiment 1A) or expression (Expression 2) of the target face. The results from Experiments 1A and 2 were consistent in that face matching performance was better in younger than older adults, particularly when face images contained a change in visual appearance, such as a novel expression or view.

However, further analysis revealed that matching novel face images improved when the target face was presented as a moving compared to a static sequence of images, particularly for older adults. Furthermore, matching performance was similar for static and dynamic learning conditions, when the task was to match a learned image of the face. Importantly, the dynamic enhancement for matching novel images of faces was evident when faces were learned moving in either a rigid (i.e. viewpoint) or a non-rigid (i.e. expression) manner. Conversely, when faces were inverted (Experiment 3) no such benefit was observed for learning moving relative to static faces on matching inverted faces for either the younger or older adults. Moreover, a similar cost in matching facial identity across changes in expressions was found for both age groups, when faces were inverted.

The finding that older adults are relatively worse at matching unfamiliar faces than younger adults, particularly across changes in visual appearance, is consistent with previous research. For example, Habak et al. (2008) noted that compared to younger adults, performance in older adults was poorer when matching identities across changes in view. Moreover, they found that this cost in performance could not be reliably reduced in the older adult group by increasing the exposure duration of the target image, unlike in the younger adult group. For the younger adults, performance continued to improve with increasing exposure duration (i.e. there was a consistent improvement in face discrimination performance across changes in viewpoint with increasing exposure of 200ms, 500ms or 1000ms respectively). Habak et al. (2008) concluded that there may be a limit to the amount of form information which older adults can encode for matching facial identity across views. Furthermore, they noted that the viewpoint cost is unlikely to be due to a general slowing in processing speed with age, as younger adults could do the task at even shorter exposure durations, but is more likely to reflect changes in the ability to associate different facial views with the same identity. However, Habak et al. (2008) increased the exposure duration of a target image from one viewpoint only, therefore such exposure is likely to elicit a pictorial or image-based encoding of that face and consequently render face matching more image-dependent in the older adult group. In contrast, the current findings suggest that an age-related cost in matching facial identity across views can be reduced when a face was learned moving in a rigid manner, relative to a series of static face images, suggesting that motion can facilitate the association of facial views in memory. Indeed, matching performance to novel views of the face was equivalent to the best performance in this group (i.e. matching learned views in either static or dynamic target learning conditions, see Table 2.1) when the face was learned moving rigidly. The results from studies involving younger adults (e.g. Pilz et al., 2006; Thornton & Kourtzi, 2002) suggest that viewing a face in motion is likely to elicit a face-based, rather than

image-based encoding of identity. As such, these findings suggest that dynamic information can assist in linking facial identities across views (i.e. allowing novel viewpoints and expressions of the face to be more readily linked to the facial representation) and that the ageing brain can exploit these cues to form a more robust structural representation of the face. However, it cannot be completely ruled out that the additional static information in the dynamic sequence may have contributed to these effects (videos, by default, contained more static frames per second than the multiple static control condition). Nonetheless, the viewpoint dependency displayed in older adults' face matching performance, even when multiple static images of the face were available during learning, would suggest that it is the combination of form and motion information which may maintain efficient face perception with older age. This combination of cues may become more important as the ability to efficiently encode form information declines.

Interestingly, this dynamic advantage was also observed in older adults when the face was learned moving in a non-rigid manner. Previous studies have suggested that non-rigid motion may support recognition through the extraction of a dynamic facial signature (e.g. the way someone may move their forehead in thought), which is processed independently of form information (e.g. O'Toole et al. 2002). The results from Longmore & Tree (2013) support this idea. They observed that both developmental prosopagnosics and age-matched controls (mean age 50 years) were better at matching the identity of a face across changes in facial viewpoint when the face displayed the same non-rigid motion during the learning and test phase, compared to when both the learned and test images were static in nature. This suggests that the ability to extract characteristic motion cues to support recognition may remain relatively well preserved with increasing age and lends favour to the supplemental information hypothesis proposed by O'Toole and colleagues. However, the current findings suggest that non-rigid motion may also support face processing in a manner which is consistent with the representation enhancement hypothesis. It was found that when the target face was learned moving in a non-rigid manner (i.e. an expression change), that this was sufficient to improve matching performance for novel expressions of the face in the older adult group (by about 10%). Here, novel image matching was superior in the dynamic relative to static condition, an effect which was apparent when face matching performance was normalised to each individual's best performance (i.e. learned expressions). Thus the current results extend Longmore & Tree (2013) findings and suggest that viewing a face in non-rigid motion also enhances subsequent facial identity matching, even when the image is static in nature.

Schultz & Pilz (2009) demonstrated that non-rigid motion increased activation in both form and motion areas of the face processing network. Such increased activation may be

mediated by perceptual experience, as we more often view faces in motion and are therefore likely to have a larger neural network responsive to moving than to static faces images (also see Fox, Iaria, & Barton, 2009). It is possible that the role of motion cues may become more pronounced with age due to changes in the integrity of the underlying neural structures in the face processing network (Daniel & Bentin, 2012; Grady, 2002; Lee et al., 2011), such as the reduced selectivity of neurons in functional areas, such as the FFA (Goh, Suzuki, Park, 2011; Lee et al., 2011), or the decline in visual contrast sensitivity which may reduce the availability of spatial information pertaining to facial configuration (Goffaux & Rossion, 2006; Norton et al., 2009; Owsley et al., 1981).

In contrast, it was noted that younger adults' face matching performance was consistently very good when matching identity across both novel and learned images of the face. Moreover, for this group both reaction time and accuracy performance were similar when the target face was learned in motion or through a static image sequence. As such, the current results failed to observe a dynamic enhancement which has been previously reported in other perceptual matching studies involving younger adult groups (i.e. Pilz et al., 2006; Thornton & Kourtzi, 2002). This may relate to the nature of the static presentation in the current study. Specifically, previous studies which have reported a dynamic enhancement in matching performance (e.g. Pilz et al., 2006; Thornton & Kourtzi, 2002) typically contrasted performance in that condition with matching performance when a single static image, presented for the same duration, was learned. However, this study attempted to present the same amount of form information about the face in both the dynamic and static conditions.

Furthermore, other studies which presented multiple static views of the target face, or have presented the face rotating in the direction of facial views to be identified, have also reported no relative benefit for rigid motion on subsequent face recognition or matching performance in younger adults (Lander & Bruce, 2003; Lee, Habak, & Wilson, 2010). Moreover, when static images were presented in the current study, the natural sequence for view change was preserved, and this preserved spatiotemporal alignment has previously been demonstrated to be an important factor in both the learning and recognition of faces (Lander & Bruce, 2000, 2003; Schultz et al., 2013). In contrast, a cost in face recognition was reported when the spatiotemporal alignment was disrupted (Wallis & Bülthoff, 2001). Indeed, it was also observed that the time to match novel relative to learned views of the face was significantly slower in younger adults, when the spatial alignment of the static images was disrupted (i.e. jumbled). Preserved spatiotemporal alignment may have been sufficient for younger adults to integrate the multiple views into a robust percept of the face. Furthermore, Schultz et al. (2013) noted that

motion processing areas of the face network were more responsive to a 'natural' motion sequence (i.e. when the spatiotemporal alignment of image frames is preserved) in non-rigidly moving faces, even when the motion was at a rate of 12.5 Hz rather than a more familiar rate of 25 Hz. However, other studies on the role of ageing on basic motion perception *per se* suggest that older age may be associated with a decline in motion perception (Roudaia et al., 2010; Snowden & Kavanagh, 2006). Thus it is possible that in the static condition older adults may have perceived each face image as a discrete event and may have had difficulty combining all images into a single robust representation of the face. This suggests that as we age face recognition may benefit from cue combination, specifically the combination of spatial and temporal cues when representing faces in memory.

The result that face perception in older adults can benefit from dynamic cues, yet older adults also show a decline in other non-facial motion perception tasks, e.g. in discriminating the direction or speed of motion in random-dot patterns or moving gratings (often requiring greater differences in speeds to reliably distinguish between them) (Bennett, Sekuler, & Sekuler, 2007; Snowden & Kavanagh, 2006) has important implications for our understanding of the role of motion in the face processing network. Recent fMRI evidence has demonstrated that facial form areas, largely the occipital face area (OFA), may modulate the transmission of dynamic facial cues from motion sensitive area V5 to the superior temporal sulcus (STS) (Furl, Henson, Friston, & Calder, 2014). Importantly, Furl et al. (2014) suggest that motion sensitivity in STS does not arise due to a single motion based pathway but rather through interactions between form and motion pathways. Furthermore, they noted that activation in the STS region did not appear to be sensitive to motion in random-dot patterns, or dynamic objects. These ageing results compliment these findings and demonstrate that the dynamic enhancement observed in this group likely reflects some level of face specificity. Furthermore, when face images were inverted (i.e. rotated 180° in the picture plane) motion had no facilitative effect on face matching performance relative to static presentations for either the older or younger adults. Thus the benefit of motion on matching faces across changes in visual appearance, which was evident in the older adults' performance when the face was upright, is likely to relate to a motion encoding system which is tuned to upright faces and is sensitive to configural processing. For example, Thornton et al. (2011) noted that the time taken to detect the sex of a moving (compared to static) face increased by an average of 100 milliseconds when the face was inverted relative to upright. This effect appeared to be specific to the processing of faces, as the cost in performance was not evident when processing other classes of moving stimuli, including bodies or animals, and thus cannot be explained as a general difficulty in decoding motion information under inverted

conditions. Thornton et al. (2011) suggested that as inversion interferes with configural face processing, recognition becomes more dependent on feature recognition. The addition of motion may thus make it more difficult to encode and identify such features when the face is upside down.

In summary the findings from this chapter provide further insight into how motion may facilitate face processing and provide evidence in support of the idea that both rigid *and* non-rigid motion can enhance the representation of faces in memory. Furthermore, the results suggest that this dynamic enhancement is related to processing which is particular to faces, rather than a general benefit for additional information or a change in attentional state created by the presence of motion during face learning.

Chapter 4

The Effect of Motion on Face Perception in Developmental Prosopagnosia

4 The Effect of Motion on Face Perception in Developmental Prosopagnosia

4.1. Abstract

There is growing evidence to suggest that facial motion is an important cue for face recognition. However, it is poorly understood whether motion is integrated with facial form information, or whether it provides an independent cue to identity. Here this question was specifically addressed by comparing the effect of motion on face perception in two developmental prosopagnosics and age-matched controls. Participants learned faces presented dynamically (video), or in a sequence of static images, in which rigid (viewpoint) or non-rigid (expression) changes occurred. Immediately following learning, participants matched a static test image to the learned face. Test images varied by viewpoint (Experiment 4) or expression (Experiment 5) and were learned or novel. Similar performance was found for prosopagnosics and controls in matching faces across changes in viewpoint when the learned face was shown moving in a rigid manner. However, non-rigid motion interfered with face matching across changes in expression in both individuals with prosopagnosia compared to performance of the control participants. In contrast, non-rigid motion did not differentially affect the matching of facial expressions across changes in identity for either prosopagnosics (Experiment 6). These results suggest that whilst the processing of rigid motion information may be preserved in developmental prosopagnosia, non-rigid motion can specifically interfere with the representation of structural face information. Taken together, these results suggest that both form and motion cues are important in face perception and that these cues are integrated in the representation of facial identity.

4.2. Introduction

As highlighted in Chapter 2 and 3, studies investigating the processing of both familiar and unfamiliar faces have overwhelmingly relied on the use of static images as stimuli, likely driven by models of face recognition which have argued that facial motion cues were more relevant to social signal processing, rather than identity (Bruce & Young, 1986). Yet, recent evidence suggests that dynamic cues can enhance the processing of facial identity in neurotypical younger adults (Lander & Bruce, 2000, 2003; Lander, Christie, & Bruce, 1999; Lander & Chuang, 2005; Pilz, Thornton, & Bülthoff, 2006; Pilz, Bülthoff, & Vuong, 2009; Thornton & Kourtzi, 2002; for a recent review see Xiao et al., 2014). However, how exactly motion may contribute to the processing of facial identity remains unclear (O'Toole, Roark, & Abdi, 2002; O'Toole & Roark, 2011; Roark, Barrett, Spence, Abdi, & O'Toole, 2003).

On the one hand, as mentioned in Chapter 3, it is argued that motion may afford a salient cue for recognition which is processed independently from facial form information; this is referred to as the supplemental information hypothesis (SIH). Specifically, O'Toole et al. (2002) suggested that motion may provide unique 'dynamic identity signatures' to a person's facial identity which can act as stand-alone, i.e. supplemental, cues for recognition. O'Toole et al. (2002) proposed that these dynamic signatures are likely processed in posterior superior temporal sulcus (pSTS), an area of the face processing network (see Haxby, Hoffman, & Gobbini, 2000) which is primarily concerned with processing the changeable aspects of the face (Haxby, Hoffman, & Gobbini, 2000; Pitcher, Dilks, Saxe, Triantafyllou, & Kanwisher, 2011; Schultz & Pilz, 2009). The dynamic signatures are learned through repeated exposure to the face and thus facial motion may be more relevant for the recognition of familiar faces rather than the learning of new facial identities (O'Toole et al., 2002; Roark et al., 2003). Indeed, studies with neurotypical younger adults have consistently demonstrated that motion benefits *familiar* face recognition when available form cues are degraded (Knight & Johnston, 1997; Lander et al., 1999; Lander & Chuang, 2005). This dynamic enhancement appears to be modulated by the type of facial motion, being more pronounced for non-rigid than rigid motion and also by the degree of idiosyncrasy of the motion displayed (Knappmeyer et al., 2003; Lander & Chuang, 2005). Thus motion (non-rigid) which was once assumed to convey purely social information can provide a supplemental cue to support facial identity processing.

Hill & Johnston (2001) also provided evidence in support of the SIH using a novel paradigm to assess the role of facial motion in discriminating between facial identities. In that study, the authors used motion capture to animate an 'average face' with different dynamic facial

identities. They observed that although the face stimuli provided no reliable visual form cues, observers performed above chance in categorizing and discriminating between facial identities based on the motion cues alone. It is important to note that the observers were unfamiliar with the identities displayed. Thus, although Hill & Johnston (2001) provide evidence for the SIH, demonstrating that facial motion can provide a salient independent cue for recognition, their results also suggest that these dynamic cues are rapidly acquired and are relevant for distinguishing and also learning new facial identities (see also Steede, Tree, & Hole, 2007a).

One additional avenue of research which has also provided support for the SIH comes from a small number of studies which have examined dynamic face processing in prosopagnosia. As highlighted in Chapter 1, prosopagnosia is a disorder characterised by the inability to recognise the identity of an individual from their face alone. Although the disorder can result from damage to an already established face processing system (Bodamer, 1947; Farah, 1990), more recent evidence has highlighted that atypical face recognition can emerge in development in the absence of brain insult i.e. *developmental prosopagnosia* (DP) (Duchaine, Germine, & Nakayama, 2007; Bradley Duchaine, 2008; Susilo & Duchaine, 2013). To date, prosopagnosia has been extensively studied through the use of static face images. These studies have demonstrated that the processing of static form cues in the face is significantly impaired in such individuals (e.g. Bowles et al., 2009; Duchaine et al., 2007; Duchaine & Nakayama, 2005; Németh, Zimmer, Schweinberger, Vakli, & Kovács, 2014; Towler, Gosling, Duchaine, & Eimer, 2012). However, a small number of studies have demonstrated that the ability to extract idiosyncratic motion cues to support identity processing may remain, to some extent, preserved in prosopagnosia (Lander, Humphreys, & Bruce, 2004; Longmore & Tree, 2013; Steede et al., 2007b).

For example, Lander, Humphreys, & Bruce (2004) observed that HJA (a patient who acquired prosopagnosia and visual agnosia following occipito-temporal damage) was unable to use dynamic cues to support familiar face recognition or the learning of new facial identities. Nevertheless, the authors reported that HJA could match the identity of sequentially presented dynamic faces in comparison to static faces. In other words, HJA could use dynamic information for the purpose of face perception but not face recognition. This performance in matching dynamic faces is consistent with studies which reported that HJA was not impaired at matching face parts, relative to whole faces (Boutsen & Humphreys, 2002). Previous studies have suggested that motion perception was unimpaired in HJA (Humphreys, Donnelly, & Riddoch, 1993), therefore HJA may have been able to exploit motion information, independently from facial form, for the purpose of face matching. Other evidence from studies involving developmental prosopagnosics has largely supported Lander and colleagues original findings. Specifically,

although evidence for a benefit for motion on face memory has been inconsistent (Esins, Bülthoff, & Schultz, 2014; Longmore & Tree, 2013; but see Steede et al., 2007a), the ability to match moving faces has been reliably observed. For example, Longmore & Tree (2013) reported better face matching performance across changes in viewpoint in individuals with DP, when the same idiosyncratic non-rigid motion was available in the face stimuli during the learning and test conditions, compared to when all images were static in nature. In addition, Steede et al. (2007a) observed that CS, a developmental prosopagnosic, could reliably discriminate between facial identities when only motion cues in the face were available (see also Steede, Tree, & Hole, 2007b), irrespective of whether the motion was rigid or non-rigid. Taken together these results suggest that the ability to extract motion information for the purpose of perceiving unfamiliar faces (i.e. to match and discriminate newly learned facial identities) may remain relatively intact in cases of DP. On the other hand, the evidence suggests that facial motion may not facilitate memory for faces in DP, suggesting that facial motion may be difficult to represent in this cohort (Lander et al., 2004).

In terms of neurotypical face processing, the findings from Chapter 3 suggest better perception of unfamiliar faces in older adults when the face was initially learned in motion relative to a sequence of static images. Ageing is often associated with a decline in the ability to perceive and recognise unfamiliar faces from static cues alone (Habak, Wilkinson, & Wilson, 2008; Lee, Grady, Habak, Wilson, & Moscovitch, 2011). However, the results from Chapter 3 demonstrated that when an unfamiliar face was presented in motion during learning, this motion subsequently benefited the matching of novel (static) images of that face. Thus, the performance enhancement observed for learning faces in motion suggests that facial motion may combine with available form information to create a more robust structural representation of the face, i.e. preserving identity matching across image transformations of the face (representation enhancement hypothesis or REH), at least over short term intervals. This demonstrates that motion does not only act as a source of information which is extracted independent of facial form information (SIH).

The processing of motion for the purpose of face perception may be underpinned by unique neural substrates such as the pSTS. The pSTS is considered an area within the face processing network (see Haxby, Hoffman, & Gobbini, 2000) which is primarily concerned with processing the changeable aspects of the face (Haxby et al., 2000; Pitcher et al., 2011; Pitcher, Duchaine, & Walsh, 2014; Schultz & Pilz, 2009). However, as highlighted in Chapter 3, recent neuroimaging evidence also suggests that interactions between form and motion areas in the face processing network may be more pronounced than previously assumed (Furl, Henson, Friston, &

Calder, 2014; Schultz & Pilz, 2009; Schultz, Brockhaus, Bülthoff, & Pilz, 2013). For example, using functional magnetic resonance imaging Schultz & Pilz (2009) confirmed prior observations that STS plays a pivotal role in processing dynamic aspects of the face, with enhanced cortical activation to faces moving in a non-rigid manner, compared to static face images (see also Fox, Iaria, & Barton, 2009; Pitcher et al., 2011) and similar findings have also been suggested for rigid motion (Lee et al., 2010). Interestingly they also observed that classically defined static form processing areas (e.g. fusiform gyrus) also exhibited this enhanced functional activation profile (albeit to a lesser extent than STS) for dynamic compared to static faces, leading the authors to conclude that the integration of facial form and motion likely occurs in both ventral and dorsal areas of the face processing network. However, very recently it has been argued by Pitcher, Duchaine, & Walsh (2014) that the dissociation between static and dynamic face processing may be strict. Specifically, they observed that disrupting the processing of occipital face area (OFA) in neurotypical individuals, through the delivery of theta burst transcranial magnetic stimulation (TBS), did not affect the functional response profile in pSTS to moving face images, yet it did reduce the pSTS response profile to static face images. However, as there were no task demands in that study it is difficult to conclude what aspects of face processing may have been sub-served through the functional activity in pSTS. Moreover, Furl et al. (2014) argued that facial motion sensitivity in STS does not arise due to a single motion-based pathway but rather through interactions between form (OFA) and motion (V5) pathways (Furl et al., 2014). Indeed, these network interactions have been proposed as a possible explanation for the functional selectivity of different areas in STS to a range of stimuli (Furl et al., 2014; Hein & Knight, 2008); including moving bodies (Blake & Shiffrar, 2007; Grossman & Blake, 2002), voice information (Belin et al., 2011, 2004) and face and voice integration (Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004). Therefore, interactions between form and motion areas of the face processing network may provide an important insight into our understanding of how face recognition arises in the neurotypical brain.

In summary, there is some evidence to suggest that facial motion is an important source of information for face processing, with studies suggesting that facial motion can improve the perception of both familiar and unfamiliar faces. In the case of prosopagnosia, there is some suggestion that access to facial motion processing may be preserved, suggesting that motion may support identity processing in DP (Longmore & Tree, 2013; Steede et al., 2007a). Thus, in the same way that dynamic cues can support face recognition in neurotypical younger adults when available form cues are degraded (e.g. Lander et al., 1999) this processing of independent sources of information may benefit face perception in prosopagnosia. Taken together, these results

support a possible dissociation between the mechanisms involved in static and dynamic face processing (i.e. SIH). However, the results from Chapter 3 and a number of previous studies (e.g. Pilz et al., 2006; Thornton & Kourtzi, 2002) also demonstrate that learning a face in motion can enhance the structural (i.e. form) representation of the face, suggesting that static and dynamic cues are likely to interact when encoding facial identity, in line with the representation enhancement hypothesis (REH). It is therefore, unclear whether motion information will enhance or interfere with the processing of facial information in individuals with DP since, to the best of our knowledge, no study has explicitly tested how facial motion may affect the representation of facial form in this population.

4.2.1. Outline of the Experiments

The series of face matching experiments reported below examined how learning a face in motion, relative to static presentation, may affect the subsequent ability to match an image of that face across changes in viewpoint and expression. Performance on these tasks was tested in two apperceptive developmental prosopagnosics, UM and PL (see Chapter 2 for full characterisation of each prosopagnosic), and compared to the performance from a group of age-matched neurotypical individuals (control group). A matching paradigm was used as this has been previously shown to be sensitive to the effects of motion on face learning in neurotypical younger (Pilz et al., 2006; Thornton & Kourtzi, 2002) and older (Chapter 3) adults, as well as being sensitive to detecting idiosyncratic motion processing in prosopagnosia (Lander et al., 2004; Longmore & Tree, 2013). Separate experiments investigated how rigid motion (Experiment 4) and non-rigid motion (Experiment 5) may affect face perception using an identical immediate matching paradigm for both types of motion. It was also examined if (non-rigid) motion cues in the face can be used to support the perception of social information from a face, rather than identity, in DP (Experiment 6).

4.3. EXPERIMENT 4

Experiment 4 tested participants' ability to match unfamiliar faces depending on whether the face was initially viewed moving or not. Specifically, both UM and PL and an age-matched group of participants first learned an unfamiliar face shown either moving (rigid motion of the head) or through a sequence of static images extracted from the same moving sequence. In each trial, following the learned face, participants were then immediately provided with two static images of faces and were asked to identify the face image which matched the previously learned face. The target face image was shown from a learned or a novel viewpoint. In order to assess

how motion affects the structural representation of the face, matching accuracy was compared for novel relative to learned face images, across static and dynamic learning conditions, for both prosopagnosics and controls. Face recognition can be said to occur when identity can still be matched despite changes in viewpoint or expression, as this type of matching relies on the ability to represent the essential structure of the face, rather than image characteristics of the learned face (Bruce, 1982).

It was predicted that if form and motion information are related cues for perceiving faces, then matching static face images may benefit from the presence of motion in the learned face. In neurotypical individuals at least, learning a face in motion may enhance, or at least not detract from, encoding the structural representation of the face (Chapter 3). On the other hand, in the case of prosopagnosia it is not clear whether motion may enhance or further impair face perception. For example, face motion may provide an intact source of information for learning faces, and can improve matching performance across moving faces (i.e. motion to motion matching), as previous studies have shown. However, given that the current task involves matching faces based on their form information only (i.e. the test images are static), then it is possible that the presence of motion during face learning may detract from, rather than enhance, the structural representation of facial identity in DP. Moreover, these effects might be modulated by the type of facial motion i.e. rigid or non-rigid (Hill & Johnston, 2001; Lee, Habak, & Wilson, 2010).

4.3.1. Method

Participants

The data reported here for control participants are those reported in Chapter 3, Experiment 1A for the older adult group. As such, the same older adult data was used as a contrast for both UM and PL's performance on this task. Note that the older adult group are aged-matched (mean age = 72.9 years) to both UM (76 years) and PL (73 years).

Stimuli and Apparatus

The face stimuli are as described in Chapter 3, Experiment 1A. Briefly, the stimuli consisted of video clips of 20 different actors, moving their head in a rigid manner. Each video consisted of a motion sequence in which the actor's head faced the front (0°) and rotated to their right (45°) referred to as a '*turn to right*' sequence or one in which the actor faced the left (45°) and rotated their head towards a frontal view (0°), referred to as a '*turn from left*' sequence. A

static image sequence (3 images) was extracted from each sequence. Static test images contained a learned 0° or a novel 45° image.

Design

The design is as described in Chapter 3, Experiment 1A. Specifically, the experiment consisted of 80 trials in total. In half of the trials the target face was presented dynamically (i.e. a video), while in the other half of the trials 3 sequential static images of the target face were shown. As mentioned in Chapter 3, during the presentation of the target face, half of the age-matched control participants were shown the *turn to right* sequence in the learning phase and then presented with the 45° left view of the actor's face as the novel viewpoint in the test, while the other half were shown the reverse i.e. the *turn from left* sequence as the target face and then the 45° right view of the actor's face as the novel viewpoint, as results from Chapter 3 indicated that both sequences produced similar performance. However, for consistency both the prosopagnosic participants were exposed to the *turn to right* sequence in the learning phase and then presented with the 45° left view of the actor's face as the novel viewpoint in the test. The two test face images were always presented from the same view, and one always matched the target face. In half of the trials the test images were presented from a frontal, or 0° view, which was a view always shown during the presentation of the learned face (i.e. learned). In the other half of trials, test images were of a 45° view of the face which was a view not previously presented in the target image (i.e. novel). An illustration of a trial including example stimuli is shown in Chapter 3, Figure 3.1.

Procedure

The procedure is as described in Chapter 3, Experiment 1A. In brief, all participants were instructed to learn the face presented in the centre of the screen and, following a short interval (500ms), to indicate which of two face images matched the identity of the face they had learned i.e. a two alternative forced choice task. One of the images presented always matched the learned facial identity. Test images were always static and displayed either a learned (40 trials) or novel (40 trials) viewpoint of the target face.

Analysis

Viewpoint Change Cost

Due to the variability in performance between the age-matched older adults and developmental prosopagnosics, similar to Chapter 3, performance was calculated on this task as the *relative* difference in matching accuracy (percent correct) for the learned and novel viewpoints, normalised to each individual's own 'baseline' accuracy performance for matching the learned viewpoints $[(\text{learned} - \text{novel}) / \text{learned} * 100]$. This score (*viewpoint change cost*, see Chapter 3, Experiment 1A) was calculated for both prosopagnosics, as well as each individual in the age-matched group. The absolute accuracy scores for matching learned and novel viewpoints of the face, across static and dynamic learning conditions for each prosopagnosic and the age-matched control group can be viewed in Table 3.1.

A paired t test was used to compare the viewpoint change cost for both static and dynamic target face learning conditions, for the age-matched older adult group. The Revised Standardized Difference Test (RSDT) (revised see Crawford et al., 2010; original see Crawford & Garthwaite, 2005) was used to compare the viewpoint change cost in prosopagnosics relative to controls, for both the static and dynamic learning conditions. The RSDT assesses whether the observed difference between a case's score on two tasks X and Y, in this case static or dynamic learning conditions, differs from the observed difference between task scores in X and Y in a control or normative population. It is an appropriate test even with a modest control sample size and has been used in the prosopagnosic literature to compare performance across different experimental conditions, as applied in the current context (e.g. Longmore & Tree, 2013). The test provides a t score and a point estimate of the effect size for the difference between the case and controls (Z-DCC) and an accompanying 95% confidence interval.

4.3.2. Viewpoint Change Cost Results

The cost in face matching performance with a change in viewpoint is shown for each of the prosopagnosic individuals and their age-matched controls in Figure 4.1. In terms of age-matched controls performance analysis revealed that the cost associated with matching a face across novel viewpoints was similar across both learning conditions [$t(13)=1.48$, $p=0.16$] although it was more pronounced for the static (M cost=8.9%) than dynamic (M cost=2.8 %) learning conditions. The results demonstrated that both UM [$t(13)=0.35$, $p=0.37$] and PL [$t(13)=0.96$, $p=0.18$] showed a similar pattern of performance to that of the control group, such that the cost associated with matching the face across a novel viewpoint was similar for both learning

conditions (see Figure 4.1, 4.2). Note that Figure 4.1 plots UM and PL’s viewpoint change cost and the mean performance of the age-matched controls. However, in Figure 4.2 both UM and PL’s viewpoint change cost is plotted in relation to each individuals’ score in the age-matched control group.

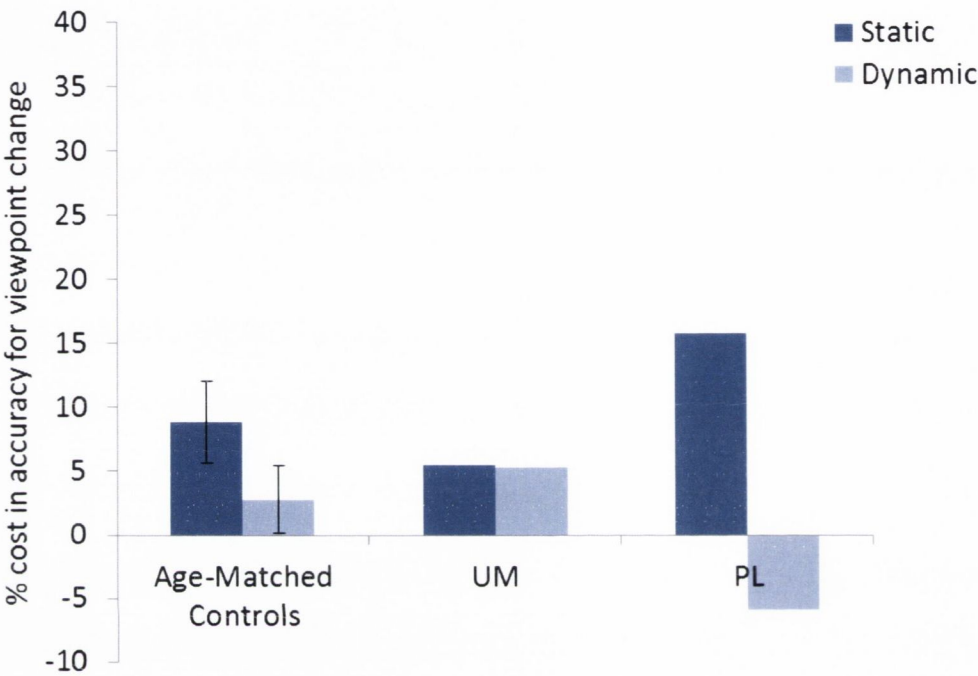
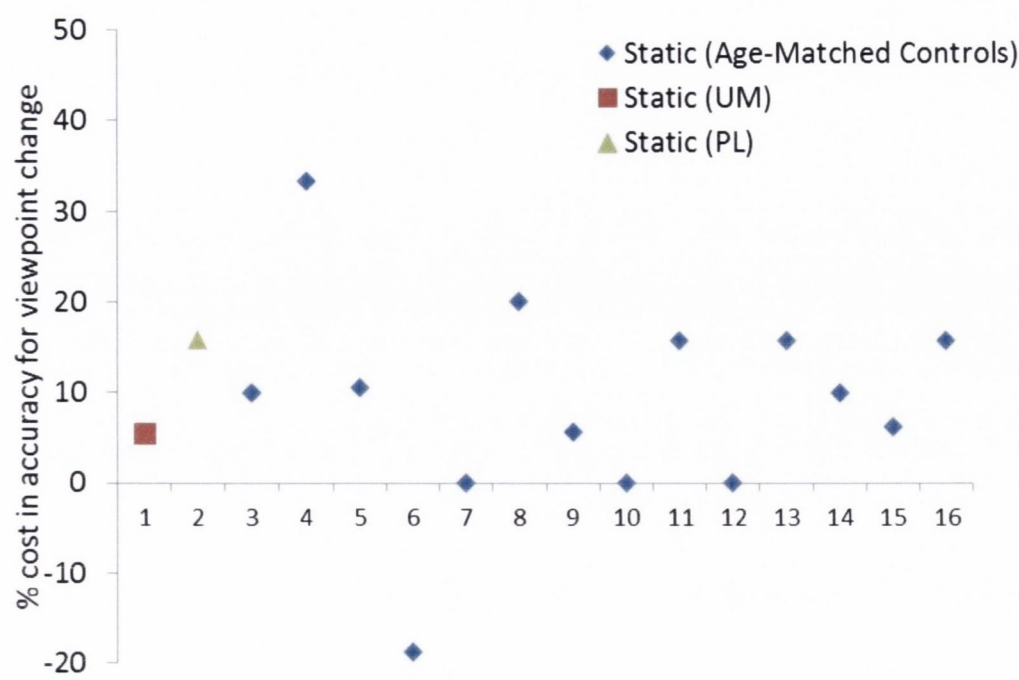


Figure 4.1. Plot showing the mean cost in accuracy for matching a novel relative to a learned viewpoint of the target face, across each of the prosopagnosics and their age-matched controls tested in Experiment 4. The results are shown for each of the static or dynamic (rigid motion) target face learning conditions. Error bars show ± 1 standard error of the mean (SEM).

a)



b)

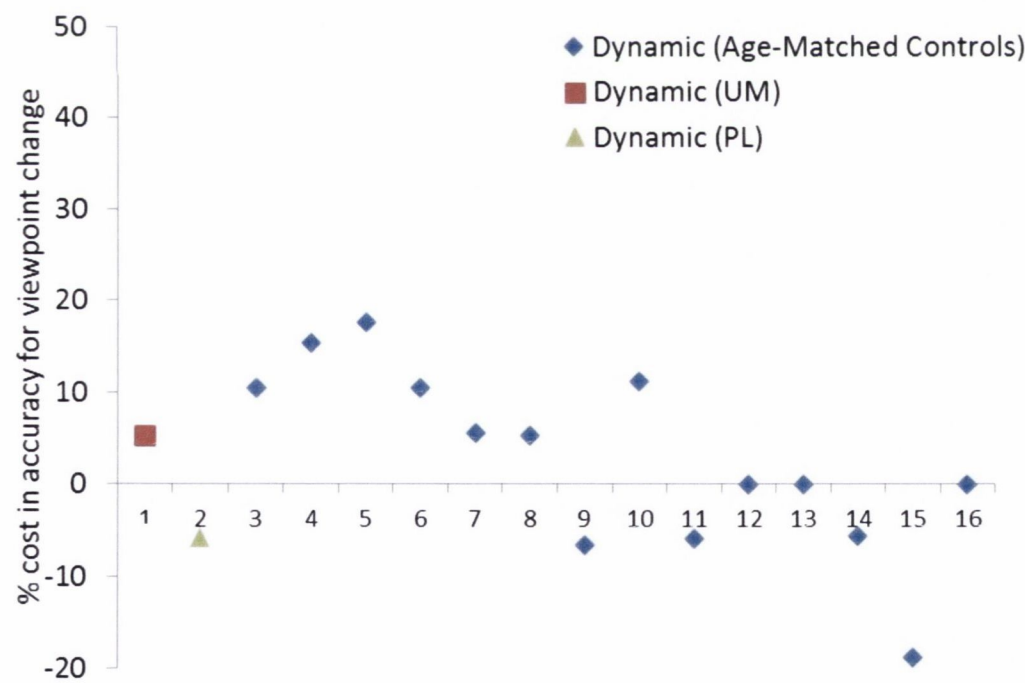


Figure 4.2. Plot showing the cost in accuracy for matching a novel relative to a learned viewpoint of the target face, for each of the age-matched controls (n=14) and both prosopagnosics tested in Experiment 4 (rigid motion). The Y axis represents each participant tested. The results are shown separately for a) the static (upper plot) or b) dynamic (lower plot) target face learning conditions.

4.3.3. Discussion

The results suggest similar face matching performance across both age-matched controls and the prosopagnosic individuals. Specifically, performance for matching faces across novel viewpoints was similar when the face had been learned moving in a rigid manner or from a series of static images. Furthermore, performance was relatively good across these matching conditions in all individuals (> 80% in any condition). This suggests that rigid motion did not detract from the encoding of facial form cues in DP. However, it is possible that this may relate to the type of motion used. Specifically, rigid motion may still facilitate image matching strategies as the internal features of the face remained constant. Indeed, other studies have demonstrated that prosopagnosics can face match, although they may engage unusual, or non-face specific strategies to do so, such as direct feature matching (e.g. McKone et al., 2011).

Table 3.1. Individual accuracy scores for UM and PL and mean percent accuracy scores for age-matched controls (with standard deviations) for matching facial identity across learned and novel views (Experiment 4) and expressions (Experiment 5) of the face.

	UM		PL		Age-Matched Controls	
	Static	Dynamic	Static	Dynamic	Static	Dynamic
<i>Experiment 1: Rigid Motion</i>						
Learned View	90	95	95	85	89.6 (10.5)	87.5 (10.0)
Novel View	85	90	80	90	81.4 (12.0)	85.0 (11.9)
<i>Experiment 2: Non-Rigid Motion</i>						
Learned Expression	95	95	95	90	91.1 (8.4)	91.1 (6.8)
Novel Expression	85	65	75	55	76.1 (16.4)	83.9 (8.1)

4.4. EXPERIMENT 5

The purpose of Experiment 5 was to examine how learning a face in non-rigid motion, compared to a static image sequence, may affect the subsequent matching of the facial identity across changes in expression in UM and PL, relative to age-matched controls. It is likely that, compared to rigid motion, non-rigid motion may present a particular challenge for DPs in the current matching task. In contrast to the rigid motion, non-rigid motion is characterised by the change in the *internal* features of the face over time. As such it may be difficult to produce normal

matching performance when the face has been learned in non-rigid motion, compared to rigid motion, by employing an image or feature matching strategy (Thornton, Mullins, & Banahan, 2011). Indeed, while Steede et al. (2007b) noted that developmental prosopagnosic CS could use facial motion for discriminating and learning identities, they suggested that CS may have been using a learning strategy which differed from controls; one which was particularly centred on rigid (e.g. head shaking), compared to non-rigid motion (e.g. idiosyncrasy of a smile). As such, in comparison to the results from Experiment 4 the presence of non-rigid motion during face learning, may render face identity matching across novel expressions, based on form cues only (i.e. the test images are static), error prone in this population.

4.4.1. Method

Participants

In addition to UM and PL, 14 age-matched older adults participated in this experiment (9 female, all right handed, mean age= 70.2 years; SD= 6.1 years). Seven of these participants are reported in Chapter 3, Experiment 2. As such, these same 7 datasets were used as a comparison for UM and PL's performance. These participants had learned the 'happy' expression as the target face in the 2AFC face matching task. Ten additional older adults were also recruited for this task, however 3 datasets were subsequently removed as it was revealed that the individuals had cataracts at the time of testing. New participants were recruited in order to maintain consistency for the target learning condition (i.e. 'happy' expression as the target face). None of these participants had participated in any of the experiments reported in Chapter 3. All participants scored above the cut off of ≥ 24 of 30 on the MoCA (mean score=27, SD=0.6). Older participants had normal visual acuity for their age, as measured at 4 meters (mean binocular LogMar score= 0.05, SD= 0.04) and 40cm (mean binocular LogMar score= 0.02, SD= 0.06). Contrast sensitivity was also normal for age and was assessed using the Pelli-Robson Contrast Sensitivity Test (mean score= 1.9, SD= 0.08). All participants wore appropriate corrective lenses, if needed, at the time of testing.

Stimuli and Apparatus

The stimulus set consisted of 20 facial identities described in detail in Chapter 3, Experiment 2. Briefly, each video, or 3 static image sequence, displayed the actor transitioning from a neutral to a happy or angry expression. Test images were static images of the apex of the expression (happy or angry).

Design

All other details of the design are as reported in Experiment 4, with the exception that target and test images varied by expression and not viewpoint. The two test faces were always static images and were taken from the last frame of the video sequence. In half of the trials the test images were of the apex of the expression shown during the target face presentation. In the other half of trials the images were of a novel unseen expression which was not seen during the target face presentation. All age-matched control and prosopagnosic participants, were shown the 'happy' expression as the target face and the angry expression as the novel expression (with identity unchanged in matching target and test face image). An illustration of a trial including example stimuli can be seen in Chapter 3, Figure 3.4.

Procedure

The procedure was mainly as described in Experiment 4 with the exception that participants were required to match the identity of a previously learned face (moving or static) in face images in which the facial expression was the same or novel.

Analysis

Expression Change Cost

The analysis of the accuracy data is as described in Experiment 4 with the exception that a relative *expression change cost* was calculated for each age-matched control and prosopagnosic participant. This score consisted of the relative difference in accuracy for matching faces across learned and novel expressions, normalised to each individual's own 'baseline' accuracy performance for learned expressions of the face i.e. $[(\text{learned} - \text{novel})] / \text{learned} * 100$. This was calculated separately for both static and dynamic trials.

4.4.2. Expression Change Cost Results

The expression change cost was compared for static and dynamic learning conditions in age-matched controls using a paired *t*-test. This analysis revealed that the cost associated with matching the face across novel expressions was significantly reduced when the face had been learned in motion relative to static presentation [$t(13)=2.43$, $p=0.03$]. In line with Experiment 4 the expression change cost was also compared for each prosopagnosic individual across each of the learning conditions (i.e. in motion or through static images), to the expression change cost for

the age-matched controls performance using the Revised Standardised Difference Test (Crawford et al., 2010). Performance across each prosopagnosic individual and the mean performance for the control group is shown in Figure 4.3. UM and PL's performance is also plotted with each individual's ($n=14$) expression change score in the control group, for static and dynamic learning conditions in Figure 4.4. These analyses revealed that for both UM [$t(13)=3.37$, $p=0.005$, $ZD_{CC}=-3.80$, 95% $CI=-5.60$, -2.28] and PL [$t(13)=3.45$, $p=0.004$, $ZD_{CC}=-3.88$, 95% $CI=-5.86$, -2.21], the cost associated with matching the face across a novel expression was significantly increased when the face had been learned in motion rather than static images. In other words, the performance of both prosopagnosic individuals was the reverse pattern of performance to that of the control group whose performance benefitted from the provision of motion during face learning.

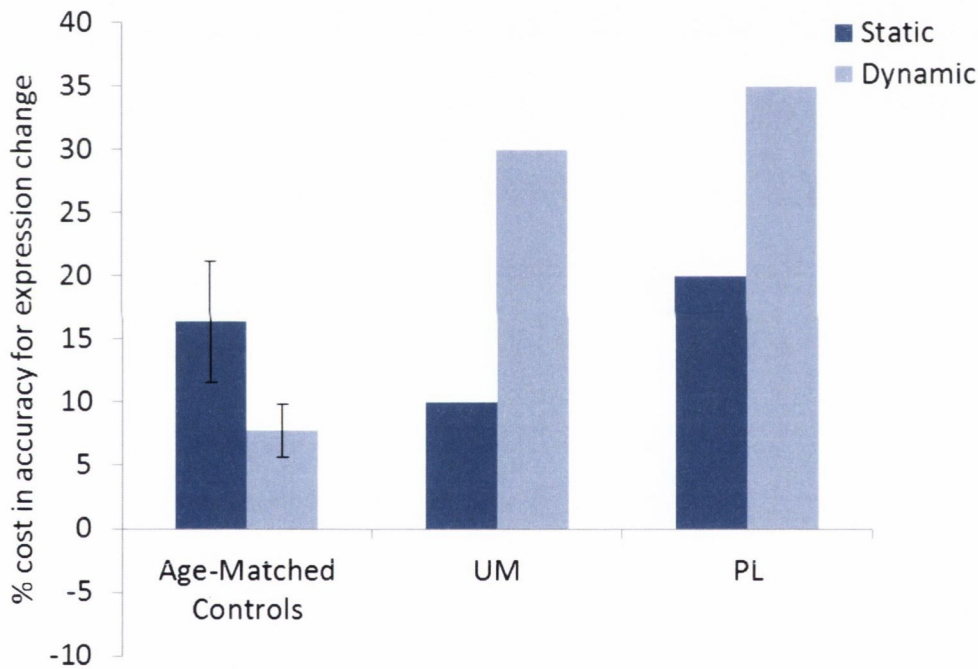
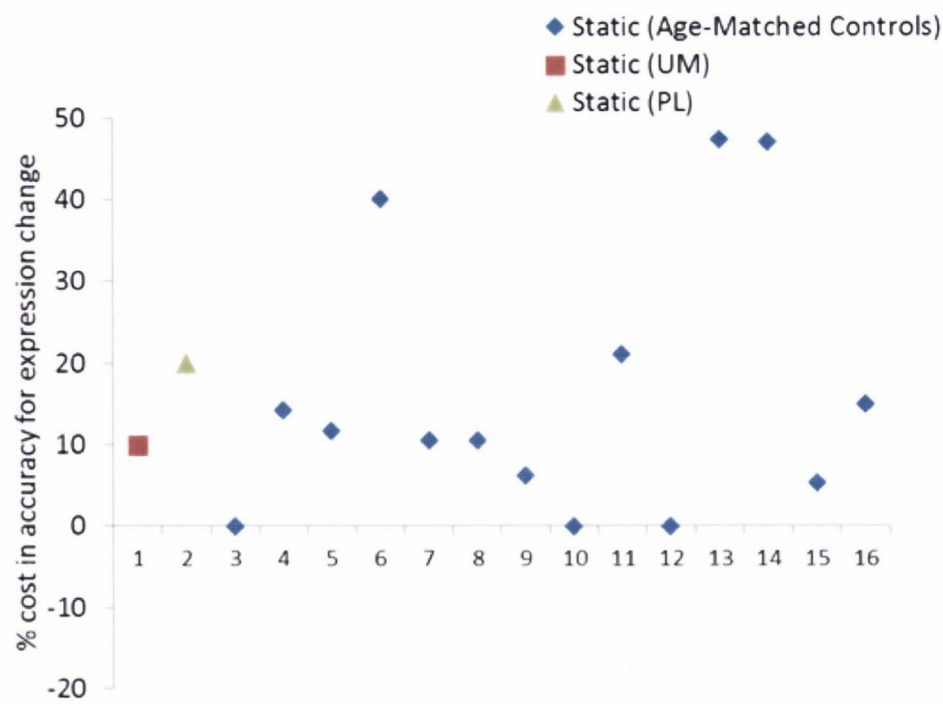


Figure 4.3. Plot showing the mean cost in accuracy for matching a novel relative to a learned expression of the target face, for age-matched controls and prosopagnosics tested in Experiment 5. The results are shown for each of the static or dynamic (non-rigid motion) target face learning conditions. Error bars show ± 1 standard error of the mean (SEM).

a)



b)

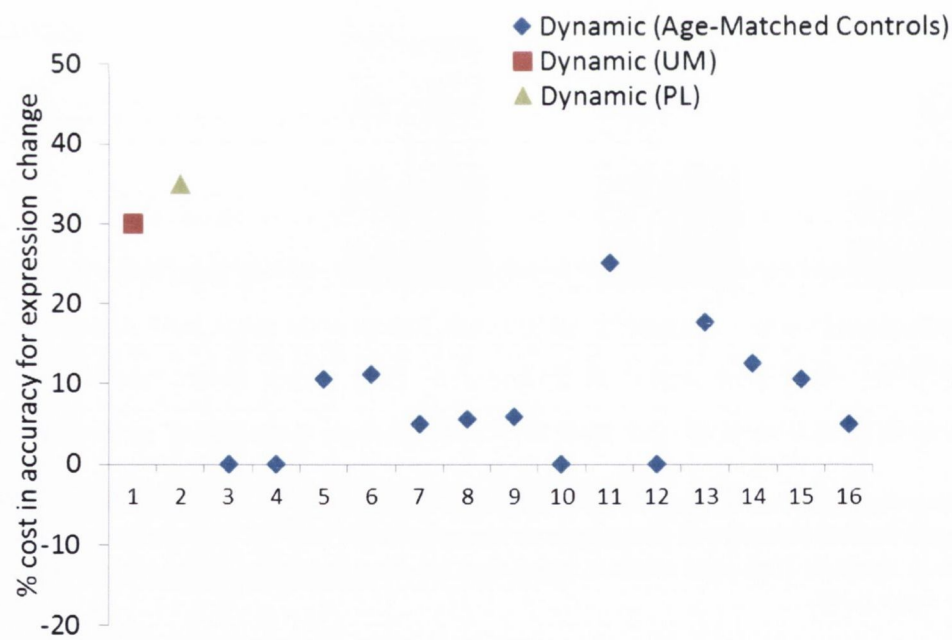


Figure 4.4. Plot showing the cost in accuracy for matching a novel relative to a learned expression of the target face, for each of the individual age-matched controls (n=14) and both prosopagnosics tested in Experiment 5 (Non-Rigid Motion). The Y axis represents each participant tested. The results are shown separately for the a) static (upper plot) or b) dynamic (lower plot) target face learning conditions.

4.4.3. Discussion

The results from Experiment 5 demonstrate that in neurotypical older adults the cost in accuracy performance when matching facial identity across changes in expression can be reduced when the face was previously learned in motion rather than through static snapshots. Indeed, the cost in performance decreased by almost 10% (i.e. from 16.4 % in the static learning condition, to 7.8 % in the dynamic learning condition). Thus, it appears that motion may promote face constancy across incidental changes in image characteristics of the face, particularly in an older age group. Indeed, this result is in line with the older adult results from Chapter 3, Experiment 2, where a similar reduction in error associated with matching facial identity across novel expressions was observed i.e. approximately 10 %. Furthermore, this dynamic enhancement appears greatest when the learned face is presented moving in a non-rigid manner, in comparison to rigid motion of the face (which incurred a non-significant enhancement of approximately 6%, as shown in Experiment 4). However, in contrast to the performance of the control participants, non-rigid motion significantly interfered with matching identity across changes in expression in both UM and PL. For both of these individuals, the number of face-matching errors increased when the learned face was moving rather than when it was presented through multiple static images. This finding suggests that non-rigid facial motion detracted from the encoding of facial form information in this population.

4.5. EXPERIMENT 6

The results from Experiment 5 suggested that non-rigid motion interfered with prosopagnosics' ability to match facial identity across changes in expression. It is possible that this may relate to a general deficit in the processing of non-rigid motion, or to a specific deficit in the processing of non-rigid motion for facial identity only. In an attempt to tease apart these possibilities a similar matching paradigm to that reported in Experiment 5 was used, with an essential change to the task instructions: in the following task both PL and UM were required to match the learned expression of the face, across static and dynamic learning conditions, from two test face images which displayed either the same facial expression or not (irrespective of facial identity).

4.5.1. Method

Design

In line with the design of Experiment 5, both UM and PL were presented with the 'happy' expression as the target expression. Similarly, in half of the trials, faces were presented in motion during learning, while in the remainder of the trials, the target face was displayed as a static image sequence. As before, a factorial design was used during the test phase between facial identity and facial expression. As such, half of the test faces (40 trials) were matched on facial identity and facial expression, while the other half of the test faces (40 trials) were matched on facial expression but not on identity. When the face matched on identity and expression (i.e. matched to the target face), half the trials contained a distractor test image of a same (20 trials, see A of Figure 4.5) or different identity (20 trials, see C of Figure 4.5) to the target that did not match the facial expression of the target face. When the face did not match on identity, but only on expression, half the trials contained a distractor test image of a same identity (20 trials, see B of Figure 4.5) or different identity (20 trials, see D of Figure 4.5). During the test phase, faces were presented in pairs based on a 2AFC protocol as before, and only one face in each pair matched the expression of the learned face, irrespective of identity. There were 80 trials in total. See Figure 4.5 for an illustration of the four possible test image combinations.

Procedure

The procedure is as described in Experiment 5, with the exception that the participant matched the expression (as opposed to the identity) of the learned face by choosing from two available static images onscreen.

4.5.2. Results and Discussion

Both UM and PL scored an average approaching 100 percent correct in identifying the target expression across all possible test image combinations. Furthermore, performance across static ($M=97.5\%$, 100% for UM and PL, respectively) and dynamic ($M=100\%$, 100% for UM and PL, respectively) conditions was equivalent. This finding suggests that facial motion did not interfere with the processing of social signals, such as facial expression. Moreover, the participants' performance suggests that motion does not detract from encoding a representation of the face which is used to support expression analysis, as both prosopagnosics could reliably match

expressions across identities even when the face had been learned as moving rather than static sequence.

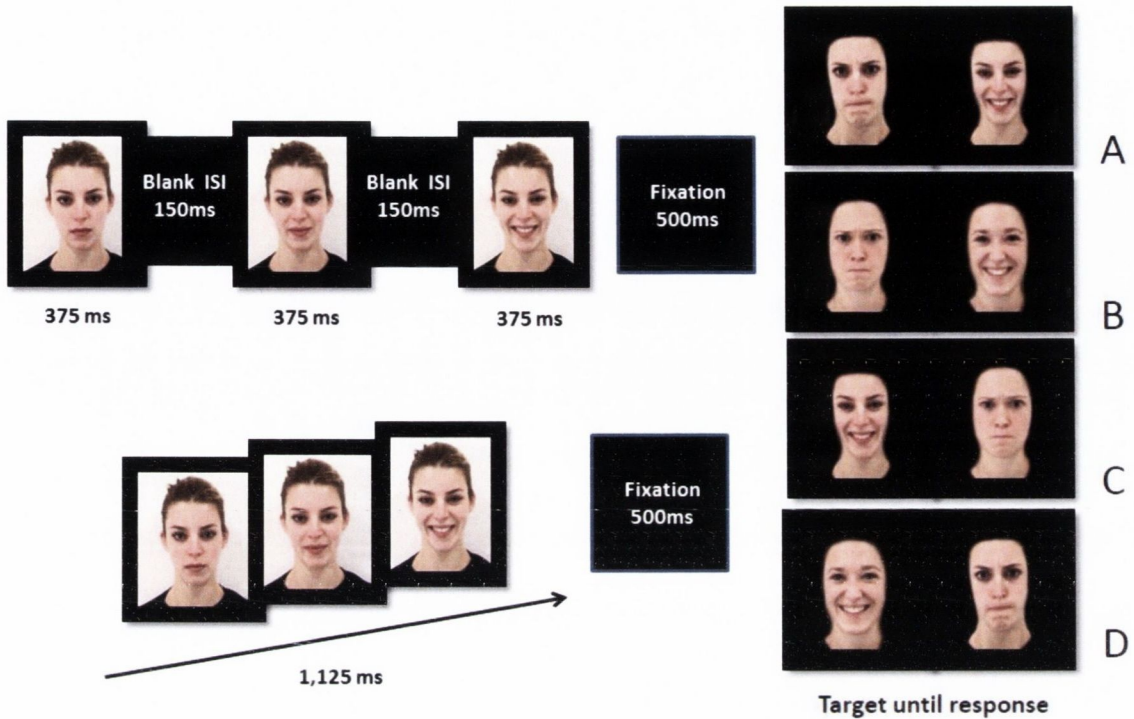


Figure 4.5. An example of a trial which illustrates the face images shown in the static (upper panel) and non-rigid motion (lower panel) training conditions used in Experiment 6. The pairs of images (A-D) shown on the right depict the different test phases and include an image of the target and the distractor expressions, with different identity combinations. In A both test images depict the same identity as the learning condition, while in B both images depict a different identity to the learning condition. In C and D both test images depict the same and different identities to the learning condition. Note that expression and identity overlap in C and not D.

4.6. General Discussion

A small number of studies have investigated facial motion processing in DP, with the purpose of examining whether motion may provide a cue for identity which is accessed independently from facial form information (Longmore & Tree, 2013; Steede et al., 2007b). In contrast, the aim of the present study was to examine how facial motion and form may *interact* in face processing in DP by explicitly examining how learning a face in motion affected the structural representation of the face in this population. Here, two developmental prosopagnosics, UM and PL, and a group of age-matched controls learned a face displaying rigid (Experiment 5) or non-rigid facial changes (Experiment 6). Crucially, the learning condition was manipulated such that the face was either learned in motion or through a sequence of static snapshots extracted from the moving sequence. Participants were then tasked with matching the face identity from two

available static images, which displayed either a learned or a novel image of the face. The results from Experiment 5 demonstrated that matching the face across novel viewpoints was similar when the face had been learned in rigid motion or from a static presentation, for both prosopagnosics and age-matched controls. In marked contrast, the presence of non-rigid motion during learning significantly impaired both UM and PL's ability to match the face identity across novel expressions. Importantly, this performance profile was reversed for age-matched controls, such that learning the face in non-rigid motion significantly enhanced subsequent matching performance for novel expressions of the face relative to when the face was learned as static images. The results from Experiment 6 demonstrated that both UM and PL could reliably match facial expressions, both within and across identities, when the face displayed the same non-rigid expression changes which were present during the identity matching task in Experiment 5.

The finding that developmental prosopagnosics can reliably use rigid motion, to the same degree as controls, when encoding structural facial form information is consistent with previous reports of preserved idiosyncratic rigid motion matching in DP (Steede et al., 2007b). This similar matching performance across prosopagnosics and controls may relate to the nature of the face processing which is elicited by this type of motion sequence. For example, it has been argued that rigid motion may bias the processing of feature based, rather than holistic, face information (Xiao, Quinn, Ge, & Lee, 2012). Thus, it is possible that when learning faces shown moving in a rigid manner, prosopagnosics may have been able to recover feature based information about a face, since the internal features of the face remained constant as the face moved from side to side. Importantly, for the control group the benefit on matching the face across changes in visual appearance was relatively small for faces learned in rigid, in comparison to non-rigid motion. Matching the face across novel expressions appears to be a harder task (see Table 3.1). Indeed, it is difficult to produce good performance on this task by adopting a feature matching strategy, as internal deformations of the face render this strategy less reliable. Rather it is possible that non-rigid motion may confer an advantage in face perception which is largely restricted to the main mechanisms underlying face processing i.e. holistic processing (Yin, 1969). For example, as mentioned in Chapter 3, Thornton et al. (2011) demonstrated that non-rigid motion interfered with matching the sex of a face, by increasing reaction times, when face stimuli were inverted relative to upright. However, when the face was presented as a static image there was no such interference. They argued that when face images are inverted the perceptual system is biased towards feature based processing, rather than holistic processing, which occurs when the face is upright. However, internal deformations of the face that occur as a consequence of non-rigid

motion make it more difficult to encode such features, likely resulting in the further response delay to categorise the sex of the face.

In the current study non-rigid motion significantly enhanced the representation of the face in age-matched controls, resulting in more accurate performance for matching the face across novel expressions, compared to when the face had been learned through a sequence of static images. However, the presence of non-rigid motion during face learning made it difficult for both prosopagnosics to recover facial form information, rendering matching across novel expressions relatively more error prone in this condition. Taken together, these seemingly conflicting results give rise to the same conclusion, that these form and motion cues likely influence each other at the encoding stage at least. However, in the case of DP, motion detracts from encoding facial form, while it enhances the representation of the face in neurotypical controls. This suggests behaviourally relevant correspondence between static form and dynamic facial cues in neurotypical face processing.

This result is consistent with recent evidence supporting the idea that there is a significant degree of overlap between dynamic and static representations of identity in the neurotypical face processing system (Jeffery, Petrovski, & Rhodes, 2014; O'Toole et al., 2014). For example, O'Toole et al. (2014) observed similar patterns of neural activity in ventral and dorsal areas of the face network, in response to faces either presented as moving or as static images. Furthermore, this pattern was bidirectional, such that the neural patterns for representations created from static face images transferred to those elicited by dynamic faces and vice versa. While the pattern of activation was transferable it is important to note that the response magnitude of the activation varies across face areas depending on the dynamic nature of the stimulus (e.g. stronger activation in STS to moving relative to static face images). Nevertheless, O'Toole et al. (2014) argue that there is some degree of redundancy and transfer across representations created from static or dynamic faces. It is possible that these effects may arise due to interactions between form and motion processing areas (e.g. Furl et al., 2014) or due to the sensitivity of both ventral and dorsal areas to dynamic face information, as highlighted in Chapter 3 (Fox et al., 2009; Schultz & Pilz, 2009; Schultz et al., 2013). This suggests that it may be the interactions between form and motion areas which may be compromised in DP, as both UM and PL were unable to recover structural information to support identity processing when the face had been learned in motion. In contrast, when the face had been learned through static presentation, their performance for matching the face across novel expressions was comparable to controls (see Table 3.1).

Indeed, this conclusion parallels a case of acquired prosopagnosia reported by Cowey & Vaina (2000). While most cases of acquired prosopagnosia have documented the individual's inability to recognise a face from static face images, with a self-reported deficit in recognising even the most familiar faces, Cowey & Vaina (2000) report a rather unique case of AL, a 65 year old lady who suffered extensive damage to the ventral extrastriate cortex as the result of a stroke. After the stroke AL presented with severe object agnosia and prosopagnosia. However, her ability to recognise facial identity gradually returned but with a rather profound limitation: while, AL could recognise the identity of a person's face when it was static, she found it impossible to recognise a face when it was in motion, although she could see and describe the movement patterns (intact idiosyncratic motion processing). As such, AL's deficit in the ability to recognise the face in motion was not related to an inability to process its static structural form, or from an inability to decode facial movement, rather it was the interactions between these two sources of visual information which gave rise to her significant behavioural impairment. Indeed, Cowey & Vaina (2000) concluded that AL's deficit in visual perception likely arose from a dysfunctional connectivity in the brain between motion processing areas and areas in the temporal lobe which decode form from motion. As such, while idiosyncratic motion processing may remain relatively intact in DP (Longmore & Tree, 2013; Steede et al., 2007b), it cannot be reliably concluded that this is evidence for a strict dissociation between static and dynamic processing, as found in the current study.

Interestingly, the results from the expression matching task reported above demonstrated a preserved ability for both prosopagnosics' to use non-rigid dynamic information for the purpose of matching expressions both within and across facial identities. Indeed, performance was similar across both static and dynamic learning conditions. This finding is consistent with that of Thornton & Kourtzi (2002) who demonstrated that non-rigid motion conferred an advantage in face matching performance in neurotypical individuals when the task was to match identity, rather than to match expression. This finding, of intact non-rigid motion processing for expression analysis, may help reconcile recent evidence of preserved functional activation in STS to moving face images, following delivery of TBS to visual form area OFA, reported by Pitcher et al. (2014). Indeed, many have argued that STS is primarily concerned with the processing of social signals in the face, e.g. expression or gaze direction (Haxby, Hoffman, & Gobbini, 2002; Hoffman & Haxby, 2000; Pitcher, 2014; Winston, Henson, Fine-Goulden, & Dolan, 2004). The current results suggest that it is possible that STS may be able to represent the changeable aspects of the face which support social signal processing, possibly independent of visual form areas. However, it is likely that interactions between STS and OFA may explain the

ability to represent facial identity information which changes over time, a task at which both prosopagnosics in the current study were impaired on when the face was learned moving in a non-rigid manner. This interpretation is also consistent with recent findings reported by Fox, Hanif, Iaria, Duchaine, & Barton (2011) for patient RST1, who presented with a lesion to pSTS but with preserved functioning in visual form areas, OFA and FFA. Behavioural testing indicated that the patient could match facial identity but not expression, but the former was only true when facial expression was held constant. In marked contrast, RST1 showed poor identity matching for images of the same face which varied by expression, leading the authors to conclude that STS likely contributes in some way to identity processing across changes in expression.

Moreover, STS has also been shown to be sensitive to the fluidity of non-rigid motion. In contrast, visual form areas are sensitive to the quantity of static information in a moving display, rather than the spatiotemporal dynamics of the motion sequence (Schultz et al., 2013). The results from the younger adults reported in Chapter 3 (Experiment 1B) showing faster reaction times to matching novel viewpoints of the face when the face had been learned through a fluid, rather than a jumbled, sequence of images argue that face information reflecting changes over time is relevant for identity processing. Thus it is likely that STS serves some functional role in representing the changeable aspects of the face within the same identity. Thus it remains to be seen whether the preserved functional activation in STS reported by Pitcher et al. (2014) may relate to the changeable aspects of the face per se, or whether the area may still be sensitive to changes both within and across different facial identities, following TBS to visual form areas.

However, it is also important to note that the current expression matching task was arguably easier than the facial identity matching task. Although, both UM and PL could match the expression across identities, it is possible that they may have engaged a feature matching strategy to support this performance as the distractor images were arguably perceptually distinct from the matching expression (happy versus angry). Indeed, Calder & Young (2005) argued that the apparent dissociation between identity and expression processing in prosopagnosia may arise due to the fact that both tasks are not equated for task difficulty. It may be possible to overcome this by using expressions which are more perceptually matched as distractor images e.g. by parametrically varying the similarity between the target and distractor expression along a critical dimension (e.g. morphing the distractor expression with the target in degrees of e.g. 10-100%). However, other findings have demonstrated that developmental prosopagnosics can reliably discriminate between morphed expressions (Duchaine, Parker, & Nakayama, 2003). Moreover, in terms of the current study, it is important to consider that even if a feature matching strategy had been engaged, non-rigid motion did not detract from encoding these features.

Taken together, the results from the series of experiments reported here demonstrate that dynamic cues, particularly non-rigid motion cues, support the representation of the facial identity in normal face processing. Furthermore, they point towards the suggestion that poor encoding of facial form information in DP may be related to the relative inability to integrate both form and motion cues. Further studies examining whether these deficits extend to individuals with associative, rather than apperceptive, prosopagnosia may help unravel the heterogeneous nature of this disorder (e.g. Fox, Iaria, & Barton, 2008; Grüter, Grüter, & Carbon, 2008; Le Grand et al., 2006).

Chapter 5

Interactions between the Face and Voice in Person
Perception

5 Interactions between the Face and Voice in Person Perception

5.1. Abstract

In parallel to face recognition, information from the voice can be used to support identity processing. For example, results from younger neurotypical adults' demonstrate that familiar speaker recognition is enhanced when the voice is accompanied by its corresponding facial identity and that these interactions occur at the early stages of information processing in the brain. However, to date, ageing research has largely focused on how dynamic visual information can enhance auditory speech, rather than speaker perception per se. To address this issue, the first part of Chapter 5 reports a series of experiments which examined how the face and voice may interact to support identity processing in older age. Here, younger and older participants learned a voice from either an auditory only (Experiment 7) or bimodal stimulus comprising of a voice accompanied by either a static image or a synchronised video of the corresponding facial identity (Experiment 8). Following a short delay, participants were presented with a two-interval forced choice (2IFC) in which they had to match the test voice to the learned identity. Matching vocal stimuli contained either same or different speech content. Overall the results from both experiments demonstrated that speaker matching performance was poorer in older than in younger adults. However, the cost associated with matching speaker identities across changes in speech content was similar for both age groups (Experiment 7). When the voice was learned as a bimodal stimulus older adults were faster at matching the voice across different speech content when the voice was presented with a dynamic, rather than a static, face during learning. No effect of the face (static or dynamic) was evident on younger adults' speaker matching performance, even when the task was made more difficult by extending the interval between the learned and test stimulus (Experiment 8). However, when task demands were increased by intermixing trials in which the face or the voice had to be matched across different speech content or a different facial pose (Experiment 9) there was evidence for better speaker matching performance when the voice had been learned in the presence of its dynamic, rather than static, face counterpart. These findings suggest that dynamic facial cues can affect speaker matching performance, but that these effects may be more readily observed when voice representations are weaker, such as occurs with older age, or when the task demands for younger adults are increased.

Although these findings suggest that identity information is shared across the senses, very little is known about how developmental prosopagnosia, a disorder characterised by poor face processing, may affect the processing of identity from the voice. As such the second part of

Chapter 5 turned to address speaker matching performance in both UM and PL in comparison to a group of age-matched controls (Experiment 10). Here, as in Experiment 7, participants listened to a voice uttering a sentence and following a short delay matched the learned identity to a test voice in a 2IFC design, in which matching vocal stimuli again contained either same or different speech content. In comparison to controls UM demonstrated very poor speaker matching accuracy, for both same and different speech content trials. However, PL's overall matching performance was similar to controls. It is argued that these results support the idea that the face and voice interact to support identity processing and that poor processing in one modality, i.e. the face, may affect processing in another, i.e. the voice. However, follow up tests would be needed to fully characterise how this deficit may emerge. Taken together, the results from Chapter 5 suggest that normal identity processing is likely achieved through a sharing of information across the senses and that, in turn, atypical face processing may affect how speaker identity is represented.

5.2. Introduction

Like the face, the voice conveys information which can support person recognition (Belin et al., 2004). Indeed, the perceptual system is tasked with a similar problem in decoding identity in either the voice or the face. Like faces, voices share the same invariant perceptual features across individuals (acoustic parameters such as pitch and timbre), and thus the perceptual system must represent a unique ‘vocal signature’ for each human voice in memory, by perceiving subtle differences in these features across individuals (Belin, Bestelmeyer, Latinus, & Watson, 2011; Belin et al., 2004). Furthermore, as identity must be recognised across changes in the visual appearance of the face, so too must the voice be recognised across changes in speech utterances. However, although both the face and the voice appear to be represented in memory in a similar manner (Belin et al., 2004; Campanella & Belin, 2007), a number of studies have demonstrated that our ability to recognise voices is not as efficient as our ability to recognise faces (Hanley, Smith, & Had, 1998; Hanley & Turner, 2000; Joassin et al., 2011; Stevenage et al., 2012; Stevenage, Hale, Morgan, & Neil, 2014; Stevenage, Howland, & Tippelt, 2011). In particular, unfamiliar voice recognition (i.e. voices heard once or twice) appears to be particularly error prone (e.g. Papcun, Kreiman, & Davis, 1989; Read & Craik, 1995).

However, in the real world it is important to consider that voices are rarely heard in isolation. Rather when we consider a typical communication scenario we are exposed to both the voice *and* its corresponding facial identity. Moreover, there is a close spatial and temporal relationship between these identity cues, as the dynamic articulating face changes over time in direct correspondence with the speech signal (Schweinberger, Kloth, & Robertson, 2011; Schweinberger, Robertson, & Kaufmann, 2007; Schweinberger, Kawahara, Simpson, Skuk, & Zäske, 2014). As such, voices are likely to become familiar within the context of their corresponding facial identity. Indeed, in a study investigating familiar (i.e. personally familiar) voice recognition, von Kriegstein, Kleinschmidt, Sterzer, & Giraud (2005) reported that while both familiar and unfamiliar voices induced activation in temporal voice areas (TVA) in the superior temporal sulcus (STS), familiar voices also activated visual form areas typically associated with representing facial identity (i.e. fusiform face area or FFA), suggesting the existence of established *multisensory* representations for each familiar identity. In a follow up study, von Kriegstein and Giraud (2006) extended these findings to demonstrate that these multisensory associations can be rapidly acquired (see also von Kriegstein et al., 2008). In that study, participants learned an unfamiliar speaker with either its corresponding facial identity or a name. They observed increased accuracy for recognizing voices that had been originally paired in the learning phase with their associative faces. Moreover, these multisensory associations were also evident at a

neural level, with increased functional connectivity between voice and face areas for this learning condition. Other studies have replicated and extended von Kriegstein & Giraud's findings (Joassin, Pesenti, et al., 2011; O'Mahony & Newell, 2012; Sheffert & Olson, 2004). For example, O'Mahony and Newell (2012) observed that participants who were trained on face and voice pairings were faster in detecting the familiarity of a person when they were presented with congruent, rather than incongruent, face and voice information at testing stage. Intriguingly, in line with von Kriegstein and colleagues findings, they did not observe the same recognition enhancement for congruent face and name pairings, suggesting that it is the naturalistic coupling of face and voice information which enhances person recognition and not the pairing of the face with additional arbitrary information i.e. the name of the person. Indeed, further studies revealed direct structural connections between FFA and TVA (Blank et al., 2011), suggesting that it is the sharing of identity information across sensory cortices (Ghazanfar & Schroeder, 2006), rather than top-down modulation (e.g. face 'imagery' being elicited in response to the voice) which likely mediates these behavioural effects.

Importantly, many of these behavioural effects have resulted from learning paradigms in which the voice was learned with its dynamic face counterpart. Indeed, the spatiotemporal relationship between these two cues appears to affect how both the face and voice are integrated to support recognition. This is highlighted by findings from Schweinberger, Robertson, and Kaufmann (2007) who presented individuals with dynamic face and voice pairings, i.e. preserving the cross-modal spatial and temporal alignment of the faces and voices, while manipulating the congruency of identity in these face and voice pairings (such that a voice could be presented with a non-corresponding facial identity but the face would appear to be articulating with the voice in synchrony). They noted behavioural benefits and costs for identifying familiar voices, when they were respectively accompanied with either a congruent or incongruent facial identity. Yet importantly, this was particularly true for voices accompanied with time synchronised dynamic facial information, static facial information either reduced or removed the effects of cross-modal modulation. This suggests that the face was not simply acting as an additional 'cue' to identity rather it is likely that the spatial and temporal alignment of the face and voice is crucial for audio-visual integration in person recognition (see also Robertson & Schweinberger, 2010).

These results have important implications for traditional models of person recognition which have argued that face and voice information is processed independently and combined only at a post-perceptual stage in person-identity nodes (PIN), upon which access to semantic information pertaining to the identity becomes available (Bruce & Young, 1986; Burton, Bruce, & Johnston, 1990). However, the evidence supporting audio-visual interactions between dynamic

face and voice pairings, reported by Schweinberger and colleagues, argue that these interactions likely arise at a perceptual stage during the encoding of identity (Robertson & Schweinberger, 2010; Schweinberger et al., 2011). This argument is supported by evidence suggesting that observers cannot suppress accompanying incongruent visual information in favour of available voice information (mandatory interactions) and that congruent face and voice pairing enhance perception. While Schweinberger et al. (2007) argued that these effects were likely modulated by previously established, i.e. familiar face and voice representations, it is important to note that they also observed a slightly similar, albeit weaker, trend for a dynamic face benefit on *unfamiliar* voice detection. They noted faster reaction times when judging whether a voice was unfamiliar or not when the voice was presented with a dynamic rather than static unfamiliar face. Although these effects were weaker than that observed for familiar voice recognition they do suggest that there may be some benefit on voice recognition from dynamic face information even for representations which are yet to be established.

Indeed, Kamachi, Hill, Lander, & Vatikiotis-Bateson (2003) demonstrated that observers showed above chance performance in matching silent moving unfamiliar faces to their corresponding speaker identities and vice versa. While this matching ability held across changes in speech content, matching performance was impaired for dynamic faces in which the fluidity of the motion was disrupted. Kamachi and colleagues also noted that these cross-modal matching effects were reduced (performance fell below chance) when the face was presented as a static image (see also Lachs & Pisoni, 2004). However, recently Mavica & Barenholtz (2013) reported identity matching performance which was greater than chance, between unfamiliar face images (i.e. static photographs) and recorded voices. These results suggest that the perceptual system is sensitive to cross-modal identity information, but that these effects may be more consistently observed when the spatiotemporal cues in the face and voice remain intact. Moreover, it appears possible to match the face and voice for identity when motion cues are presented in the relative absence of form information i.e. as point light displays (Rosenblum, Smith, Nichols, Hale, & Lee, 2006), rather than when both form and motion cues are available (e.g. Kamachi et al., 2003; Lander, Hill, Kamachi, & Vatikiotis-Bateson, 2007). This sensitivity likely arises due to the physical constraints linking faces to voices (Kamachi et al., 2003). For example, Lander et al. (2007) demonstrated that changing the manner in which a sentence was said, by altering either the expressiveness or rhythmic structure of the speech signal, significantly disrupted matching performance. Altering the manner between target and test presentation likely makes matching more difficult as it consequently alters the spatiotemporal relationship between the movement of the face and the corresponding voice.

5.2.1. *Interactions between the face and voice in ageing*

There is some evidence that speaker recognition (i.e. long term memory recall of voice identity) is poorer in older than in younger adults (Clifford, 1980; Yonan & Sommers, 2000). For example, Yonan & Sommers (2000) observed that compared to younger adults, older adults were worse at identifying previously familiarised voices, across changes in speech content, after 1 or 2 day intervals between training and test. In that study, participants were trained on voice-name associations and asked to link the voice to the name during testing. Indeed, as discussed earlier, learning the face with its corresponding name is likely to be a difficult task, as it does not exploit cross-modal redundancies between the face and voice signals (von Kriegstein & Giraud, 2006). However, our understanding of how the face and voice may interact to support person perception in older adults is poor. As mentioned in the opening chapter, much of the literature on ageing has focussed on how dynamic face cues can support *speech*, rather than *speaker*, perception (Gordon & Allen, 2009; Maguinness et al., 2011; Setti, Burke, Kenny, & Newell, 2013; Sommers et al., 2005; Tye-Murray, Spehar, Myerson, Sommers, & Hale, 2011).

The findings from these studies with older adults have largely supported the classic finding reported by Sumby & Pollack (1954) that dynamic facial cues can enhance speech perception particularly when the auditory speech signal is degraded (Gordon & Allen, 2009; Sommers et al., 2005; Tye-Murray et al., 2011). However, unlike younger adults the benefit for older adults is largely limited to non-degraded, i.e. super threshold, visual cues from the face (e.g. Gordon & Allen, 2009). Nevertheless, in parallel to the findings of Kamachi et al. (2003) these findings suggest that the perceptual system is sensitive to, and can exploit, redundant cross-modal spatiotemporal cues, at least to support speech perception, in older age. Moreover, others have argued that these face and voice interactions may become more pronounced with increasing age (Maguinness et al., 2011; Setti et al., 2013). For example, Maguinness et al. (2011) demonstrated that compared to younger adults, older adults' immediate speech recall for semantically unpredictable, relative to predictable, sentences was enhanced when both the voice and the moving face of the actor was available during the encoding of the speech signal. Furthermore, Setti et al., (2013) reported that older adults were more susceptible to the cross-modal illusion known as the "Mc-Gurk" or "Mc-Gurk-Mac Donald" effect (McGurk & MacDonald, 1976), in that they tended to integrate incongruent auditory and visual signals and report significantly more illusory percepts, than their younger adult counterparts. Taken together, these results suggest that older adults may rely more on the availability of the visual dynamic face when encoding the auditory signal.

Johnsrude et al. (2013) also demonstrated that the familiarity of a speaker can enhance speech perception in older adults (Pilotti, Beyer, & Yasunami, 2001; Yonan & Sommers, 2000). Specifically, they observed that older adults were better at tracking a target voice, in the presence of a secondary talker, when the target was that of a familiar (voice of a spouse) relative to unfamiliar speaker (see Schall & von Kriegstein, 2014 for neurological underpinnings of such effects). However, others have observed that this benefit of voice familiarity on speech recognition may be reduced in older, compared to younger adults. Pilotti et al., (2003) suggested that “poorer sensory uptake”, i.e. poorer processing of the auditory signal, may mean that perceptual features pertaining to the identity of the voice may be discarded in favour of the linguistic content when the task is to focus on encoding the content of the speech.

Taken together, the results from the voice recognition and speech perception literatures demonstrate that older adults are sensitive to identity cues within the speech signal, however these effects may be modulated by the familiarity of the speaker (e.g. Johnsrude et al., 2013), such that recognition is poorer for relatively unfamiliar voices (such as those learned in a training stage, rather than representations of personally familiar individuals which were previously established in memory), particularly when matching across changes in speech content (Yonan & Sommers, 2000). This cost in unfamiliar voice recognition is consistent with findings from the face perception literature of poorer matching and recognition of identity of unfamiliar faces, particularly across changes in visual appearance of the face (see Chapter 2 and 3). However, given that the perceptual system can utilise redundant spatiotemporal cues across the face and the voice to enhance the perception of the speech signal, it is possible that these cues may also support the perception of the speaker in older adults. Indeed, the findings from Chapter 3 suggest that older adults can exploit dynamic facial cues to enhance the representation of facial identity when form cues are not sufficient. Furthermore, as the representations of voices may be relatively weaker in this cohort than in younger adults it is possible that the role of dynamic visual cues on unfamiliar voice perception may be more readily observed in older, relative to younger adults performance (Schweinberger et al., 2007).

5.3. Outline of the Experiments

As such, the first part of this Chapter examined the role of the face during the learning of speaker identities in younger and older adults. Specifically, it investigated if speaker *matching* (i.e. matching voices on the basis of their identity), in the absence of long term memory demands, was affected by the ageing process (Experiment 7) and if the presence of a dynamic, rather than a static, face identity may enhance the representation of the voice in older and younger adults

(Experiment 8). It was also investigated if learning the voice with its corresponding synchronised facial identity would assist in matching both the voice and the face across changes in either speech content or the face image in the younger adult group only (Experiment 9).

5.4. EXPERIMENT 7

In Experiment 7 both younger and older adults were required to match a series of unfamiliar voices (auditory only) on the basis of their identity. In each trial, participants first learned the target voice and were then required to match the learned identity by choosing from two sequentially presented test voices. The test voices contained either the same or different sentence content to what had been learned. In accordance with Chapter 3 and 4, an immediate matching paradigm was employed and as such this task likely assesses speaker perception rather than recognition. However, the interval was longer in duration between the learning and test phase (i.e. 1,500ms) than that reported in Chapter 3 (500ms). This interval was chosen as it has been previously used in voice discrimination studies (e.g. Kreiman & Papcun, 1991). Furthermore, the suitability of this interval was confirmed through older adult participant feedback, as it was expressed that shorter durations made it difficult to differentiate between the learning and the test phase of the experiment. It was expected that, consistent with previous findings reported in the voice recognition literature, voice matching would be poorer in the older than the younger adult group. Furthermore, this age effect may be more evident when the task is to match the voice across changes in speech content.

5.4.1. Method

Participants

Sixteen young adults, recruited from the student population at Trinity College Dublin (10 female, 15 right handed, mean age= 22.5 years, SD= 4.3 years) participated in this experiment and were given course credits in return for their time. All younger adults reported normal or corrected to normal vision. Eighteen, community-dwelling, older adults were also recruited for this task.

All participants were naive to the purpose of the experiment. All younger and older adults were asked to subjectively rate their hearing by answering the question: "Is your hearing (with or without a hearing aid) Excellent/Very Good/Good/Fair/Poor?" This question was taken from the Irish Longitudinal Study on Ageing Questionnaire (Kenny et al., 2010) and has been shown to be a valid indicator of objective hearing measures in older age, demonstrating a 94.67% specificity rate

(Kenny Gibson, Cronin, Kenny, & Setti, 2014). Kenny Gibson et al. (2014) reported the 94.67% specificity rate based on the observation of $n=150$ older adults. The specificity rate refers to those who identified that they did not have a hearing impairment, rating their hearing as either 'Excellent', 'Very Good', 'Good', and then passed an objective hearing test ($n=142/150$). All younger adults rated their hearing as either 'Excellent' or 'Very Good' and all older adults reported 'Excellent' or 'Very Good' hearing.

Data collected from two of the older adults was subsequently removed from analysis as they did not understand the task instructions and as such did not complete the task correctly. Data from the other 16 older adult participants were included in subsequent analyses (11 female, all right handed, mean age = 70.7 years; $SD = 4.9$ years). All older adults scored within the normal range (i.e. ≥ 24 of 30) on the MoCA (Nasreddine et al., 2005), a reliable measure of global cognitive function. All older participants had normal visual acuity (mean binocular LogMar score = 0.06, $SD = 0.1$ and LogMar score = 0.02, $SD = 0.1$ at 4m and 40 cm respectively) and contrast sensitivity (mean score = 1.99, $SD = 0.1$, assessed using the Pelli-Robson Contrast Sensitivity Test) for their age. All participants wore appropriate corrective lenses, if needed, at the time of testing.

Stimuli and Apparatus

The stimuli consisted of audio recordings of 10 male and 10 female native English speakers articulating a set of three sentences: "People read books", "This is my house", and "The window is closed". All were variations of sentences previously rated as semantically neutral in content i.e. they had no unique emotion ascribed to their wording (see Adolphs, Damasio, & Tranel, 2002; Russ, Gur, & Bilker, 2008), thus controlling for arousal induced by the semantics in the speech signal. The recordings were made in a quiet room using a Sony ICDSX68 digital voice recorder. During the recording session speakers were asked to memorize each sentence and then to utter the sentence aloud in a neutral tone and at a natural speech rate. Sentences were subsequently edited using Adobe Audition® and were roughly equated for volume and sentence length. The sentence "People read books" was approximately 1.5 seconds in duration, while sentences "This is my house" and "The window is closed" were both approximately 1.8 seconds in duration. The experiment was programmed in and run using Presentation® software and was presented on a Dell OptiPlex 755 PC and displayed on a HP L1710 17" LCD colour monitor. Auditory stimuli were presented over Sennheiser HD 202 closed back on-ear stereo headphones at a super-threshold decibel level.

Design

The experiment was based on a 2-way, mixed, factorial design with Age (younger or older adults) as a between-subjects factor and Sentence content (same or different) as a within-subject factor. The experimental protocol was based on a delayed matched to sample, two-interval forced choice paradigm, where a target voice was presented, followed by two sequentially presented test voices. As such, this study adopted the same sequential matching protocol as previously described by Lander et al. (2007). The participant was required to learn the identity of the speaker in the first voice clip and to choose which of the two test voices matched the identity of the previous speaker. The dependent variables were response accuracy and response time taken to match the identity of the target speaker.

The experiment consisted of 40 trials in total. In half of the trials the speech content i.e. the sentence wording, was the same across the target and the test voices ("People read books"). In the other half of the trials the speech content was different across the target and test voices ("This is my house" or "The window is closed"). Each speaker appeared once as a target and distractor voice in the *same* sentence content trials and once as a target and distractor in the *different* sentence content trials. For different sentence content trials, half of the participants learned "This is my house" as the target sentence, while the test sentences consisted of "The window is closed"; while the other half of participants learned "The window is closed" as the target sentence and "This is my house" as test sentences. All 40 trials were presented in a single test block and the order of trials was randomized across participants. On each trial one of the two test voices always matched the identity of the target speaker and one was a distractor voice. The distractor voice always matched the sex of the target speaker. The order of the correct match test item was counterbalanced across trials, appearing directly after the target voice in only half the trials.

Procedure

Participants wore headphones for the duration of the experiment. In any one trial, they were first presented with a voice recording (target voice) and were instructed to learn the identity of this initial voice presentation. The target voice was followed by a 1,500ms inter stimulus interval (ISI) which displayed a blue fixation cross in the centre of the screen. The blue cross then disappeared and participants then heard two voice clips presented sequentially, i.e. one directly after the other. Participants were instructed to listen to each voice clip and decide whether the first or the second voice matched the identity of the target speaker previously heard. Participants

were alerted to the fact that the speech content may change, but that their task was to focus on the speaker identity. A response screen appeared immediately after the test voices were presented which instructed participants to press button ‘z’ on the keyboard if they thought Voice 1 matched the target identity or ‘m’ if Voice 2 matched the target speaker identity. Participants were instructed to always listen to both test options and to respond as quickly as possible when they saw the response screen. A response triggered the offset of the current trial and the onset of the next trial. Each trial was preceded by an inter-trial interval of 500ms. See Figure 5.1 for an illustration of a sample trial.

Four practice trials preceded the main experiment (2 same content and 2 different speech content trials). The voice identities used in the practice trials did not appear within the main experiment. Participants were given verbal feedback as to their accuracy in performance during the practice period and were allowed to repeat these 4 trials as many times as needed until they were comfortable with the testing procedure.

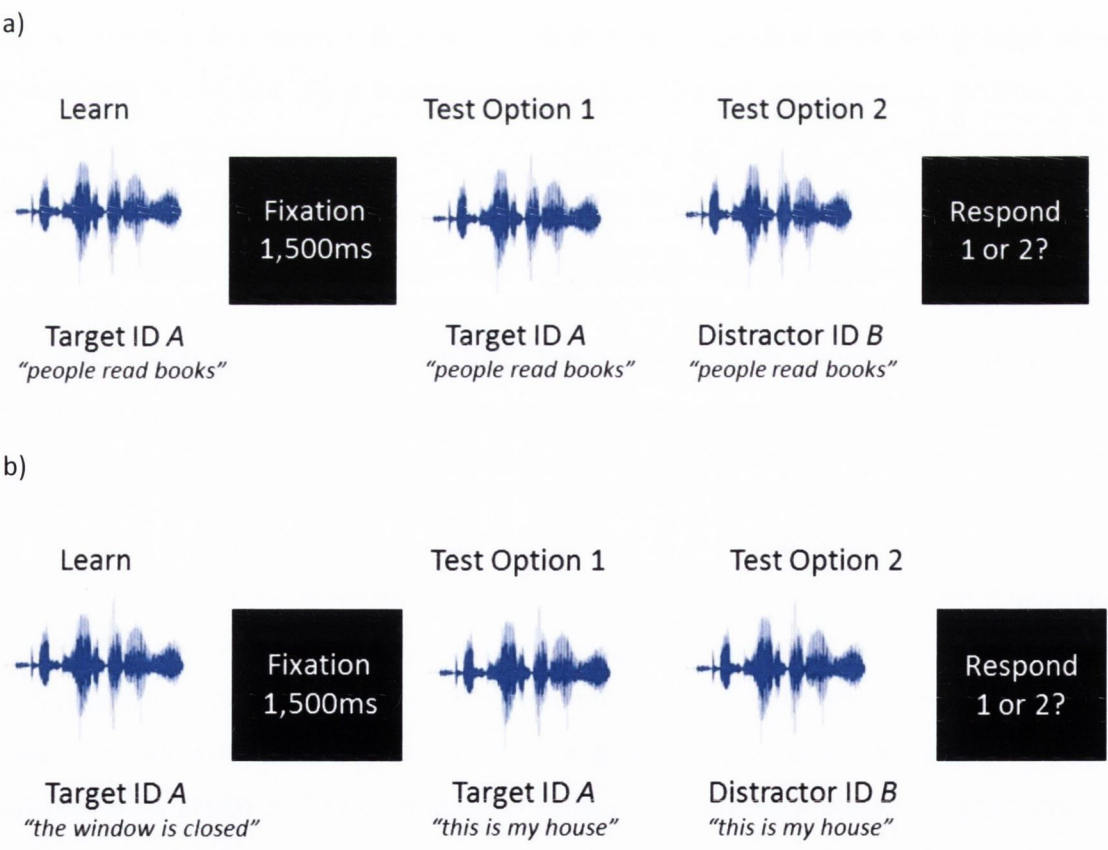


Figure 5.1. Illustration of different trials in Experiment 7. a) Depiction of a same sentence content trial (upper panel) and b) different sentence content trial (lower panel). Note that *both* the target learning and test conditions are auditory (voice) only.

Analysis

Similar to the face matching tasks reported in Chapter 3 and 4 there was a large range in speaker matching performance evident in the older adult group. However, it was noted that performance for same-sentence content trials was largely predictive of equivalent speaker matching performance for novel speech content trials ($r=0.75$, $p<0.001$). As such, the normalizing procedure described in Chapter 3 and 4 for matching novel relative to learned face images was not applied in this instance. Accuracy was therefore calculated as the percentage of voices correctly matched for speaker identity for each participant and for each sentence content condition (i.e. same or different sentence content trials). For the purpose of our analysis, reaction times (RTs in milliseconds) to correct matches only were included, for learned and novel content trials.

5.4.2. Results

Accuracy

Accuracy was calculated as the percentage of voices correctly matched for speaker identity for each participant and for each sentence content condition. The mean percentage of correct responses across participants was subjected to a 2×2 mixed design analysis of variance (ANOVA), with Age (younger or older) as a between-subject factor and Sentence content (same or different) as a within-subject factor. A main effect of Age was observed [$F(1,30)=21.29$, $p<0.001$, $\eta_p^2=0.42$], with poorer speaker matching accuracy in the older ($M=79.9\%$) than younger ($M=96.2\%$) adults. A main effect of Sentence content was also found [$F(1,30)=4.1$, $p=0.05$, $\eta_p^2=0.11$], with better accuracy for matching speaker identity across learned than novel sentence content, (see Figure 5.2). There was no evidence of an interaction between Age and Speech content [$F(1,30)<1$].

Reaction Times

RTs for correct responses only were included in analysis. Furthermore, for each of the participants RTs which were 2.5 standard deviations away from the mean, in either direction, in each sentence content condition were removed before analysis (approximately 10% of the data were removed). The mean RTs were then analysed using a 2×2 mixed design ANOVA, with Age (younger or older adults) and Sentence content (learned or novel) as a within subject factor. A main effect of Age was found [$F(1,30)=23.25$, $p<0.001$, $\eta_p^2=0.44$], with slower time taken to match speaker identity in the older ($M=666$ ms) relative to younger ($M=421$ ms) adult group. No main

effect of Sentence content [$F(1,30)<1$], or interaction between Speech content and Age [$F(1,30)<1$] was observed.

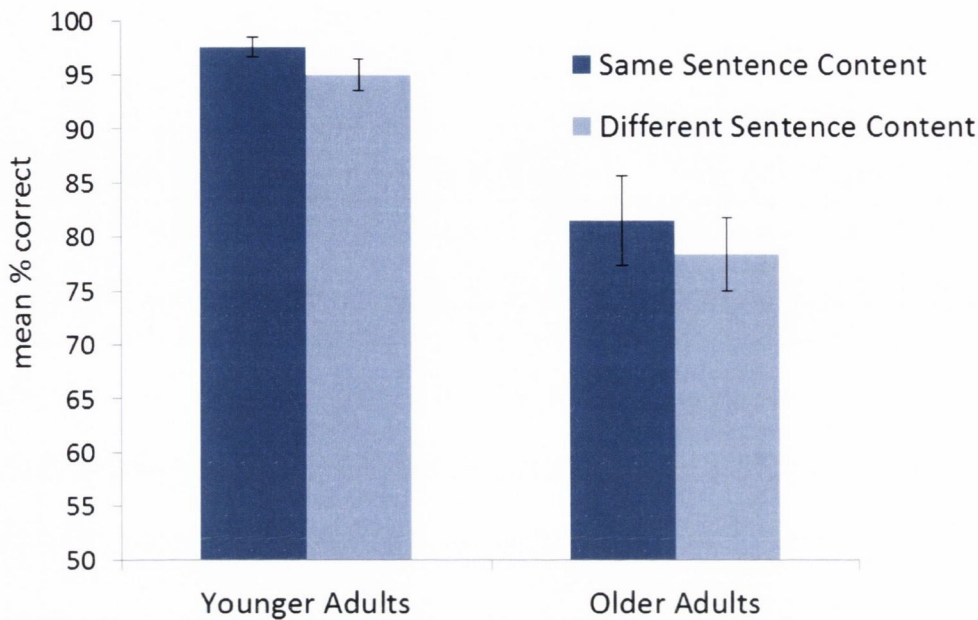


Figure 5.2. Plot showing the mean accuracy scores for matching speaker identity across the same or different speech content, for both younger and older adult groups tested in Experiment 7. Error bars show ± 1 SEM.

5.4.3. Discussion

Overall, older adults were poorer at matching speaker identities than younger adults. However, the cost in accuracy performance associated with matching the speaker identity across changes in sentence content was similar across both age groups, and there was no cost in response times for a change in sentence content. Thus, in line with previous research, the findings from Experiment 7 suggest that unfamiliar speaker matching, even within the same modality, was more error-prone when the content of the speech signal was changed. Matching across speech utterances requires that listeners extrapolate from the learned speech content the invariant, paralinguistic cues that are used to support identity matching, e.g. pitch, rhythm, independent of the speech content (Pilotti et al., 2003). However, it is possible that learning the voice within the context of its corresponding facial identity may improve speaker matching performance, particularly in the older adult group.

5.5. EXPERIMENT 8

The purpose of Experiment 8 was to examine if learning an unfamiliar voice in the presence of its corresponding static or dynamic facial identity may affect unfamiliar speaker matching performance in the older and younger adult group. Previous studies have demonstrated that younger adults can reliably match video clips of silent moving, and to a lesser extent static face images, to their corresponding vocal identities, even when the spoken sentence is different across the face and voice clips (Kamachi et al., 2003; Lander et al., 2007; Mavica & Barenholtz, 2013). This suggests that the perceptual system is sensitive to commonalities between facial patterns and vocal signatures. Furthermore, these spatiotemporal patterns appear to be reliably associated with the speaker, rather than the speech content. The cross-modal redundancies across the facial movement and vocal content may support matching performance even for identities which are unfamiliar.

As such, in the following experiment it was expected that the presence of a dynamic, rather than a static, face during learning may assist in speaker matching performance and that this effect may be more evident in the older adult group, owing to their poor auditory (voice) only matching performance in Experiment 7. Indeed, based on the findings from Experiment 7, it was predicted that younger adults would likely be able to represent the voice independently of the face under the current matching paradigm. As such, in an attempt to modulate task difficulty, the interval between learning and test was modulated in this group (see Design and Procedure section that follows).

5.5.1 Method

Participants

Sixteen younger adults, recruited from the student population at Trinity College Dublin (12 female, 15 right handed, mean age= 21.9 years, SD= 4.7 years) participated in this experiment and were given course credits in return for their time. All younger adults reported normal or corrected to normal vision. Eighteen, community-dwelling, older adults also participated in this task. Data collected from two of the older adults was subsequently removed from analysis for the following reasons: one was removed as the participant did not understand the task instructions and as such did not complete the task correctly, while the other was removed as it was subsequently reported that the participants had mild cataracts at the time of testing. Data from

the other 16 older adult participants were included in subsequent analyses (11 female, all right handed, mean age= 69.8 years; SD= 6.4 years).

All younger and older adults subjectively rated their hearing by answering the question: "Is your hearing (with or without a hearing aid) Excellent/Very Good/Good/Fair/Poor?" All younger and older adults rated their hearing as either 'Excellent' or 'Very Good'. All older adults scored within the normal range (i.e. ≥ 24 of 30) on the MoCA (Nasreddine et al., 2005), a reliable measure of global cognitive function. All older participants had normal visual acuity (mean binocular LogMar score= 0.1, SD= 0.1 and LogMar score= 0.00, SD= 0.1 at 4m and 40 cm respectively) and contrast sensitivity (mean score= 1.95, SD= 0.0, assessed using the Pelli-Robson Contrast Sensitivity Test) for their age. All participants wore appropriate corrective lenses, if needed, at the time of testing. None of the younger adults, but four of the older adult participants had participated in Experiment 7. However, note there is no overlap in the experimental stimuli, as all actors used in Experiment 8 are novel compared to those reported in Experiment 7.

Stimuli and Apparatus

The stimuli consisted of audio-visual recordings of 10 male and 10 female native English speakers articulating a set of three sentences: "People read books", "This is my house", and "The window is closed". All actors differed from Experiment 7, however all details of sentences used are as previously described. The video recordings were made in a quiet room using Canon 550d camera, recording at a rate of 29.97 frames per second, and a Rode directional microphone. The recording volume was adjusted using Audition software for each actor to maintain consistency of loudness across all recordings samples. During the recording session actors were asked to memorize each sentence and to then look directly at the camera, which was positioned at eye-level in front of them. They were instructed to keep their head facing forwards at all times and to try and avoid any rigid head movements (i.e. nodding or head rotations). When they felt comfortable, each actor uttered each sentence aloud four times in a neutral tone and at a natural speech rate. Two samples of each sentence "People read books" and one sample of each sentence "This is my house" and "The window is closed" were recorded from each actor and were subsequently edited for use in the main experiment. The audio and visual samples were chosen as they were devoid of artefacts and the speech rate was relatively consistent across all sentence types.

The final set of videos were edited using Adobe Premiere® to include a black mask which occluded any non-facial features which were visible during the recording session, such as hair. The

frame length of each masked video was edited using VirtualDub software (www.virtualdub.org). The sentence clip “People read books” was edited to be on average 54 frames in length, while sentence clips “This is my house” and “The window is closed” were approximately 57 frames in length. Using the same software, the audio stream for each video file was saved separately as a ‘wav’ file and a version of each video, where the audio signal was removed, was also created. Thus, an audio-visual, audio-only and visual-only version (this visual-only version was used in Experiment 9) of each sentence clip was created. In addition, static images were also extracted from each video sequence. For the sentence clip “People read books” one static image which displayed the viseme /p/ was extracted (on average frame 5). For sentences “This is my house” and “The window is closed” frames which displayed the viseme /h/ of ‘house’ (on average frame 28) and /w/ of ‘window’ (on average frame 12) were extracted. These frames were chosen on the basis of previous studies which suggested that the point of articulation for visemes /h/ and /w/ renders both images visually dissimilar, but also relevant to the content (see Cappelletta & Harte, 2011, 2012; Potamianos et al., 2003). The experiment was programmed in and run using Presentation® software and was presented on a Dell OptiPlex 755 PC and displayed on a HP L1710 17” LCD colour monitor. Auditory stimuli were presented over Sennheiser HD 202 closed back on-ear stereo headphones at a super-threshold decibel level.

Design

The experiment was based on a factorial design with Sentence content (same or different) and Face (dynamic or static) as repeated factors for the older adult group. For the younger adult group the experiment was based on a mixed factorial design, with both Sentence content (same or different) and Face (dynamic or static) as within subject factors and Interval duration (short or long) as a between subjects factor (see Procedure section for further details). The experimental protocol was based on a delayed matched to sample, two interval forced choice paradigm, where a target voice was presented with its corresponding facial identity (static or dynamic) during the learning stage of the trial and this was followed by two sequentially presented test voices (i.e. without faces). The participant was required to learn the identity of the speaker, but to also look at the facial identity and to then choose which of the two test voices matched the identity of the previous speaker. The dependent variables were response accuracy and response time taken to match the identity of the target speaker.

The experiment consisted of 80 trials in total. In half of the trials the voice was presented during the learning stage with its corresponding dynamic facial identity which moved in synchrony with the speech i.e. as a video. While in the other half of the trials the voice was presented with a

static image of its corresponding facial identity. The static image was always extracted from the same sentence clip and was displayed for the same duration as the voice clip. Furthermore, in half of the *dynamic* and half of the *static* face-voice pairings the sentence was the same across the target presentation and the test voice (e.g. “People read books”). However, although the content was the same the test voice was a different recording of the same sentence clip. This helped to ensure that speaker identity matching was being performed based on the perceptual properties of the speech clip, and not based on direct auditory feature matching to the original recording (similar to image matching in face perception). In turn, in the other half of the trials the content of the sentence was different between the target and test voices (“This is my house” or “The window is closed”). Each voice was used twice as a target voice and twice as a distractor voice in the *dynamic* face condition and twice as a target voice and twice as a distractor voice in the *static* face condition. For different sentence content trials, half of the participants learned “This is my house” as the target sentence, while the test sentences consisted of “The window is closed”; while the other half of participants always learned “The window is closed” as the target sentence and “This is my house” as test sentences. All 80 trials were presented in a single test block and the order of trials was randomized across participants. On each trial one of the two test voices always matched the identity of the target speaker and one was a distractor voice. The distractor voice was always matched to the sex of the target speaker. See Figure 5.3 for an illustration of a sample trial. The order of the correct match test item was counterbalanced across trials, which meant it appeared directly after the presentation of the target voice in only half the trials.

Procedure

Participants were presented with a bimodal face-voice identity and were instructed that they should learn the target voice but also look at the face during learning. Following presentation of the face-voice pairing, a blue fixation cross appeared in the centre of the screen. For the older adult group the fixation cross was always displayed for 1,500 milliseconds (similar to Experiment 7). However, for the younger adult group the fixation cross was displayed for 1,500 milliseconds (short interval) for half of participants, while the other half of participants were shown the fixation for 5,000 milliseconds (long interval). This was done in an attempt to manipulate task difficulty in this group, by extending the interval between learning and test items. All participants were instructed that after the blue cross disappeared they would hear two voice clips presented sequentially, in the absence of a face, and that they should listen to each voice clip and decide whether the first or the second voice matched the identity of the voice previously learned. A response screen appeared immediately after the two test voices were presented which instructed

participants to press button ‘z’ on the keyboard if they thought Voice 1 matched the target identity or ‘m’ if Voice 2 matched the target speaker identity. Participants were instructed to always listen to both test options and to respond as quickly as possible when they saw the response screen. A response triggered the offset of the current trial and the onset of the next trial. Each trial was preceded by an inter-trial interval of 500ms. See Figure 5.3 for an example of trial. As Experiment 7, four practice trials preceded the main experiment: 2 same content (static or dynamic face) and 2 different (static or dynamic face) speech content trials. The face and voice identities used in the practice trials were unfamiliar to participants and were not part of the stimulus set used in the main experiment. In line with Experiment 7, participants were given verbal feedback as to their accuracy in performance during the practice trials. Furthermore, all participants were given the opportunity to repeat these 4 trials as many times often as was required for them to become comfortable with the testing procedure.

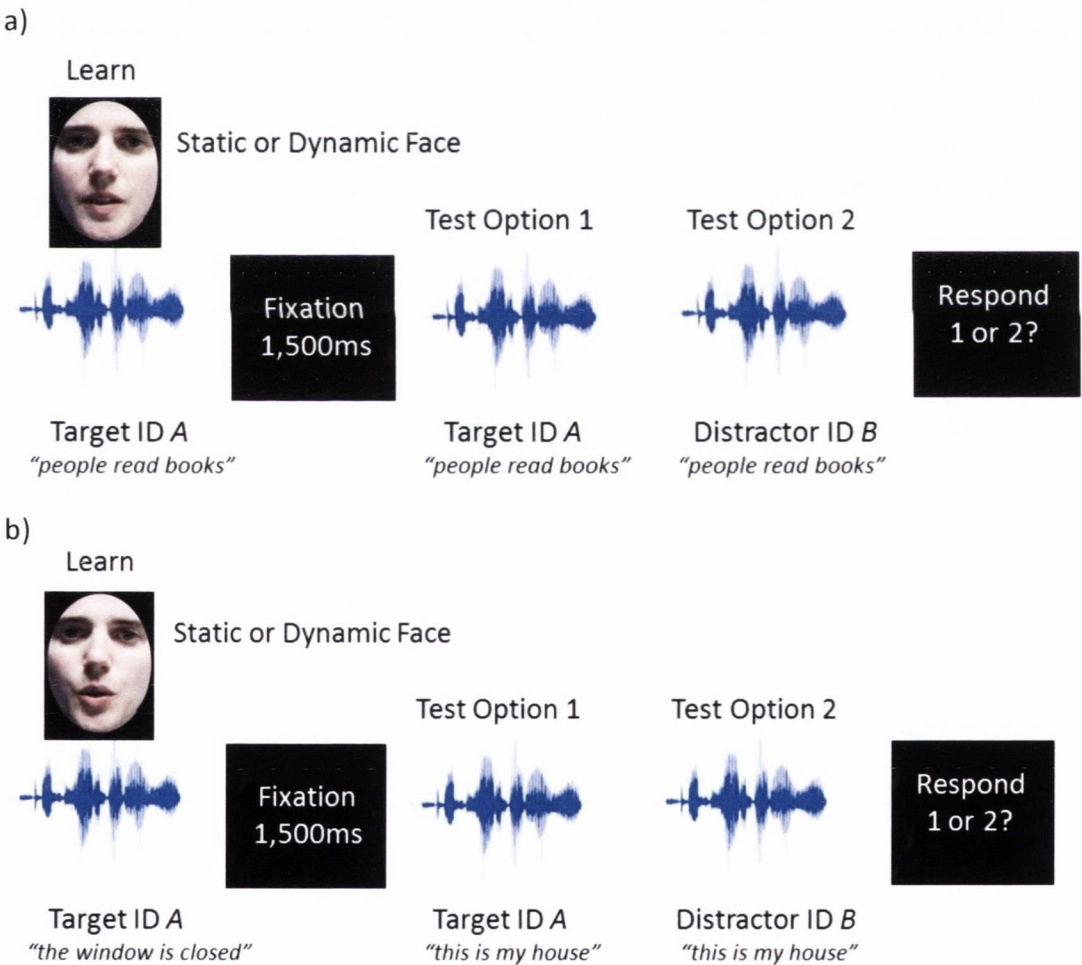


Figure 5.3. An illustration of a sample audio-visual speaker learning trial, where the voice is paired with either its static or dynamic face counterpart during learning. Note that both test options are auditory only. The upper panel (a) represents a same sentence content trial, while the lower panel (b) represents a different sentence content trial.

5.5.2. Results

Due to the manipulation of the timing interval between the learning and test, results were analysed separately for the older and younger adult group.

Accuracy Older Adults

Accuracy was calculated as the percentage of voices correctly matched for speaker identity for each participant for each condition. The mean percentage of correct responses across participants was subjected to a 2x2 repeated measures analysis of variance (ANOVA), with Face (static or dynamic) and Sentence content (same or different) as repeated factors. No main effect of Face [$F(1,15)<1$] or Sentence content [$F(1,15)<1$] was observed. There was no evidence of an interaction between Face and Sentence content [$F(1,15)<1$]. Mean accuracy performance for matching speaker identity in the presence of a static or dynamic face for same (different) content trials was 83.4% (81.9%) and 84.1% (80.6%), respectively.

Reaction Times Older Adults

For the purpose of this analysis, RTs to correct matches only were included. Moreover, RTs which were greater or less than 2.5 deviations from the mean RT for each condition were not included in the analyses. This resulted in an exclusion of approximately 8 % of the data. RTs were analysed using a 2 x 2 repeated measures ANOVA, with Face (static or dynamic) and Sentence content (same or different) as repeated factors. A main effect of Face was found [$F(1,15)=4.33$ $p=0.05$, $\eta_p^2=0.49$], with faster response times to matching speaker identity when the voice was learned with its dynamic ($M=639$ ms) rather than static ($M=698$) facial counterpart. A main effect of Sentence content was also observed [$F(1,15)=5.69$, $p=0.03$, $\eta_p^2=0.22$], with slower response times to match speaker identity across different ($M=692$ ms) relative to the same Sentence content ($M= 644$ ms). There was also a significant interaction between Face and Sentence content [$F(1,15)=11.95$, $p=0.004$, $\eta_p^2=0.44$]. A *post hoc* Tukey analysis revealed that when the same Sentence content was presented at test there was no difference in reaction time performance for matching speaker identity when the voice was presented with a static ($M= 637$ ms) or dynamic ($M= 652$ ms) face during learning. However, when the task was to match the voice identity across changes in sentence content, then performance was slowest when the voice was learned with a static face pairing ($M= 759$ ms) than when the same sentence content was spoken. This cost in reaction time performance for matching speaker identity across novel or same sentence content was not evident for dynamic face and voice pairings ($M= 626$ ms), see Figure 5.4.

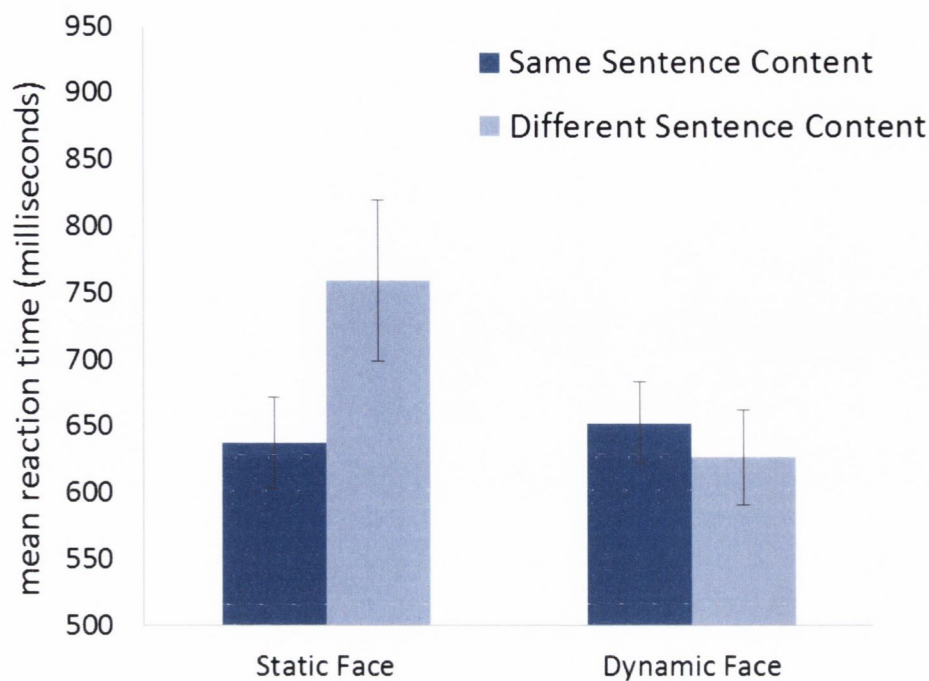


Figure 5.4. Plot showing the mean reaction times for older adults to match speaker identity across the same, relative to different, sentence content trials for static (left of figure) or dynamic (right of figure) face-voice pairings in Experiment 8. Error bars show ± 1 SEM.

Accuracy Younger Adults

Accuracy was calculated as the percentage of voices correctly matched for speaker identity for each participant for each condition. The mean percentage of correct responses across participants was subjected to a 3x2 mixed design analysis of variance (ANOVA), with Face (static or dynamic) Sentence content (same or different) as a within-subject factors and Interval duration (short or long) as a between-subject factor. There was no main effects of Interval duration [$F(1,14)<1$] or Sentence content [$F(1,14)=1.13, p=0.31$]. However, an interaction between Interval duration and Sentence content [$F(1,14)=10.16, p=0.007, \eta_p^2=0.42$] was found, with lower accuracy for matching the speaker across changes in Sentence content ($M= 94.7\%$) for the shorter, rather than the longer ($M= 97.8\%$) interval between learning and test conditions (see Figure 5.5). No main effect of Face [$F(1,14)<1$] or an interaction between Face and Interval duration [$F(1,14)<1$] were found. Furthermore, there was no interaction between Face and Sentence content [$F(1,14)<1$] or between Face, Sentence content and Interval [$F(1,14) <1$].

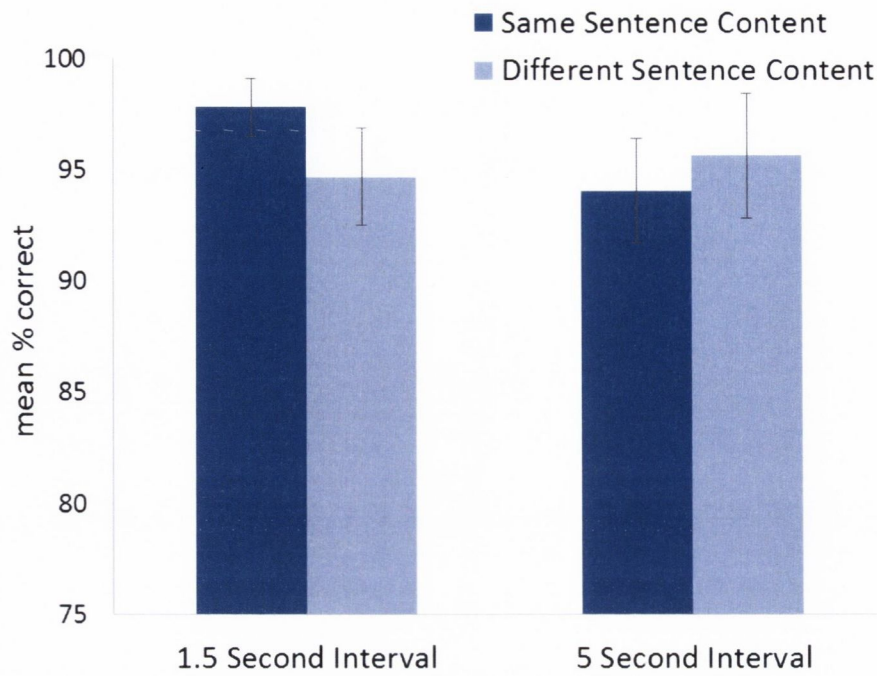


Figure 5.5. Plot showing the interaction between the sentence content and interval on the mean accuracy scores for younger adults in Experiment 8. Error bars show ± 1 SEM.

Reaction Times Younger Adults

The RTs to correct matches only were included in analysis. Furthermore, RTs which were 2.5 standard deviations away from the mean RT (in either direction) for each condition, for each group (short or long interval), were excluded (approximately 7 % of the data). Participants mean RTs were subjected to a 2x2 mixed design analysis of variance (ANOVA), with Face (static or dynamic) and Sentence content (same or different) as a within-subject factors and Interval duration (short or long) as a between-subject factor. The analyses revealed that there was no main effects of Interval duration [$F(1,14)=2.65$, $p=0.133$] or Sentence content [$F(1,14)<1$], or interaction between Interval duration and Sentence content [$F(1,14)=3.33$, $p=0.09$] found. No main effect of Face [$F(1,14)<1$] or interaction between Face and Interval duration [$F(1,14)<1$] was found. Furthermore, there was no interaction between Face and Sentence content [$F(1,14)= 1.26$, $p=0.28$] or between Face, Sentence content and Interval duration [$F(1,14)= 1.40$, $p=0.26$].

5.5.3. Discussion

The accuracy results for the older adult group in Experiment 8 revealed that the addition of a visual face did not enhance accuracy performance for matching speaker identity, relative to when the voice was presented in isolation (see accuracy results Experiment 7). Indeed, accuracy scores were roughly equated to Experiment 7, such that performance was at approximately 84%

for same content trials and 80% for different speech content trials, regardless of whether the face was learned within the context of its static or dynamic face counterpart. However, interestingly in this group reaction time scores revealed a different pattern of performance compared to when the voice had been learned in isolation. Specifically, when the voice was presented with a static face, the time taken to match speaker identity across novel speech utterances was significantly slower, than when the voice had been paired with its dynamic counterpart. Therefore, it is likely that when the voice was presented with its corresponding, synchronized facial identity, older adults could more readily encode the face and the voice into a single percept. This effect was more pronounced when the task demands were difficult i.e. matching the voice across sentences which were not presented during learning. This is arguably a harder task than matching identity across the same sentence content, as it requires the participant to extract the invariant, paralinguistic identity cues which support voice perception independent of what was being said.

Although the presence of a matching face appeared to modulate the encoding of speaker identity in the older group, both the accuracy and reaction time performance of the younger adults revealed no effect for the presence of the face during voice learning, even when the interval between learning and test was extended. Extending the interval is likely to increase memory demands, making this arguably a more difficult task. This finding is different from previous studies which have indicated that learning the voice with a face can enhance longer-term voice recognition (e.g. von Kriegstein & Giraud, 2006). Nevertheless, accuracy results did reveal that matching the speaker across changes in sentence content was poorer than matching based on the learned sentence content, particularly when the interval was relatively short between learning and test (1.5 seconds versus 5 seconds). It is likely that at shorter intervals it was easier to represent the voice for 'same' rather than 'different' sentence content trials. However, this benefit for trials with the same sentence content may have been reduced at the more extended interval owing to a possible overall decay in memory for the voice (approx. 94 % for both same and different sentence content). As such, it is possible that matching the speaker across changes in sentence content, under more challenging conditions, may reveal how the face may support voice processing in younger adults.

5.6. EXPERIMENT 9

Experiment 9 was designed to examine how the presence of a dynamic, relative to static face, during learning would affect identity matching in younger adults under more challenging task conditions. Specifically, the current experiment contained test trials which required the participant to match either the face *or* the voice to the learned identity. Furthermore, due to the

results of relatively poorer performance in matching the voice across different sentence content trials, test items always contained a change in sentence content or a change in facial motion or pose, relative to what was learned. As such, unlike Experiment 8, this task cannot be completed by attending to one sensory modality alone. Rather, performance must be achieved by attending to both signals. Under such circumstances it is possible that dynamic, relative to static, facial cues may facilitate the representation of the face *and* the voice as a *single* identity percept (Schweinberger et al., 2011, 2007). Indeed, Schweinberger et al. (2011) noted that the N170 response, an event-related potential (ERP) component which is associated with and consistently evoked by the visual presentation of a face (review see Eimer, 2011), was larger when a dynamic face was presented with its corresponding vocal identity, relative to when the face was presented in isolation. One possible explanation offered by Schweinberger et al. (2011) for such an effect, was that the presence of a time-synchronised voice may drive attention to the articulating face which, in turn, may generate an increase in the face selective response in the ERP signal. As such, it was predicted that the performance for matching the face and the voice would be more efficient when the spatiotemporal relationship between the face and the voice is maintained, rather than removed (i.e. static face) during learning.

5.6.1. Method

Participants

Sixteen younger adults, recruited from the student population at Trinity College Dublin (12 female, 12 right handed, mean age= 23.6 years, SD= 4.6 years) participated in this experiment and were compensated with a 5 Euro book voucher for their time. All younger adults reported normal or corrected to normal vision and all rated their hearing as either 'Very Good' or 'Excellent' based on the questions outlined in the Participants section of Experiment 7.

Stimuli and Apparatus

The stimulus set is as described in Experiment 8.

Design

The experiment was based on a within-subject factorial design with Face (dynamic or static) and Test item (face or voice) as repeated factors. As Experiment 8, the experimental protocol was based on a delayed matched to sample, two interval forced choice paradigm, where a target voice and face pairing (static or dynamic) was displayed, followed by two sequentially

presented test items. However, unlike Experiment 8, test items could consist of either the voice or the face. In each trial the participant was required to learn the audio-visual identity and to then match the learned identity to a test voice or face. The dependent variables were response accuracy and response time taken to match the identity of the target. In terms of analysis, and in line with the hypothesis outlined above, an a priori decision was made to contrast face and voice matching performance separately for *static* face-voice pairings and *dynamic* face-voice pairings using planned comparisons. However, for completeness the overall ANOVA results are also reported.

As Experiment 8, the experiment consisted of 80 trials in total. In half of the trials both the face and its corresponding voice identity were presented in spatial and temporal alignment during learning (i.e. as a video of a speaking actor). However, in the other half of the trials a static face was presented with its corresponding voice during learning. The static image was always extracted from the same sentence clip (see Experiment 8) and was displayed for the same duration as the voice clip. Furthermore, in half of the trials the test items consisted of voice recordings of a sentence which was always different to the target voice (“This is my house” or “The window is closed”) (40 trials). In the other half of the trials, test items consisted of face images. These images were either of a *dynamic* face clip which varied from the target face clip (silent video from “This is my house” or “The window is closed”) (20 trials) or a *static* face image which differed between the target and test presentation (/h/ “This is my house” or /w/ “The window is closed”) (20 trials). Test faces were either static or dynamic in order to equate the test items for face and voice test trials. Specifically, for face matching trials the face at test always reflected the type of presentation (static or dynamic), just as the auditory only voice matching trials always reflected the method of presentation of the auditory signal. Thus any face matching effects cannot be attributed to a change in sensory cue at test. Each face or voice was displayed for the same duration as it was presented during the learning stage. Furthermore, each face (voice) was used once as a target face (voice) and once as a distractor face (voice) in the audio-visual dynamic learning condition and once as a target face and once as a distractor face in the audio-visual static learning condition. Half of the participants learned the face-voice pairing of “This is my house” as the target stimulus, while the test face images (static or dynamic) or voice clips were taken from the sentence “The window is closed”. In turn, the other half of participants always learned the face-voice pairing of “The window is closed” as the target stimulus and test face images (static or dynamic) or voices were taken from “This is my house” sentence clips.

All 80 trials were presented in a single test block and the order of trials was randomized across participants. On each trial one of the two test faces or voices always matched the identity

of the learned target face or voice and one was a distractor stimulus. Participants always matched face identity to a static (or dynamic) test face, if a static (or dynamic) face image was displayed during audio-visual learning. Distractor faces and voices were always matched to the sex of the target identity. The order of the correct match test item was counterbalanced across trials, such that it appeared directly after the target face in only half the trials. Participants were not alerted as to whether a trial would contain a face or voice item at test.

Procedure

In each trial, participants were first presented with a bimodal face and voice from the same actor identity. They were instructed to learn both the target face and voice during presentation. After the bimodal stimulus was presented a blue fixation cross appeared in the centre of the screen. The fixation was displayed for 1,500 milliseconds. Participants were instructed that after the blue cross disappeared they would see two faces or hear two voices presented sequentially, (i.e. they would view one face directly after the other, or hear one voice directly after the other). They were informed that the face and voice test trials would be intermixed during the testing session and that they would not be alerted as to whether a trial would contain face or voice items at test. However, they were told that the test stimulus would differ to what had been learned, such that the sentence content would change or the test face would contain a novel pose or movement. They were instructed that their task was to decide whether the first or the second face (or voice) matched the previously learned identity. A response screen appeared immediately after the two test faces or voices were presented which instructed participants to press button 'z' on the keyboard if they thought the first test option matched the target identity or 'm' if they thought the second option matched the target identity. Participants were told to always view both test options and to respond as quickly as possible when they saw the response screen. A response triggered the offset of the current trial and the onset of the next trial. Each trial was preceded by an inter-trial interval of 500ms.

Before beginning the main experiment participants completed a practice block of four trials which consisted of 2 face (static or dynamic learning) and 2 voice (static or dynamic learning) matching trials. All face and voice identities used in the practice block were unfamiliar and were not part of the main experiment stimulus set. In parallel, to Experiment 7 and 8 participants were given verbal feedback as to their accuracy in performance during the practice trials. All participants were given the opportunity, if needed, to repeat these 4 trials until they became comfortable with the testing procedure. See Figure 5.6 for an example of trial.

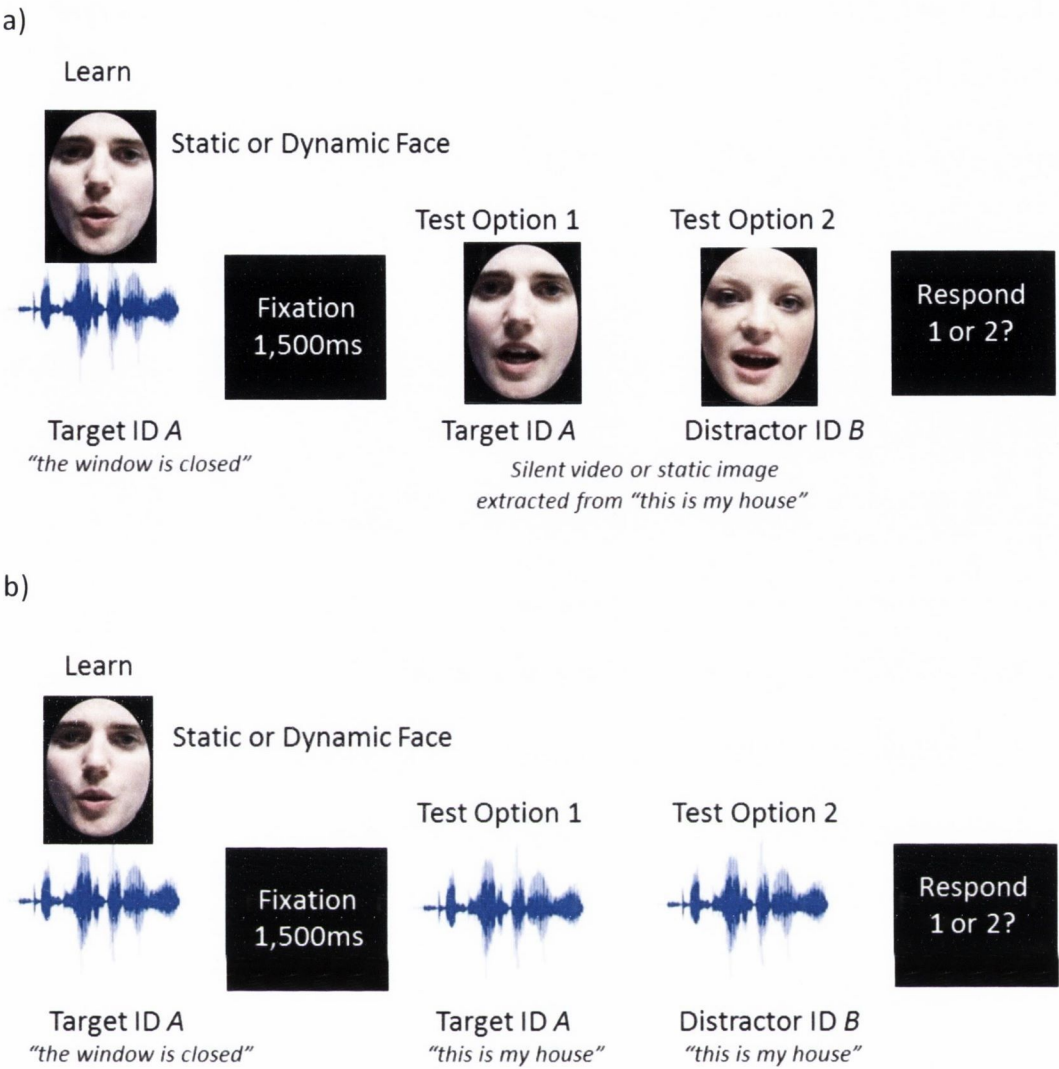


Figure 5.6. An illustration of a sample audio-visual speaker learning trial, where the voice is paired with either its static or dynamic face counterpart during learning. The upper panel a) represents a face matching trial in which the test face images (match or distractor) were always static or dynamic, while the lower panel b) represents a voice matching trial.

5.6.2. Results

Accuracy

Accuracy was calculated as the percentage of correctly matched faces or voices in each condition (i.e. *static* or *dynamic* face-voice pairings at learning) for each participant. The mean percentage of correct responses across participants was subjected to a 2x2 repeated measures ANOVA, with Face (static or dynamic) and Test item (face or voice) as repeated factors. Analysis revealed that there was no main effect of Face [$F(1,15)<1$]. A main effect of Test item was observed [$F(1,15)=7.65$, $p=0.01$, $\eta_p^2=0.38$], with better matching performance for the face

($M=97.7\%$) than voice ($M=94.5\%$) at test. There was no interaction observed between Face and Test item [$F(1,15)=1.42, p=0.25$].

Planned comparisons revealed that there was a difference between face and voice matching performance for static face-voice pairings [$F(1,15)=10.97, p=0.004$]. Here, voice matching ($M=93.7\%$) was less accurate than face matching ($M=97.8\%$). However, there was no difference observed between face and voice matching performance for dynamic face-voice pairings [$F(1,15)=2.09, p=0.17$]. In this condition voice matching ($M=95.3\%$) was comparable to face matching ($M=97.5\%$) performance (see Figure 5.7).

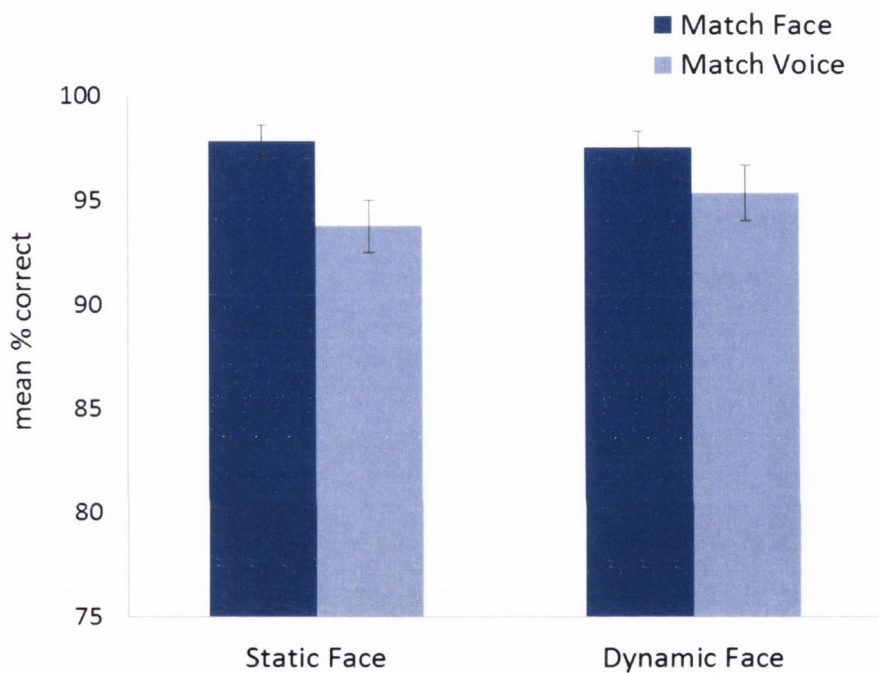


Figure 5.7. Plot showing the mean accuracy scores for matching the face or the voice, when the target identity was presented as a static face-voice pairing (left of figure) or a dynamic face-voice pairing (right of figure) during learning. Error bars show ± 1 SEM.

Reaction Times

The RTs taken to correct matches only were included in analysis. Furthermore, RTs which were 2.5 standard deviations greater or less than the mean for each condition, for test item (face or voice), were excluded from analysis (approximately 4 % of the data). The RTs were analysed using a 2x2 repeated measures ANOVA, with Face (static or dynamic) and Test item (face or voice) as repeated factors. This analysis revealed no main effect of Face [$F(1,15)<1$]. A main effect of Test item was found [$F(1,15)=10.37, p=0.006, \eta_p^2=0.41$], with slower reaction times to match the face ($M=460$ ms) than the voice ($M=399$ ms) at test. An interaction between Face and Test item was also observed [$F(1,15)=7.7, p=0.02, \eta_p^2=0.34$].

Planned comparisons revealed that there was no difference between the reaction times taken to match the face ($M=412$ ms) and voice ($M=443$ ms) for *static* face-voice pairings [$F(1,15)=2.92$, $p=0.12$]. However, there was a difference observed between reaction time performance for face and voice matching for *dynamic* face-voice pairings [$F(1,15)=13.56$, $p=0.002$]. In this condition, the time taken to match the voice ($M=387$ ms) was faster than the time taken to match the face ($M=477$ ms) (see Figure 5.8).

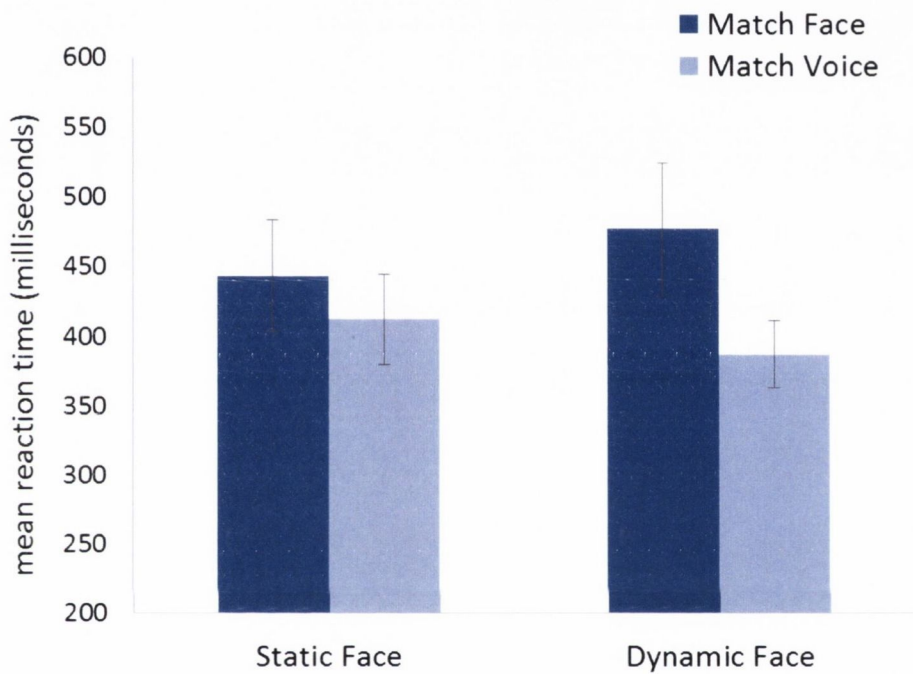


Figure 5.8. Plot showing the mean reaction times for matching the face or the voice, when the target identity was presented with a static face-voice pairing (left of figure) or a dynamic face-voice pairing (right of figure) during learning. Error bars show ± 1 SEM.

5.6.3. Discussion

The results from Experiment 9 revealed that younger adults’ performance was less accurate for matching the voice across changes in speech utterances, than for matching the face across a novel pose or motion profile. However, planned comparisons revealed that this poorer speaker matching performance was mediated by the learning condition. Specifically, speaker matching performance was poorer than face matching when the voice had been learned in the presence of a static, rather than a dynamic face. The finding that the voice may act as a weaker cue to identity than the face, is in line with previous findings (e.g. Stevenage et al., 2014; Stevenage, Hugill, & Lewis, 2012). However, the current results extend previous reports and highlight that the voice is more likely to emerge as a weaker identity cue when the natural spatio-temporal relationship between the face and voice is disrupted. Indeed, voices were also

responded to faster than faces in this dynamic face-voice condition. Taken together, these results suggest that the face and the voice are encoded more efficiently as a single identity percept when the spatiotemporal relationship between the face and voice information is preserved.

5.7. EXPERIMENT 10

5.7.1. Face and voice interactions in developmental prosopagnosia

As mentioned previously, the voice appears to be a weaker cue to identity than the face, possibly as it is a less reliable signal owing to the variations which can occur through both speech manner (e.g. pitch, timbre, and rhythm) and speech content. For example, Stevenage et al. (2014) demonstrated that although cross-modal priming effects were evident in the processing of familiar faces and voices (see also Stevenage, Hugill, & Lewis, 2012), these effects were more robust when the face was used to prime the voice, rather than when the voice was used to prime the face. This finding, along with findings of enhanced speaker identification following audio-visual learning (e.g. von Kriegstein & Giraud, 2006), argue that the coupling of face and voice information may enhance the representation of speaker identity in the neurotypical brain. In turn, this raises an interesting point pertaining to how speaker identity might be represented when face recognition is compromised, such as in the case of developmental prosopagnosia. While a number of studies on DP have documented that such individuals report that they rely on the voice as a cue for person identity (Grueter et al., 2007; Kennerknecht et al., 2006; Kennerknecht, Kischka, Stemper, Elze, & Stollhoff, 2011; Le Grand et al., 2006; McConachie, 1976), only a small number of studies have explicitly tested familiar voice recognition in this population (Jones & Tranel, 2001; Schall & von Kriegstein, 2014; von Kriegstein, Kleinschmidt, & Giraud, 2006).

One such study which did investigate voice recognition in a single case of DP (referred to as participant 'SO') demonstrated that their familiar and unfamiliar voice recognition accuracy was *worse* than that of age-matched controls (von Kriegstein et al., 2006). Interestingly, in that study von Kriegstein et al. (2006) noted that the functional response profile in areas TVA to voice stimuli and the functional connectivity between FFA and TVA to familiar voices was preserved in SO. However, they argue that this was not sufficient to support the representation of familiar speakers. Rather, they suggested that the unusual response activations in area FFA to individual face identities in DP, rather than to faces at a categorical level per se, likely underpinned SO's poor ability to represent speaker identity at an individual level. As such, von Kriegstein et al. (2006) suggested that SO's deficit in familiar voice recognition likely arose in the context of her

poor face processing abilities, as she was unable to use her experience based multisensory representations of identity, which are employed in neurotypical processing.

If prior multisensory experience is important in shaping voice recognition ability, then it is possible that both voice recognition and *voice matching* may be impaired in DP. Specifically, it is conceivable that voice processing may be weaker in a DP cohort due to an inability to use the dynamic face as a way of refining or shaping voice processing abilities (von Kriegstein et al., 2006). Based on the findings from neurotypical studies on cross-modal matching performance for silent dynamic face and voice recordings (Kamachi et al., 2003; Lander et al., 2007; Rosenblum et al., 2006), it is possible that dynamic face information may modulate auditory processing, possibly by biasing the observer to attend to specific features in the speech signal, (e.g. rhythmic structure), which is carried in both the voice and its dynamic face counterpart. Indeed, the results from Chapter 4 argue that apperceptive developmental prosopagnosics may have difficulty in representing, or at least tracking, the changeable aspects of the face during face motion. Indeed, other studies have also reported poor lip reading in this cohort (De Gelder & Vroomen, 1998; Lange et al., 2009). It is therefore possible that UM and PL may present with a general deficit in voice matching ability, owing to the fact that they may find it more difficult than neurotypical individuals to encode the relevant perceptual cues which can be used to represent voice identity.

To that end, Experiment 10 examined speaker matching ability in UM and PL using the same paradigm and the same older adult participant data for comparative performance, to that reported in Experiment 7. Unfortunately, neither UM or PL could partake in the follow up audio-visual version of the experiment⁵ (i.e. Experiment 8) and as such results are based on both prosopagnosics matching performance when the voice was learned in the context of an auditory-only stimulus only. Note that both UM and PL reported that they had 'Very Good' hearing.

5.7.2. Method

Procedure

The procedure was as described in Experiment 7.

Stimuli and Apparatus

The stimuli and apparatus are as described in Experiment 7.

⁵ Both UM and PL indicated that they may possibly return at a later stage in the year for follow up testing.

5.7.3. Analysis and Results

See Figure 5.9 for a plot illustrating UM and PL's individual scores and the age-matched controls' mean accuracy performance (note control data is the same data as that plotted in Figure 5.2). Accuracy performance for both prosopagnosics on the same and different sentence content trials was compared to performance by their age-matched control group using a series of modified *t*-tests described by Crawford, Garthwaite, & Porter (2010). The test is designed to compare whether a single case score is significantly different from a control or normative sample and has been used in the literature for assessing deficits in prosopagnosia (as highlighted in Chapter 1 and 4) and recently for phonagnosia (Garrido et al., 2009). These analyses revealed that UM showed significantly poorer speaker-identity matching performance for same [$t(15)=-1.77, p=0.04, Z_{CC}=-1.82, 95\% \text{ CI}=-2.62, -1.0]^6$ and for different [$t(15)=-1.74, p=0.05, Z_{CC}=-1.80, 95\% \text{ CI}=-2.58, -0.98$] sentence content trials (performance approached chance for different content trials, see Figure 5.9). Further analysis using the Revised Standardised Difference Test (Crawford et al., 2010) revealed that although UM's performance was poorer than that of the control group, her performance pattern for matching speaker identity across same, relative to different, sentence content trials was similar to controls [$t(15)=0.10, p=0.49$]. In terms of PL's performance, the *t*-test analysis revealed that speaker matching accuracy was similar to controls for both same [$t(15)=0.57, p=0.29$] and different sentence content trials [$t(15)=0.11, p=0.46$]. The accuracy performance of PL was of a similar pattern to controls for matching the identity of the speaker across same relative to different sentence content [$t(15)=0.60, p=0.56$].

However, reaction time analysis revealed that PL was significantly slower at matching identity for both same ($M=1096 \text{ ms}$) [$t(15)=1.82, p=0.04, Z_{CC}=1.88, 95\% \text{ CI}=1.04, 2.70$] and different ($M=1293 \text{ ms}$) [$t(15)=2.61, p=0.01, Z_{CC}=2.69, 95\% \text{ CI}=1.61, 3.75$] sentence content trials than controls. Furthermore, UM showed this same slower response profile for same ($M=1108 \text{ ms}$) [$t(15)=1.88, p=0.04, Z_{CC}=1.93, 95\% \text{ CI}=1.08, 2.77$] but not for different ($M=943 \text{ ms}$) [$t(15)=1.17, p=0.10, Z_{CC}=1.20, 95\% \text{ CI}=0.54, 1.84$] sentence content trials.

⁶ Z_{CC} or Z_{DCC} = Effect Size for difference between case and controls; CI= Confidence Interval

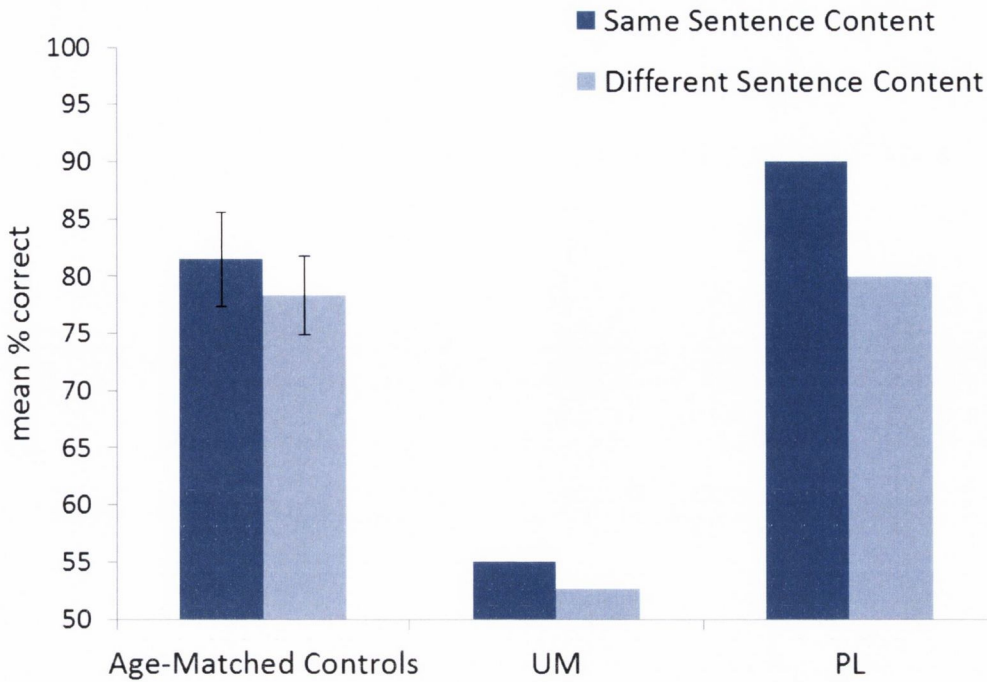


Figure 5.9. Plot showing the mean accuracy for matching speaker identity across same and different sentence content trials, for age-matched controls and prosopagnosics tested in Experiment 10. Error bars show ± 1 SEM.

5.7.4. Discussion

The results indicated that PL displayed similar accuracy in matching speaker identities to age-matched controls. Moreover, PL's performance for matching speaker identity across same, relative to different, sentence content was similar to the performance of the age-matched control group. As such, it appears that PL has a preserved ability to match vocal identity in spite of poor face perception and recognition abilities. However, in general both UM and PL's reaction time performance appeared to be slower than that of the mean control performance. In terms of UM's accuracy performance, matching speaker identity across same, relative to different, sentence content was similar to controls. However, in marked contrast to the performance of either PL or the control group, UM's speaker identity matching performance was significantly more error prone. Indeed, UM's accuracy levels for both same and different sentence content trials approached chance performance in this task (see Figure 5.9). As such it appears that UM may have an overlapping impairment for matching facial identity and speaker identity. Indeed, UM stated that she does not rely heavily on the voice as a cue to identity. Importantly, UM was impaired in a task involving voice perception rather than voice recognition. Thus she demonstrated a perceptual deficit in pattern recognition and featural analysis of the voice signal (similar to her perceptual deficit in matching faces), rather than a deficit in explicit recognition,

which may on some occasions indicate a multimodal semantic deficit occurring at later stages of processing (Liu, Pancaroglu, Hills, Duchaine, & Barton, 2014).

5.8. General Discussion

The aim of the study reported in this Chapter was to examine the role of facial motion on the perception of unfamiliar voices in older, younger and developmental prosopagnosic individuals. Previous ageing studies have largely investigated how dynamic facial cues can support speech, rather than speaker, perception. Furthermore, how disorders of face processing, particularly DP, may impact on voice processing abilities remains largely unexplored. Here participants learned a speaker identity from an auditory clip (Experiment 7, Experiment 10) or from a bimodal stimulus comprising of a static face-voice or dynamic face-voice pairing (Experiment 8). Following a delay, participants matched the target speaker by choosing from two available voices clips, which contained either the same or different sentence content. The results from Experiment 7 revealed that speaker matching was poorer in older than younger adults, but that the change in sentence content had a similar effect on performance in both groups. For UM, a developmental prosopagnosic, there was a similar decrease in performance for matching different relative to same sentence content, however compared to the older adult control group her overall speaker matching performance was extremely error prone (Experiment 10). The results from Experiment 8 (where the speaker was presented as a bimodal stimulus) demonstrated that older adults' accuracy in matching speaker identity was not enhanced when the face was learned with its dynamic, relative to static corresponding facial identity. Nevertheless, matching the voice across changes in speech content was significantly slower when the voice was learned in the presence of a static snapshot of the face identity, rather than a time synchronised audio-visual display. However, in contrast to the older adult group, learning the speaker with a dynamic, relative to static face, had no effect on either accuracy or reaction time performance for the younger adult group, even when the interval between the learning and test stage was extended (Experiment 8). However, when the task was made more difficult such that younger participants had to match *either* the face or the voice at test (Experiment 9), a benefit for the dynamic face-voice learning emerged. Here, speaker matching was equitable to face matching performance only when the voice had been learned in the presence of its dynamic, rather than static, face counterpart.

The finding that older adults are poorer at matching speaker identities than younger adults, is consistent with the reported ageing effects in the face processing literature (Lee et al., 2011, 2014). Indeed, it has been argued that both the face and the voice are represented in a

similar manner in the neurotypical brain (Belin et al., 2004; Campanella & Belin, 2007). It is therefore possible that, in line with the face processing literature, older age may be associated with more broadly tuned face (Lee et al., 2014) *and* voice representations. As such, older adults may be less able to represent the subtle variation in perceptual features which can be used to support a unique vocal signature for each identity that is represented in memory. Importantly, these age-related effects were observed when memory demands were low, suggesting changes in voice processing at the level of perceptual encoding, rather than later stages of memory recall. Prior research has suggested that older adults may be less sensitive to paralinguistic speech cues when the task is to represent the linguistic content of the speech signal (Pilotti et al., 2003). However, the current findings demonstrate that these effects can still emerge when older adults are advised not to attend to the content of the speech signal only (as it can change between learning and test) but rather to focus on the paralinguistic cues which can support identity processing across changes in the linguistic content.

A significant role for the face was observed when the task was to match the voice across novel, rather than learned sentence content. In Experiment 8, older adults were slower to match speaker identities when the voice had been learned with a static snapshot of the face, rather than a time synchronised dynamic face (i.e. video). Thus it is likely that shared cross-modal, spatiotemporal cues may allow for the face and the voice to be represented as a single identity percept in this condition (Kamachi et al., 2003; Schweinberger et al., 2007). The availability of the dynamic face may help to better represent the paralinguistic cues which support speaker perception in memory, allowing for better speaker matching across changes in speech utterances. Indeed, older adults reported that the presence of a static face was a ‘distraction’ when trying to attend to the speech signal, suggesting that the static face-voice pairings were likely perceived as discrete or unrelated events. However, owing to the relatively robust encoding of the unisensory signal, it is likely that younger adults could reliably represent the voice even in the presence of a static, relative to dynamic face. Schweinberger et al., (2007) reported that familiar voice recognition was compromised when the voice was presented with a time synchronised but non-corresponding facial identity. Specifically, participants were slower to respond to the familiarity of the voice in this condition. However, this response profile was eliminated when the voice was presented with a static, non-corresponding face identity. Schweinberger and colleagues suggested that when the face was static observers could reliably separate the two sensory cues allowing for accurate recognition of the auditory signal in the incongruent condition. In contrast, when the face was perceived to be articulating in synchrony with the voice the two sensory cues were likely integrated, making it difficult to independently monitor the auditory speech signal for identity.

The current study extends these findings to demonstrate that learning new speaker identities is facilitated when the voice is learned with its corresponding dynamic, relative to static face identity. In this condition the two signals can be reliably represented as a single percept. Furthermore, these effects may be more apparent as the ability to represent speaker identity declines and they appear to be evident during the *encoding* of the identity, indicating face and voice interactions at the early (Schweinberger et al., 2007), rather than later (Bruce & Young, 1986), stage of sensory processing.

While the presence of the face appeared to modulate speaker matching performance, it is important to note that it did not enhance older adults overall matching performance relative to when the voice was learned as an auditory-only stimulus. Moreover, younger adults matching performance was not affected by bimodal learning, even when the interval between learning and test was extended. Therefore, it is possible that *audio-visual enhancement*, that is better matching performance when a multisensory stimulus is learned relative to a unisensory stimulus, in speaker matching may be more readily observed for established face and voice representations. Although these representations can be rapidly acquired (von Kriegstein et al., 2008; von Kriegstein & Giraud, 2006), individuals may need to be exposed to the face and voice pairing for a longer duration during learning. Furthermore, previous studies on younger adults, which have observed these behavioural benefits have instructed the participant to learn to associate both sensory cues for later recall (see e.g. O'Mahony & Newell, 2012; von Kriegstein & Giraud, 2006).

It is important to note that younger adults performance, both in terms of accuracy and reaction time, was very good across all tasks which involved matching *only* the identity of the speaker (mean accuracy approximately $\geq 95\%$, mean reaction times approximately $\leq 450\text{ms}$). Indeed, it is possible that the static condition was not 'challenging' enough for the younger adults, as they could reliably represent the voice independent of the face information. Had the task been harder, for example if a dynamic *incongruent* condition had been introduced, it is possible that younger adults' performance may have been affected owing to the disruption in change over time information between the face and the voice signal. For example, Calvert, Campbell, & Brammer (2000) observed sub-additive responses in STS when an actors' lip movements (for a continuous stream of speech) was dubbed with an incongruent speech signal. In the current context, reversing the frame order of the video, whilst leaving the audio unaltered, would likely produce the same effect. Indeed, the results from Chapter 3 demonstrated that altering the spatiotemporal alignment of visual face images interfered with facial identity matching in the younger adult group. However, younger adults' identity matching performance was affected

when the task was made more difficult, requiring the observer to match the face or the voice in interleaving trials. Specifically, in line with previous findings (e.g. Stevenage et al., 2014, 2012) it emerged that the voice was a weaker identity cue, as indicated by greater errors when matching the voice relative to the face, when the voice had been learned with its static face counterpart. This suggests that when the spatiotemporal relationship between the face and the voice is altered, the face may dominate as the most salient identity cue (Stevenage et al., 2014, 2012), giving rise to relatively poorer speaker matching in this condition. In contrast, when the face and voice were learned in synchrony (i.e. preserved spatiotemporal alignment), both speaker *and* face matching performance was equivalent. This suggests that although the dynamic face may assist in refining or enhancing voice processing, this effect may be most readily observed in younger adults' performance under situations of task uncertainty. Indeed, it could be argued that these situations of uncertainty, i.e. when either a face or a voice is to be matched for identity, are more ecologically valid than when participants are directly told which identity cue to attend to. Indeed, in everyday encounters we may have to recognise an individual from the face or the voice at any one time.

5.8.1. Voice processing in developmental prosopagnosia

The results from this study suggest that voice matching may be, to some extent, compromised in DP. Specifically, both UM and PL were slower at matching speaker identity than their age-matched controls. Moreover, UM's performance was at chance when matching the voice identity across novel speech content and only slightly better when the same sentence was presented at test (55 %). These findings offer some indication that both voice recognition (von Kriegstein & Giraud, 2006) *and* voice matching may be compromised in this cohort. Indeed, even if age-matched 'normal' levels of performance are observed it cannot be assumed that this has been achieved through the same means as in neurotypical processing. For example, von Kriegstein et al. (2008) demonstrated that developmental prosopagnosics did not show the same behavioural enhancement as controls for auditory-only speaker recognition when the voice had been learned with its corresponding dynamic facial identity. In contrast, this enhancement was evident when the task was to recognise the content of the speech signal (von Kriegstein et al., 2008; Schall & von Kriegstein, 2014). Thus it is possible that atypical processing in one modality, e.g. failure to represent facial identity at an individual level, may modulate voice processing abilities. Moreover, although speculative, it is possible that due to a poor ability to use additional face information in this population that individuals with DP may require repeated exposure to a voice (in the presence or absence of a face) before it reaches the same level of familiarity as

occurs in neurotypical (audio-visual) processing. Although the current findings may point towards this interpretation, there are important caveats to consider. For example, follow up tests examining both speaker matching for bimodal learning (face and voice) and identity matching across visual and auditory modalities (Kamachi et al., 2003; Lander et al., 2007; Lawrence D Rosenblum et al., 2006) would offer valuable insight into the possibility of a co-occurring voice processing deficit in DP. Moreover, owing to the heterogeneous nature of DP (Le Grand et al., 2006) it is likely that voice processing may also vary across individuals.

Indeed, it is also important to consider that voice processing can reliably emerge in the absence of visual input e.g. in congenitally blind individuals. However, these cases have been associated with increased functional recruitment in voice sensitive areas of STS, known as intra-modal cortical recruitment (Gougoux et al., 2009). Specifically, Gougoux et al. (2009) noted that the level of activation in this area was reliably associated with a behavioural index of voice discrimination ability in this group. What is perhaps astounding is that Gougoux et al. (2009) also noted that fusiform areas, typically associated with face processing (Haxby et al., 2000; Kanwisher et al., 1997) also exhibited functional activation in congenitally blind individuals, leading the authors to conclude that this area may be functionally involved in the computations which support *both* face *and* voice representations, even in the absence of prior visual experience.

For the most part, the results from this series of experiments in this Chapter suggest that dynamic facial cues may play a role in both typical and atypical voice processing and that these interactions may occur at earlier stages of processing than previously assumed (Bruce & Young, 1986; Burton et al., 1990). Further studies, for example, examining how face processing is affected by the presence of the voice in individuals with developmental phonagnosia (Garrido et al., 2009) rather than prosopagnosia, may help elucidate how the face and voice are combined in normal identity processing.

Chapter 6

Vocal cues can modulate the perception of gender in point-light walker displays

6. Vocal Cues Can Modulate the Perception of Gender in Point-Light Walker Displays

6.1. Abstract

Previous findings from younger adult studies have demonstrated that observers can discriminate the sex of an individual from biological motion cues alone. In addition, information from the voice can support sex judgements, suggesting that sexually dimorphic cues are shared across the senses. While there is some evidence to suggest that non-vocal, but sex specific, auditory cues can modulate perceived gender in biological motion displays, little is known as to whether sex cues from the voice may also support sex classification in these dynamic displays. Moreover, whether the ageing process affects the ability to discriminate and combine these sexually dimorphic cross-modal cues remains unexplored. To that end, Experiment 11 assessed younger and older adults' ability to judge the sex of point-light walkers which varied in terms of the saliency of the sex cues displayed. Results indicated that older adults were poorer than younger adults in detecting sex differences between the walking profiles. However, both groups showed a similar level of performance when classifying 'gender-ambiguous' walkers. Younger (Experiment 12A, 12B) and older adults (Experiment 14) then judged the sex of gender-ambiguous walkers with or without accompanying male or female voices. The results demonstrated that younger, but not older, adults' performance was modulated by the presence of the voice. Moreover, this result could not be attributed to attending to the auditory cue alone, as there was no effect of the voice when incongruent audio-visual sex pairings were presented for 'exaggerated' male and female walkers (Experiment 12B). Furthermore, this effect was also observed when the speech content and the task were changed, such that the participant rated the perceived gender rather than the sex of the walker (Experiment 13).

The second part of this Chapter turned to assess whether developmental prosopagnosics UM and PL could classify the sex of a walker based on the dynamic visual displays described above. It was noted that UM and PL showed similar performance to age-matched controls on this task (Experiment 15A). Moreover, UM could reliably extract sex cues in the voice, such that she judged gender-ambiguous walkers to be male or female when they were paired with male or female voices respectively (Experiment 15B). Thus in line with the literature suggesting preserved sex judgements from face images in cases of DP, UM's performance profile suggests that this ability may be preserved across different sensory cues in DP. Taken together, the results extend

the findings from Chapter 5 and suggest that audio-visual interactions are not limited to the face and voice but may extend to other aspects of person perception.

6.2. Introduction

As highlighted in the preceding chapters, the perceptual system is sensitive to dynamic information in the face, which can be used to support the processing of identity (Knappmeyer, Thornton, & Bülthoff, 2003; Lander, Hill, Kamachi, & Vatikiotis-Bateson, 2007; Lander & Chuang, 2005; Longmore & Tree, 2013; Roark, Barrett, Spence, Abdi, & O'Toole, 2003; Thornton & Kourtzi, 2002). Indeed, Rosenblum et al. (2002) noted that this ability is so robust that the identity of an individual can be discriminated when the facial motion profile is presented as only a few points of moving lights onscreen. In addition, a number of studies have highlighted that we can also detect other socially relevant information from these facial motion displays, including the sex of a face (Hill & Johnston, 2001; Morrison, Gralewski, Campbell, & Penton-Voak, 2007). Moreover, this remarkable ability is not only limited to dynamic facial cues, as observers can detect these socially relevant signals from the biological motion cues displayed by the *body* in motion (Cutting & Kozlowski, 1977; Johansson, 1973; Kozlowski & Cutting, 1977).

The perception of biological motion from the body has been studied extensively through the use of point-light displays. As highlighted in Chapter 1, these displays are often created through positioning markers (or points of light) on an actor's joints. These markers can be tracked by cameras and edited at later stages of processing such that the markers are rendered white on a black background. Or when points of light are attached the luminance of the background is reduced, such that only these points of light are visible. Very little information can be extracted from these displays when shown in their static form. However, when the actor begins to move these visible points of light give rise to the perception of a human figure in motion (Johansson, 1973). Indeed, when presented with these sparse visual displays younger adult observers can often discriminate the identity (Cutting & Kozlowski, 1977; Stevenage, Nixon, & Vince, 1999; Troje, Westhoff, & Lavrov, 2005) and gender⁷ (Barclay, Cutting, & Kozlowski, 1978; Kozlowski & Cutting, 1977; Mather & Murdoch, 1994; Pollick, Kay, Heim, & Stringer, 2005; Troje, 2002) of the individual in motion (among other attributes).

However, a number of studies have demonstrated that the processing of biological motion declines in older age (Billino, Bremmer, & Gegenfurtner, 2008; Konar & Troje, 2013; Legault, Troje, & Faubert, 2012; Norman, Payton, Long, & Hawkes, 2004; Pilz, Bennett, & Sekuler, 2010). For example, compared to their younger counterparts, older adults are worse at detecting

⁷ The term 'gender' is often used throughout the biological motion literature to refer to categorisation of an individual as *male* or *female*. However, in general this term usually refers to the classification of *femininity* or *masculinity*. For example, a female/male may have a masculine/feminine appearance but still be reliably categorised as female or male, respectively.

(Billino et al., 2008) and discriminating the direction (Pilz et al., 2010) of point-light walkers, particularly when there is noise in the visual display. However, in line with the results from Chapter 3 which indicated a relatively preserved ability to process facial motion in older age, it appears that the age-related decline in biological processing is modest in comparison to the decline which is evident in other lower-level motion tasks (Billino et al., 2008; Norman et al., 2004). For example, Billino and colleagues reported that older and younger adults' thresholds for discriminating the direction of motion in noisy visual displays were more similar for biological motion, than for random-dot motion patterns, a class of stimuli commonly used to address the mechanisms underpinning lower-level motion perception (Bennett, Sekuler, & Sekuler, 2007; Snowden & Kavanagh, 2006; Roudaia, Bennett, Sekuler, & Pilz, 2010; Hutchinson, Arena, Allen, & Ledgeway, 2012). This led the authors to conclude that older age is not associated with a general global decline in motion perception. Rather they argued that the specialised neural substrates which underpin biological motion processing (see e.g. Grossman & Blake, 2002; Grossman et al., 2000) may be less vulnerable to the ageing process. Indeed, others have shown that specific types of biological motion processing may remain relatively well preserved with advancing age. For example, Konar & Troje (2013) reported similar performance for younger and older adult groups in discriminating the sex of an individual from point-light walker displays, while confirming previous findings of poorer lower-level motion processing in this cohort.

Indeed, the perceptual system appears to be sensitive to a range of cues in biological motion displays which can be exploited to categorise the sex of an individual in motion. These dynamic cues involve sometimes subtle differences in the 'lateral body sway' of a person in motion (Halevina & Troje, 2007; Mather & Murdoch, 1994). Specifically, lateral body sway varies across the sexes as, on average, men tend to move their shoulders from side to side more than their hips, while the reverse is true for their female counterparts (Mather & Murdoch, 1994). In addition, structural cues pertaining to the shape of the body are also recoverable from these moving visual displays. For example, the shoulder-to-hip ratio is, on average, larger for males than for females. Therefore, the relative motion of the shoulders and the hip joints can capture differences in the physical form or 'torso structure' of male and female bodies (see Cutting, Proffitt, & Kozlowski, 1978). However, it appears that lateral body sway, i.e. the dynamic patterns, offer the most salient cues for classifying or discriminating the sex of the walker (Halevina & Troje, 2007; Mather & Murdoch, 1994).

Although the ability to detect the sex of an individual from biological motion cues alone appears robust, there is also evidence that other sensory cues can modulate the perception of these dynamic displays. For example, studies have suggested that the presence of additional form

cues pertaining to the sex of an individual, i.e. body shape, can affect how these motion cues are perceived (McDonnell, Jörg, Hodgins, Newell, & O'Sullivan, 2009; Thornton, 2006). For example, McDonnell, Jörg, Hodgins, Newell, & O'Sullivan (2009) demonstrated that when a female (male) body form was animated with a male (female) walking profile, participants tended to rate the character as appearing ambiguous in terms of sex. In contrast, when a gender-ambiguous body form (i.e. androgynous body shape) was animated with a female or male walking profile, the character appeared more female or male, respectively. These results parallel the findings from the face literature which have shown that both form and motion cues likely *interact* when perceiving others (Knappmeyer et al., 2003). Taken together, these results suggest that the perceptual system is likely to integrate information across different sensory cues when representing the sex of an individual in motion.

In addition, there is also evidence that information from other sensory modalities may affect the perception of biological motion cues, particularly when the visual information is difficult to resolve. These studies have largely focused on interactions between vision and audition for the purpose of either detecting biological motion in noisy displays (Thomas & Shiffrar, 2013; Thomas & Shiffrar, 2010) or discriminating the direction or perceived synchrony of walking (Brooks et al., 2007; Saygin, Driver, & de Sa, 2008; Schouten, Troje, Vroomen, & Verfaillie, 2011; Wuerger, Crocker-Buque, & Meyer, 2012) or more complex action sequences (Arrighi, Burr, Nazionale, & Inoa, 2009; Petrini & Pollick, 2010). However, a small number of studies have addressed audio-visual interactions pertaining to the extraction of higher-order visual representations, such as sex or gender (McDonnell & O'Sullivan, 2010; van der Zwan et al., 2009). For example, van der Zwan et al. (2009) examined the effect of auditory sex cues on the perception of point-light walkers which were ambiguous in terms of their perceived sex. Specifically, these visual walkers transitioned from an exaggerated female to an exaggerated male walker through a series of gradual increments which alter the reliability of the information about the sex of the point-light walker (PLW) displays. Importantly, within this transition the PLW will at a certain point appear 'gender-ambiguous' as there are no reliable dimorphic cues to the sex of the character in the walks. Interestingly, van der Zwan and colleagues observed that when these 'gender-ambiguous' walkers were paired with unambiguous female auditory walking sequences (i.e. female footsteps), that the walkers were judged more often to be female, than when they were presented without any additional auditory cues (i.e. vision only).

However, sex information is also available in the auditory modality in other natural stimuli, such as the voice. Indeed, as outlined in Chapter 5, the voice offers a wealth of paralinguistic information, including, but not limited to, the sex of the speaker (Belin et al., 2004;

Pernet & Belin, 2012). Our ability to determine the sex of a voice appears to be largely determined by the perceptual system's sensitivity to variations in the fundamental frequency (f_0 , the perceived pitch) of the vocal signal (Latinus & Taylor, 2012; Schweinberger, Kawahara, Simpson, Skuk, & Zäske, 2014). Specifically, the pitch (or f_0) of the voice is, on average, higher in females than in males (Feinberg et al., 2005; Lattner, Meyer, & Friederici, 2005; Schweinberger et al., 2014). A number of studies have demonstrated that systematically manipulating voice pitch to be higher or lower, consequently increases the perceived femininity or masculinity of the voice (e.g. Feinberg, Jones, Little, Burt, & Perrett, 2005; Feinberg, DeBruine, Jones, & Perrett, 2008). Furthermore, it has been shown that these changes in auditory pitch can affect the perceived sex of face images in younger adults (Little, Feinberg, DeBruine, & Jones, 2013; Smith, Grabowecky, & Suzuki, 2007), particularly when these images are gender-ambiguous in appearance (Smith et al., 2007). However, to date it remains unexplored as to whether these sexually dimorphic vocal cues can modulate the perceived gender in point-light walker displays, in both younger and older adults.

6.3. Outline of the Experiments

The first part of this chapter reports a series of experiments which investigated whether sex information from the voice could modulate the visual perception of 'gender-ambiguous' point-light walkers, following a similar procedure to that reported by van der Zwan et al. (2009). Here, younger and older participants judged the sex of a series of point-light walkers in which the structural and dynamic cues pertaining to the sex of the walker were either gender-ambiguous, perceptually salient, or exaggerated (Troje, 2002, 2008) (Experiment 11). Younger (Experiment 12A, Experiment 12B) and older (Experiment 14) adults then judged the *sex*, i.e. male or female, of these gender-ambiguous walkers, with or without the presence of a male or female voice. In addition, younger adults rated the perceived *gender*, i.e. the perceived masculinity or femininity of these walkers, with or without accompanying male and female voices (Experiment 13).

6.4. EXPERIMENT 11

Experiment 11 was designed as a 'baseline' experiment for the audio-visual studies which follow. Specifically, it was designed to investigate whether the ability to detect the sex of a walker from biological motion displays, which were systematically manipulated to alter the sexually dimorphic structural and dynamic cues pertaining to the sex of the walker (Troje, 2002), was affected by the ageing process. Moreover, this experiment was designed to investigate if the point of subjective ambiguity concerning the perceived sex of the walker, referred to as a 'gender-

ambiguous' walker, (Troje, Sadr, Geyer, & Nakayama, 2006; van der Zwan et al., 2009) was the same in younger and older adults. Specifically, the gender continuum described by Troje (2002) has a range of walks which can be readily perceived as 'exaggerated' versions of males or females, to walkers which offer salient but less exaggerated sex cues, and also walkers which contain no reliable sexually dimorphic information. These latter walker types have been labelled as gender-ambiguous and have previously been shown to give rise to an equal proportion of male and female responses when they are classified for sex (i.e. chance performance) (Troje, 2002; van der Zwan et al., 2009).

6.4.1. Method

Participants

Fourteen young adults, recruited from the student population at Trinity College Dublin (8 female, all right handed, mean age= 23.0 years, SD= 4.1 years) participated in this experiment and were given course credits in return for their time. All younger adults reported normal or corrected to normal vision. Fourteen, community-dwelling, older adults also participated in this task (9 female, all right handed, mean age= 70.3 years, SD= 4.5 years). All had normal cognitive function, as assessed by the MoCA (Nasreddine et al., 2005) (i.e. all scored ≥ 24 of 30). All older participants had normal visual acuity (mean binocular LogMar score= 0.10, SD= 0.1 and LogMar score= 0.01, SD= 0.13 at 4m and 40 cm respectively) and contrast sensitivity (mean score= 1.97, SD= 0.08, assessed using the Pelli-Robson Contrast Sensitivity Test) for their age. All participants wore appropriate corrective lenses, if needed, at the time of testing.

Stimuli and Apparatus

Visual stimuli consisted of a set of point light walkers (PLW) made available by the author Troje (2002, 2008). Each PLW contained 15 individual point-lights positioned as if attached to the major joints of a human figure. The PLW stimuli were shown in constant motion and were perceived to be walking in a frontal direction, as if positioned on a treadmill. All lights were white and displayed on a black background. All of the PLW were the same height and shown from a frontal position. This frontal view was chosen as sex discrimination has been observed to be best for PLW shown from this frontal viewpoint (Halevina & Troje, 2007; Mather & Murdoch, 1994; Troje, 2002).

The PLW stimuli varied in the strength of the gender information displayed. In line with the stimuli used by van der Zwan et al. (2009), this resulted from the manipulation of both structural *and* dynamic cues. PLWs were extracted from different points of a gender continuum developed by Troje (2002) which ranged from -6 to 6: increasingly negative scores on this continuum correspond to PLW which are more female in appearance, while increasingly positive scores correspond to PLW which are more male in appearance. For example, point 6 on the continuum represents a walker which is 6 standard deviations from the mean in the male range. Each standard deviation manipulates both the structural (in this case the dimensions between the 2 points which represent the shoulders and the 2 points which represent the hips), as well as the dynamic cues which are associated with the sex of the walker (e.g. shoulder or hip sway). PLWs were chosen at 5 points along the continuum; -6 exaggerated female walk, -3 female walk, -1 gender-ambiguous walk, 3 male walk, 6 exaggerated male walk. See Troje (2002, 2008) for a full description on the production and creation of the PLW stimuli⁸. An illustration of the gender continuum can be seen in Figure 6.1.

As mentioned, the -1 point corresponds to a *perceptually* 'gender-ambiguous' walk. These walkers carry no reliable structural or motion cues which the perceptual system can exploit to detect the gender of the walker from the visual information only. This point has been shown to give rise to an equal number of male and female responses when young, neurotypical observers are asked to classify the sex of the walker (e.g. see van der Zwan et al., 2009) (see Figure 6.1 for an illustration of the gender continuum). To increase the number of stimuli, the walks were further varied by 5 body weights (very heavy, heavy, normal, light and very light) at each point on the gender continuum, giving rise to a total of 25 stimuli. Point-light walkers at -6, -3, -1, 3, 6 are hereafter referred to throughout this chapter as exaggerated female, female, gender-ambiguous, male and exaggerated male, respectively. All walkers completed 2 step cycles (i.e. they were presented walking for four steps) and were displayed for 2,000ms. This duration was chosen as Barclay et al. (1978) reported that participants could reliably judge the sex of a walker only when the stimulus was displayed for 1,600 milliseconds or longer and the walker completed at least 2 step cycles.

The experiment was programmed in and run using Presentation® software and was presented on a Dell OptiPlex 755 PC and displayed on a HP L1710 17" LCD colour monitor. The screen resolution was set to 1024 x 768 pixels and the video clips were presented at a resolution

⁸ Although these stimuli can be artificially manipulated to modulate the visual appearance of the walker, it is important to note that the algorithms which allow for such manipulations were originally developed based on motion capture data from real life recordings of male and female walkers.

of 540 x 432 pixels. Participants were positioned in front of the screen at a distance of approximately 60 cm. As such, the videos subtended a visual angle of approximately 6.5° horizontally and 20° vertically on the screen. The screen was always black during the presentation of a video. The experiment was conducted on-site, in a testing laboratory.

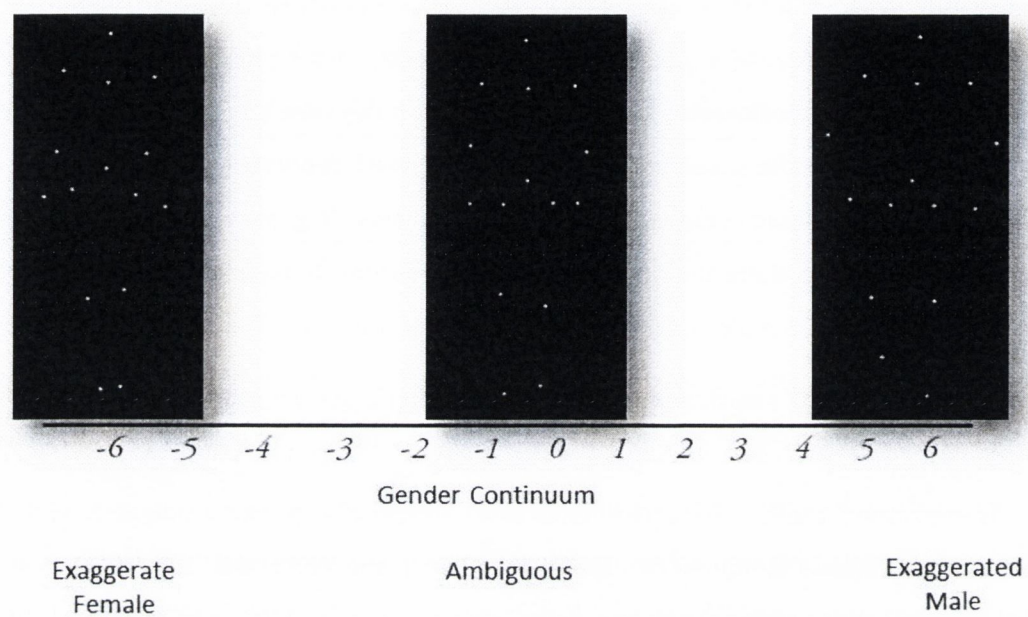


Figure 6.1. An illustration of the gender continuum described and developed by Troje (2002). Negative values (left of centre) represent walkers which are more female in appearance, while positive values (right of centre) represent walkers which are more male in appearance. For illustration purposes 3 static images, extracted from the dynamic sequences at point -6 exaggerated female, -1 gender-ambiguous and 6 exaggerated male, are depicted.

Design

The experiment was based on a mixed design with Age (younger or older adults) as a between-subjects factor and Sex Classification Score (i.e. slope, see Analysis section below) as the within-subject factor. The experimental protocol was based on a two alternative forced choice paradigm. On each trial a walker was presented in the centre of the screen and immediately following presentation the participant was required to indicate if the walker was 'male' or 'female'. Each walker type was viewed 5 times, i.e. 5 repetitions at each point on the gender continuum amounting to a total of 25 trials. All trials were presented in a single test block and the order of trials was randomized across participants.

Procedure

Each participant was positioned in front of the computer screen. They were advised that they would view a number of short video clips of moving points of light which would look like a person in motion. They were instructed that they should decide on each occasion if the presented walker was male or female. A response screen appeared immediately after the walker stimulus was presented which instructed participants to press button 'z' on the keyboard if they perceived the walker to be female or 'm' if they perceived the walker to be male. Participants were instructed to always view each video in its entirety and respond only when they saw the response screen. This was to ensure that participants were exposed to the same amount of dynamic information across all trials (Barclay et al., 1978; van der Zwan et al., 2009). In line with previous studies, the speed of response was not emphasised and the response screen remained onscreen until a button press was made. This response triggered the offset of the current trial and the onset of the next trial. Each trial was preceded by an inter-trial stimulus of 500ms. Four practice trials preceded the main experiment and the videos were randomly selected from the stimulus set on each occasion. See Figure 6.1 for an illustration of the stimuli used in the experiment.

Analysis

In line with previous studies (van der Zwan et al., 2009) performance at each point on the gender continuum was calculated as the proportion of *male* responses, that is, the number of times a participant responded 'male' divided by the number of repetitions for that stimulus type. As such, the proportion of male responses was calculated for each participant, for each walker type on the gender continuum. A logistic function was then fitted to each individual's data and the slope⁹ of this function was taken as a measure of sensitivity in classifying the sex of the walker (Sex Classification Score). Three participants (1 younger, 2 older adults) were excluded from this analysis as it was not possible to obtain a reliable estimate of the slope from the logistic function to their data. As a result, the findings reported below are based on the analysis of 13 younger adults and 12 older adults' performance. The proportion of male responses for 'gender-ambiguous' walkers was also compared to chance performance (i.e. 0.5) for both groups using a single sample *t*-test. Note that in line with previous studies, reaction time data were not used as a measure of performance in this task (Troje, 2002; van der Zwan et al., 2009). The individual slopes

⁹ The slope has been taken as an indicator of sensitivity in classifying the sex of the walker in previous studies using these similar point-light walker stimuli e.g. Konar & Troje (2013).

for younger and older adults were then subjected to a 'by group' analysis using an independent *t*-test.

6.4.2. Results

Figure 6.2 shows the mean slope of the younger and older adult group as a function of the gender continuum. As can be seen from the plot, older adults demonstrated poorer performance on this task. This age difference was confirmed using an independent *t*-test [$t(1,23) = -2.32, p = 0.02$]. Specifically, younger adults were more sensitive in classifying the sex of the walkers (M slope = 0.3), compared to older adults (M slope = 1.44). Note that in this task lower values mean better performance. As expected, a single sample *t*-test comparing the proportion of male responses for 'gender-ambiguous' walks to chance performance, confirmed that responses did not significantly differ from chance for both the younger [$t(13) = 0.09, p = 0.9$] and older [$t(12) = -1.17, p = 0.27$] adult group.

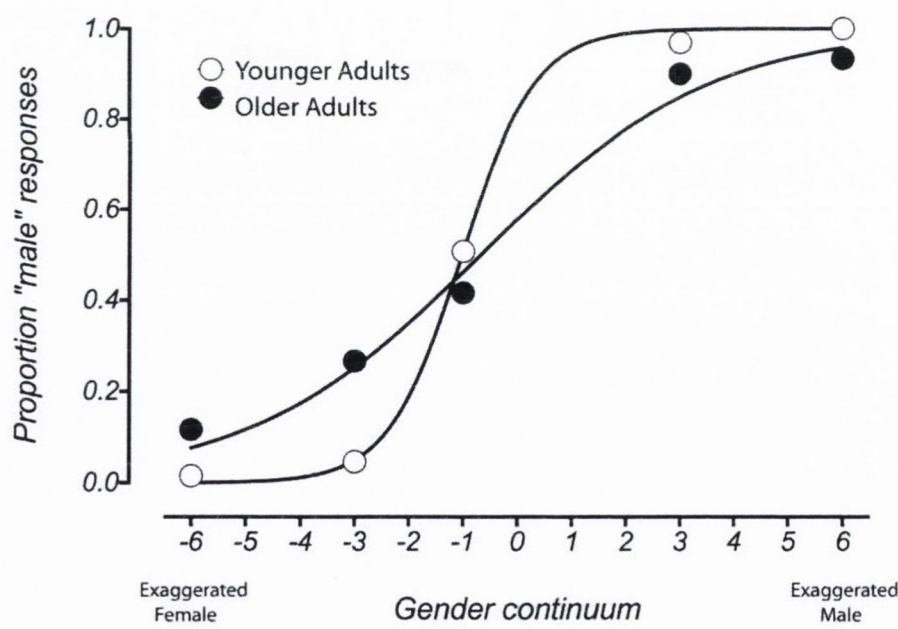


Figure 6.2. Plot showing the mean slope for younger and older adult groups as a function of PLW stimuli positioned along the gender continuum. Visual information only was presented in this experiment.

6.4.3. Discussion

Overall, compared to younger adults' performance, older adults were poorer at classifying the sex of the visually presented walkers. Specifically, as shown by the slope of the functions (see Figure 6.2) younger adults could reliably classify the sex of the walker, both when

exaggerated and less exaggerated, cues to the sex of the walker were available, i.e. they reliably classified the sex of both exaggerated male (female) and male (female) walkers. In contrast, older adults were less certain in their responses, suggesting a general decline in the ability to represent biological motion cues to the sex of an individual. This finding is different from Konar & Troje (2013) who observed that gender classification was robust in older age. However, what is interesting is that both groups demonstrated chance performance on classifying the sex of the walker when the visual information was unreliable, i.e. gender-ambiguous walkers. In this condition dynamic and structural cues are ambiguous and this suggests that both groups perceive these walkers in a similar manner. With that in mind, Experiment 12A, 12B and 13 examined younger adults' ability to classify the sex of a 'gender-ambiguous' walker when accompanied by unambiguous male or female voices, whereas older adults performance on this audio-visual task was investigated in Experiment 14.

6.5. EXPERIMENT 12A

The purpose of Experiment 12A was to examine if accompanying male and female vocal cues may affect sex judgements for visual point-light walkers in younger adults. Previous findings by van der Zwan et al. (2009) demonstrated that the presence of auditory cues to the sex of an individual, depicted by female walking sequences, modulated the perceived sex of gender-ambiguous walkers. These walkers were typically judged more often to be female when paired with an unambiguous female auditory walking sequence (i.e. female footsteps), compared to when no auditory cues were presented. As such, it was predicted that the presence of a salient sex cue available in the *voice* may modulate the perceived sex of gender-ambiguous walkers in a similar manner to that observed by van der Zwan and colleagues. Moreover, as both male and female voices were presented it was predicted that the perceived sex of the gender-ambiguous point-light displays would be biased towards the sex of the voice.

6.5.1. Method

Participants

An opportune sample of 14 younger adult participants (11 female, mean age= 28.1 years, SD= 11.5 years) took part in this experiment. All were individuals who attended the 'Love Lab: The Science of Desire' exhibition in the Science Gallery, Trinity College Dublin. Prior approval for this experiment was obtained from the School of Psychology Research Ethics Committee at Trinity College Dublin and the approval letter can be viewed in Appendix B. All participants were naïve to

the purpose of the experiment and none had participated in the previous experiments. All participants gave their informed written consent to participate and were debriefed as to the nature of the experiment upon completion.

Stimuli and Apparatus

The visual stimuli are as described in Experiment 11. In addition, a set of 10 male and 10 female vocal stimuli described in Experiment 7, Chapter 5 were used as the auditory stimuli in this study, with the exception that the sentences used in the current experiment were 'You look good today', and 'Hello there, Hi'. These sentences were social sentences and were chosen based on the brief of the exhibition which this experiment was part of; the 'Love Lab: The Science of Desire' exhibition held in the Science Gallery, Trinity College Dublin. All details of the recording including the actors used are as described in the Stimuli and Apparatus section of Experiment 7, Chapter 5. These voice clips were pre-rated for perceived gender strength and the rating procedure and the criteria for choosing the voice samples to use in the main experiment are described below.

In the pre-rating study the 20 individual voice recordings were rated for perceived gender strength by a group of 17 participants who were all unfamiliar with the speakers' identities (10 female, mean age=27.4 years, SD=5.32 years). All voice samples were presented twice in a randomized order, and one neutral sample sentence "This is my house" was always presented during the rating study. The voices were presented over speakers, in a quiet room, and 4 to 5 participants were present in the room at one time. After each voice recording was played participants were asked to indicate, using a scale of 1 to 7, the perceived gender strength of the voice, i.e. how masculine or feminine they perceived the voice to be. A rating of 1 indicated very masculine, 4 neutral, 7 very feminine. Thus, higher values corresponded to increased perceived femininity and lower scores corresponded to increased perceived masculinity in the voice. All responses were made by pen and paper and the rating scores from each participant were subsequently collated for analysis.

For both male and female voices, four voice samples were chosen from each sex (8 recordings in total). Based on a median split of the ratings for both sexes, four unique vocal identities were chosen for the female stimuli (referred to as *f-1*, *f-2*, *f-3* and *f-4*), these stimuli were reliably rated as female (mean rating *female* 5.6, SD=0.69). In turn, 4 voices were chosen for the male stimuli (referred to as *m-1*, *m-2*, *m-3* and *m-4*), (mean rating *male* 2.4, SD=0.68). Paired *t*-tests confirmed that gender ratings were significantly different across both sexes [$t(16) = -10.39$, $p < 0.001$]. Two sample sentences used in the current experiment were 'Hello there, Hi' and 'You

look good today'. Both clips were approximately 1.5 to 1.7 seconds in duration. The experiment was presented on a colour LCD monitor (1,024 x 768px) using Presentation® software running on a 3Ghz PC with 8GB RAM. Auditory stimuli were presented over Sennheiser 405 stereo Headphones at a super-threshold but comfortable decibel level.

Design

The experimental protocol was based on a two alternative forced choice paradigm. On each trial a walker was presented in the centre of the screen and immediately following presentation the participant was required to indicate if the walker was 'male' or 'female'. The experiment consisted of 85 trials, 25 of which were vision only and 60 which were audio-visual. For vision only trials each walker type was viewed 5 times as described in Experiment 11. In audio-visual trials both exaggerated male (exaggerated female) and male (female) walkers were paired with vocal information which was *congruent* with the sex of the walker and each walker type was presented 10 times. For audio-visual pairings for exaggerated female and female walkers; the voice of *f-1* and *f-2* (or *f-3* and *f-4*) was presented 6 times, while *f-3* and *f-4* (or *f-1* and *f-2*) was presented 4 times with each walker type. This was counterbalanced across participants. The same audio-visual pairing was applied to the exaggerated male and male walkers, but with the male vocal stimuli. Again this was counterbalanced across participants. Furthermore, the sentences 'Hello there, Hi' and 'You look good today' were presented an equal number of times throughout the audio-visual trials. As this study was interested in how vocal information may modulate the perceived sex of gender-ambiguous walkers, walkers at -1 on the gender continuum were paired with either *male* or *female* voices. Gender-ambiguous walkers were presented 10 times with a female voice and 10 times with a male voice (20 trials), details for these pairings remain as stated above.

Procedure

The procedure was as described in Experiment 11, with the exception that participants were informed that on some occasions they would hear a voice accompanying the presentation of the visual walk. In line with van der Zwan et al. (2009) participants were instructed to respond to the sex of the *visual* walker on all trials. Participants wore headphones for the duration of the experiment.

Analysis

Performance at each point on the gender continuum was calculated as the proportion of *male* responses reported, that is, the number of times a participant responded 'male' divided by the number of repetitions for that stimulus type. Scores were calculated separately for the vision-only and audio-visual conditions. In line with the current study, van der Zwan et al. (2009) presented observers with visual and audio-visual pairings for a number of points on the gender continuum seen in Figure 6.1. However, they based their statistical analysis on observed differences between vision and audio-visual conditions for gender-ambiguous walkers only, using a series of *t*-tests. This was based on the assumption that auditory interactions would be more readily observed when the visual stimulus was ambiguous. Here, it was decided to address the analysis in a similar way by using *t*-tests to compare sex classification for vision and audio-visual conditions at different points of the gender continuum. The proportion of male judgements for gender-ambiguous walkers was compared to chance performance, for vision only, audio-visual 'female' pairings and audio-visual 'male' pairing trials. Note that previous studies have demonstrated that sex classification does not exhibit above chance performance when these walkers are displayed in the visual only conditions (e.g. Brooks, Schouten, Troje, & Verfaillie, 2008; van der Zwan et al., 2009). For completeness, the proportion of male judgements for exaggerated male (exaggerated female) and male (female) walkers were also directly compared for the vision only and audio-visual trials for each walker type.

6.5.2 Results

The proportion of male responses for gender-ambiguous walkers was compared to chance performance (i.e. 0.5) for vision only, audio-visual (AV) female, and audio-visual male pairings using a series of one sample *t*-tests. As expected, and consistent with the results of the previous experiment, judgements of the sex of the walker in the vision only conditions was similar to chance performance, i.e. responses did not differ from 0.5 [$t(13)=-0.21$, $p=0.83$] (mean proportion 'male' responses=0.49). However, judgements of AV female-voice pairings were different from chance [$t(13)=-2.44$, $p=0.03$] (mean proportion 'male' responses=0.34). In this condition, the number of 'male' responses was relatively low such that walkers were judged to be more female in appearance, when they were paired with a female voice. The results for the audio-visual male-voice pairings demonstrated a similar trend in favour of an increase in the number of 'male' responses when gender-ambiguous walkers were paired with a male voice (mean proportion 'male' responses= 0.62). However, this trend was weak and scores in this

condition failed to differ statistically from chance performance [$t(13)=-1.78, p=0.09$]¹⁰, see Figure 6.3.

Importantly, there were no differences between the number of 'male' responses to the visual only stimuli and the responses to the AV presentation for the exaggerated male [$t(13)=1.0, p=0.33$], exaggerated female [$t(13)=1.7, p=0.1$], male [$t(13)=0.0, p=0.99$] or female [$t(13)=-0.29, p=0.77$] walkers. The proportion of 'male' responses to the exaggerated male (exaggerated female) walkers for vision and AV presentations were 0.92 and 0.90 (0.08 and 0.12), respectively. In addition, the proportion of 'male' responses for male (female) walkers for the vision-only and AV presentations were 0.9 and 0.9 (0.12 and 0.13), respectively. These results suggest that the responses of younger adults to the male or exaggerated male visual PLW were not biased by the presence of the voice.

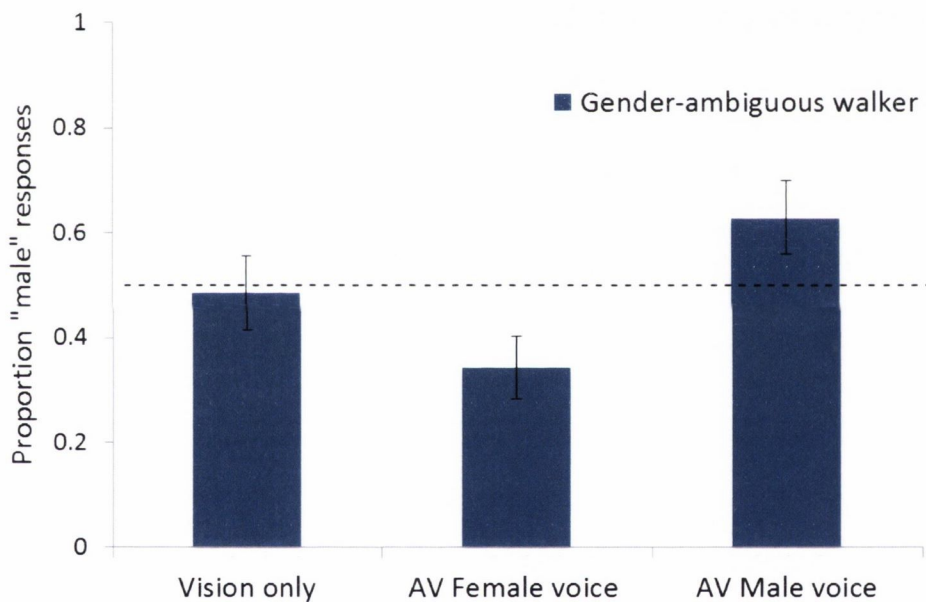


Figure 6.3. The proportion of male responses reported for gender-ambiguous walkers, when presented with no voice, a female voice and male voice in Experiment 12A. The dashed line indicates chance performance. Note that sex judgements only differ from chance performance for AV female voice pairings. Error bars represent +1 SEM.

6.5.3. Discussion

The results from Experiment 12A suggest that the presence of a voice affected the perception of sex in gender-ambiguous biological motion displays. This effect was most

¹⁰ A paired t -test did however confirm that there was a significant difference between the number of 'male' responses for AV male-voice and AV female-voice pairings for gender ambiguous PLW [$t(13)=-3.19, p=0.007$]. This suggests some level of disambiguation of the visual walker in favour of the sex of the voice presented.

pronounced when the walker was presented with a female voice, where the proportion of ‘male’ responses was significantly lower than what would be expected by chance (i.e. a decrease in favour of female judgements). However, there was a trend for the male voice to affect visual sex judgements. As such, these results parallel and expand the effects observed by van der Zwan et al. (2009) who noted that gender-ambiguous walkers were judged more often to be female in ‘appearance’ when they were paired with a female auditory walking sequence. While these results are compelling in suggesting that auditory voice cues can disambiguate, or at least modulate, the perceived gender of visual point light displays, it is important to consider there may be a number of reasons as to why this result was observed. The most obvious suggestion may be that the results reflect a response bias, rather than a perceptual effect. Specifically, it is possible that participants may have adopted a strategy of responding to the auditory modality only, rather than the results reflecting interactions between both the visual and auditory cues. This was explicitly tested in the following experiment.

6.6. EXPERIMENT 12B

The current experiment attempted to disentangle the effects observed in Experiment 12A which may have reflected a participant response bias to attend to the auditory stimulus only. Here Experiment 12A was repeated with one key alteration; exaggerated male and exaggerated female walkers were paired with both congruent and *incongruent* auditory sex cues, while all other features of the design remain as described in Experiment 12A. It was expected that if participants were responding to the auditory cue alone then the responses should be affected by the incongruent AV sex pairing. For example, participants should respond more often that an exaggerated male walker is female when it is presented with a female voice, than when it is presented with a male voice. However, no difference between the congruent and incongruent AV pairings would suggest that participants were more likely responding to the *visual* stimulus.

6.6.1. Method

Participants

An opportune sample of 15 younger adult participants (9 female, mean age= 29.6 years, SD= 15.2 years) took part in this experiment. Again all were individuals who attended the ‘Love Lab: The Science of Desire’ exhibition in the Science Gallery, Trinity College Dublin. All participants were naïve to the purpose of the experiment and none had participated in the previous

experiments. As before, all participants gave their informed written consent to participate and were debriefed as to the nature of the experiment upon completion.

Stimuli and Apparatus

The stimuli and apparatus are as described in Experiment 12A.

Procedure

The procedure is as described in Experiment 12A.

Design

The design is as described in Experiment 12A, with the exception that exaggerated male and exaggerated female walks were paired with either congruent or incongruent auditory sex cues. Specifically, on half of the AV trials exaggerated male (exaggerated female) walkers were paired with female (male) voices, while on the other half of the trials they were paired with the congruent sex, i.e. exaggerated male with male voice or exaggerated female with female voice.

Analysis

The analysis is as described for Experiment 12A with the exception that proportion of 'male' responses were also analysed for congruent and incongruent AV sex pairings for exaggerated male and exaggerated female walkers using paired *t*-tests.

6.6.2. Results

The proportion of 'male' responses for gender-ambiguous walkers was compared to chance performance (i.e. 0.5) for vision only, audio-visual female, and audio-visual male pairings using a series of one sample *t*-tests. The results revealed that for the vision only conditions participants' judgements of the sex of the walker did not differ from chance performance [$t(14)=0.71$, $p=0.48$] (mean proportion 'male' responses=0.53). However, the number of 'male' responses made to the AV male pairings did differ from chance [$t(14)=2.30$, $p=0.03$]. There was an increase in the number of 'male' judgements, such that observers tended to judge the gender-ambiguous walkers more often to be male when they were paired with a male voice (mean proportion 'male' responses=0.65). However, the results for AV female pairings suggested that the presence of the female voice did not disambiguate the visual perception of the walker, as performance did not differ from chance [$t(14)=-1.17$, $p=0.27$] in this condition (mean proportion

‘male’ responses=0.43). Nevertheless, when viewing Figure 6.4 it can be noted that there was a trend toward a lower proportion of male responses in this condition¹¹.

Results indicated that perceived sex judgements for male [$t(14)=-0.56$, $p=0.58$] and female [$t(14)=-0.53$, $p=0.61$] walkers did not differ between the vision only and AV conditions. Here the mean proportion of ‘male’ responses for female vision only and female AV conditions were 0.10 and 0.12, respectively. The mean proportion of ‘male’ responses for the male vision only and AV conditions were 0.93 and 0.95, respectively. Moreover, the results also revealed that there was no difference between sex judgements for exaggerated female walkers when the sex of the voice was congruent (mean proportion ‘male’ responses=0.07) or incongruent (mean proportion ‘male’ responses=0.09) with the walker [$t(14)=-0.65$, $p=0.52$]. In parallel, there was no difference between sex judgements for exaggerated male walkers when they were paired with male (mean proportion ‘male’ responses=0.95) or female (mean proportion ‘male’ responses=0.89) voices [$t(14)=1.58$, $p=0.14$]. These results suggest that the sex cues from the voice did not dominate in the responses made to the PLW stimuli.

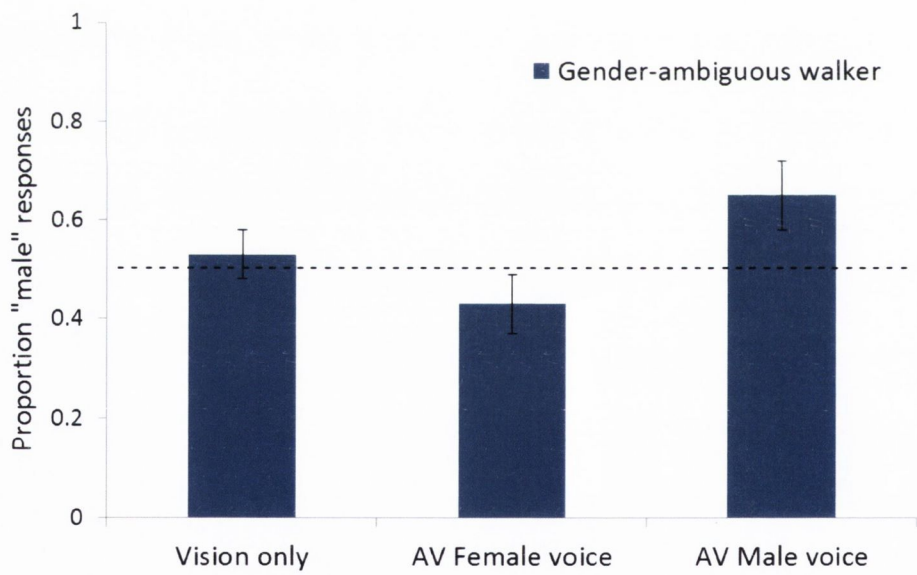


Figure 6.4. The proportion of male responses reported for gender-ambiguous walkers, when presented with no voice, a female voice and male voice in Experiment 12B. The dashed line indicates chance performance. Note that sex judgements only differ from chance performance for AV male voice pairings. Error bars represent +1 SEM.

¹¹ A paired t -test did however confirm that there was a significant difference between the number of ‘male’ responses for AV female-voice and AV male-voice pairings for gender ambiguous PLW [$t(14)=-3.31$, $p=0.005$]. Suggesting some level of disambiguation of the visual walker in favour of the sex of the voice presented.

6.6.3. Discussion

The results from Experiment 12B extend the findings from Experiment 12A. Specifically, they suggest that the modulatory effects of the auditory signal on visual sex judgements for gender-ambiguous walkers are unlikely to reflect a general response bias to attend solely to the auditory cue. If this was the case, it would be expected that performance for the exaggerated male and female AV pairings would have been affected by the presence of incongruent auditory sex cues; this was not observed. Indeed, these results parallel van der Zwan et al. (2009) findings. In that study, van der Zwan et al. (2009) noted that these audio-visual interactions were not evident when walkers were paired with walking sequences containing ambiguous gender information. Therefore, the modulatory effects cannot be explained by the presence of *any* auditory signal, rather, they may be explained by the 'gender loading' of the auditory cue, such that it can help disambiguate the perceived sex of the walker.

However, the results from Experiment 12B still do not preclude the possibility that these effects may have arisen at a later stage in the processing of information about the sex of a visual walker, such as at the response stage. An examination of whether these modulatory effects still hold across a change in task may help determine if these effects are more related to perceptual processing than changes in the response strategy of the participant. Moreover, it is possible the effects may be tied to the sentences which are used in the current study, rather than the gender cues present in the speech signal. For example, it could be argued that these 'social sentences' may have induced a state of arousal which may have affected participants' performance. This was examined in the following experiment.

6.7. EXPERIMENT 13

In the final experiment conducted with young adults reported in this Chapter, participants were asked to rate the perceived gender of the visual walker, rather than to classify the walker as male or female. It is likely that this rating may be more sensitive to differences in the perceived sex of the walker arising at a perceptual level, rather than differences at the response output. Specifically, more subtle differences may be observed where the observer is given a range of responses to choose from rather than a forced binary choice of 'male' versus 'female'. Moreover, this experiment also used sentences which vary in content, (i.e. sentence wording), to that reported in Experiment 12A and 12B. It was expected that if the cross-modal modulatory effects observed in Experiment 12A and 12B are related to the sex cues present in the speech signal, then a change in speech content should preserve these effects. Specifically, a change in speech content

should alter the linguistic, rather than the *paralinguistic* cues, pertaining to the sex of the speaker (i.e. changes in perceived pitch).

6.7.1. Method

Participants

As with the previous experiments, participants in this study comprised of an opportune sample of younger adults, who were attendees of 'Love Lab: The Science of Desire' exhibition in the Science Gallery, Trinity College Dublin. 48 younger adults (29 female, mean age=30.8 years, SD=10.2 years) took part in this experiment. None of the participants had participated in the previous experiments and all were naïve as to the purpose of the study. All participants gave their informed written consent to participate and were debriefed as to the nature of the experiment upon completion.

Stimuli and Apparatus

The visual and auditory stimuli are as described in Experiment 12A, with the exception that the sentence content of the speech samples was changed. The sentences used in this experiment were 'This is my house' and 'The window is closed'. Both sentences were approximately 1.8 seconds in duration. As mentioned in Chapter 5, these sentences have been previously rated as semantically neutral in content, i.e. they had no unique emotion ascribed to their wording (see Adolphs, Damasio, & Tranel, 2002; Russ, Gur, & Bilker, 2008).

Design

The design is as described in Experiment 12A, with the exception that the 2AFC paradigm was replaced with a gender rating scale where participants indicated the perceived masculinity or femininity of the walker.

Procedure

The procedure is as described for Experiment 12A with the exception that participants were instructed to indicate the perceived gender (i.e. level of masculinity or femininity) of the walker, rather than the perceived sex (i.e. a binary response of 'male' or 'female'). The gender scale for responses ranged from 1 to 7, where 1 represented very masculine, 4 neutral (i.e.

ambiguous) and 7 very feminine. Again, responses were indicated by button press on the keyboard and the scale remained onscreen until a response was made.

Analysis

The analysis is as described for Experiment 12A, with the exception that gender ratings rather than the proportion of 'male' responses were analysed. For consistency, the ratings for gender-ambiguous walkers for vision only and AV male and female conditions were compared to the value of 4 (i.e. the midpoint of the scale), rather than to chance performance (i.e. 0.5 as previously reported for binary judgements). Ratings for exaggerated male (exaggerated female) and male (female) walkers were also directly compared for vision only and audio-visual trials for each walker type.

6.7.2. Results

The ratings for gender-ambiguous walkers were compared to the scale midpoint (4) for vision only, audio-visual female and audio-visual male pairings using a series of one sample *t*-tests. This analysis demonstrated that participants' gender ratings for gender-ambiguous walkers did not differ from the scale midpoint of 4 for the vision only condition [$t(47)=1.92$, $p=0.07$] (mean rating response=4.2). In addition, judgements for AV male voice pairings did not reliably differ from the scale midpoint of 4 [$t(47)=-1.43$, $p=0.15$] (mean rating response=3.9). However, further analyses revealed that for AV female voice pairings observers tended to rate the gender-ambiguous walkers as more feminine in appearance. In this condition ratings significantly differed from a rating of 4, i.e. ambiguous [$t(47)=4.52$, $p<0.001$] (mean rating response=4.5)¹². See Figure 6.5.

As with the results from the previous experiments, there was no difference between the gender ratings for vision only compared to AV presentation for exaggerated male [$t(47)=0.33$, $p=0.74$] and exaggerated female [$t(47)=-1.03$, $p=0.30$], male [$t(47)=0.96$, $p=0.34$] or female walkers [$t(47)=-1.64$, $p=0.11$] walkers. The gender ratings for exaggerated male (exaggerated female) walkers for vision and AV presentations were 2.0 and 1.99 (6.0 and 6.03) respectively. In addition, the gender ratings for male (female) walkers for vision and AV presentations were 2.45 and 2.40 (5.40 and 5.54) respectively.

¹² However, a paired *t*-test did reveal that there was a significant difference between the gender ratings for AV male-voice and AV female-voice pairings for gender-ambiguous PLW [$t(47)=5.68$, $p<0.001$]. This suggests some level of disambiguation of the visual walker in favour of the sex of the voice presented.

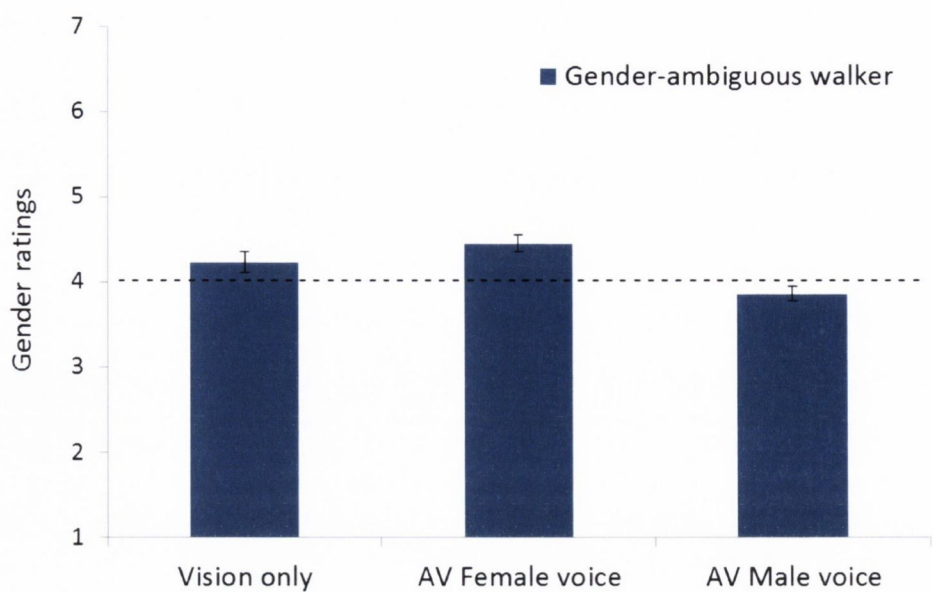


Figure 6.5. The mean gender ratings for gender-ambiguous walkers, when presented with no voice, a female voice and male voice in Experiment 13. The dashed line indicates a ‘neutral’, i.e. ambiguous rating. Note that gender ratings only differ from this neutral rating for AV female voice pairings. Error bars represent +1 SEM.

6.7.3. Discussion

The results from Experiment 13 demonstrate that the addition of auditory cues modulated the perceived *gender* of gender-ambiguous walkers. This result was most reliably observed for walkers presented with a female voice, replicating the findings for Experiment 12A. This finding suggests that observers may have perceived the visual walkers as being more feminine in appearance when paired with a female voice. However, note that looking at the Figure 6.5, there was also a weak trend for walkers to be rated more often as male when presented with a male voice, rather than through vision alone. Moreover, the modulatory effect of auditory sex cues on visual gender judgements was observed when different sentences were used. Thus, the cross-modal modulation observed for the sex judgements of perceptually ambiguous walkers in Experiment 12A and 12B are unlikely to have arisen due to a level of arousal induced by the ‘social nature’ of the sentences used in the two studies. Rather, these results suggest that it is the paralinguistic cues pertaining to the sex of the speaker which may likely modulate the perceived sex of the gender-ambiguous walkers.

6.8. EXPERIMENT 14

The results from Experiments 12A, 12B and 13 demonstrated that vocal information containing unambiguous auditory gender cues modulated the perceived sex and gender of *visual*

point-light walkers in a young adult group, but only when the gender of the walk was perceptually ambiguous. Importantly, the results from Experiment 11 revealed that while older adults were overall poorer at classifying the sex of the walker, *both* groups exhibited chance performance for classifying 'gender ambiguous' walkers. Thus it is possible that older adults' performance may also benefit from the presence of an auditory cue, such that this additional cue may modulate the perceived sex of gender-ambiguous walkers. This was tested in Experiment 14. Here, older adults judged the sex of visual walkers, with or without the addition of male and female voices in the same paradigm as reported in Experiment 12A. It was expected that sex judgements for gender-ambiguous walkers would be modulated in favour of the male or female voice pairing. Moreover, it was also predicted that these audio-visual effects might also be observed for perceptually salient walkers, owing to older adult's relatively poorer performance in classifying the sex of the walker at multiple points on the gender continuum (see Figure 6.2).

6.8.1. Method

Participants

Sixteen community-dwelling, older adults also participated in this task (9 female, all right handed, mean age=72.4 years, SD=6.9 years). All had normal cognitive function, as assessed by the MoCA (Nasreddine et al., 2005) (i.e. all scored ≥ 24 of 30). All older participants had normal visual acuity (mean binocular LogMar score=0.09, SD= 0.13 and LogMar score= 0.08, SD=0.09 at 4m and 40 cm respectively) and contrast sensitivity (mean score=1.97, SD=0.08, assessed using the Pelli-Robson Contrast Sensitivity Test) for their age. All participants wore appropriate corrective lenses, if needed, at the time of testing. In addition, participants were asked to subjectively rate their hearing by answering the question: "Is your hearing (with or without a hearing aid) Excellent/Very Good/Good/Fair/Poor?" All older adults rated their hearing as either 'Good', 'Very Good' or 'Excellent' (see the Participants section of Experiment 7, Chapter 5, for a full description of this self-report hearing measure). With the exception of 4 older adults who had participated in Experiment 11 2 months previously, no individual had participated in any of the previous experiments.

Stimuli and Apparatus

The visual and the auditory stimuli are as described in Experiment 12A and 13, respectively. As the auditory effects were shown to be irrespective of the sentence content, and for consistency with the stimuli used in Chapter 5 for the older adult group, the semantically

neutral sentences 'This is my house' and 'The window is closed' were used as auditory stimuli in this experiment. Furthermore, the experiment was presented on a Dell OptiPlex 755 PC and displayed on a HP L1710 17" LCD colour monitor. All other details remain as described in Experiment 12A.

Design

The design is as described in Experiment 12A.

Procedure

The procedure is as described in Experiment 12A.

Analysis

The analysis is as described in Experiment 12A.

6.8.2. Results

The proportion of 'male' responses for gender-ambiguous walkers were compared to chance performance (i.e. 0.5) for vision only, audio-visual female, and audio-visual male pairings using a series of one sample *t*-tests. It was observed that judgements of the sex of the walker did not differ from chance performance in the vision only condition [$t(15)=0.82$, $p=0.42$] (mean proportion male responses=0.52). In addition judgements for AV male pairings [$t(15)=1.07$, $p=0.30$] or AV female pairings [$t(15)=-0.11$, $p=0.91$] did not differ from chance performance (mean proportion 'male' responses=0.55 for male voice pairing; 0.49 for female voice pairing).¹³ On a side note, it was however observed that five of the older adult participants did show a similar pattern of performance to the younger adult group when judging the sex of gender-ambiguous walkers, with or without, a male or female voice. Specifically, while vision only performance remained at chance (mean proportion 'male' responses=0.57), the presence of a female voice modulated the sex judgements in favour of the walker appearing more female (mean proportion male responses=0.34). There was no effect of the male voice (mean proportion 'male' responses=0.5).

¹³ A paired *t*-test confirmed that there was no significant difference between the gender ratings for AV male-voice and AV female-voice pairings for gender-ambiguous PLW [$t(15)=-0.97$, $p=0.35$]. This suggests that the sex of the voice cannot be used to disambiguate the sex of walker in the older adult group.

Results also indicated that sex judgements for male [$t(15)=-0.77$, $p=0.44$] and female [$t(15)=-0.12$, $p=0.90$] walkers did not differ between vision only and AV conditions. The mean proportion of 'male' responses in the female vision only and female AV conditions were 0.19 and 0.20, respectively. The mean proportion of 'male' responses for male vision only and AV conditions were 0.90 and 0.90, respectively. In addition, there was no difference between sex judgements for exaggerated female walkers in the vision only and AV condition [$t(15)=-0.20$, $p=0.84$]. Nor was there any difference between the proportion of 'male' responses for exaggerated male walkers in the vision only or AV condition male [$t(15)=-0.07$, $p=0.94$]. The mean proportion of 'male' responses for exaggerated female vision only and exaggerated female AV conditions were 0.11 and 0.11, respectively. The mean proportion of male responses for exaggerated male vision only and exaggerated male AV conditions were 0.89 and 0.90, respectively. All in all, the results from this experiment suggest that older adults could successfully perceive the sex of the walker from the visual displays and that when visual information was reliable, there was little effect of the voice information on the responses.

6.8.3. Discussion

In marked contrast to the results observed for the younger adult groups described in Experiment 12A, 12B and 13, older adults' judgements of the sex of gender-ambiguous walkers appeared to be unaffected by the presence of either a male or a female voice. Specifically, while older adults' performance was at chance when classifying these walker types, indicating a similar point of perceptual ambiguity to the younger adult group, auditory sex cues did not modulate the perceived sex of the walker in the older adult group. This is an unexpected result, as it was predicted that older adults would benefit from the addition of auditory cues in a similar way to younger adults, particularly when the visual walker could not be reliably classified as male or female. This result seems to contradict previous findings of enhanced multisensory integration in older adults (e.g. Laurienti, Burdette, Maldjian, & Wallace, 2006; Maguinness, Setti, Burke, Kenny, & Newell, 2011; Setti, Burke, Kenny, & Newell, 2013). However, it is possible that this result may reflect older adults' poorer ability to represent multiple sensory cues, when the spatiotemporal relationship between the sensory signals is altered. Indeed, it is plausible that older adults did not associate these two cues with the same event (see also results from Chapter 5, Experiment 8), this explanation will be returned to in more detail in the General Discussion.

6.9. EXPERIMENT 15A

6.9.1. Sex classification in developmental prosopagnosia

As mentioned previously, developmental prosopagnosia (DP) is characterised by an inability to identify an individual from the face alone. However, other aspects of face processing have been shown to be largely intact in this cohort. For example, a number of studies have demonstrated that DPs can reliably detect the sex of a face, in spite of their poor ability to represent facial identity (DeGutis, Chatterjee, Mercado, & Nakayama, 2012; Le Grand et al., 2006; Nunn et al., 2001). This suggests that perceptual abilities in this cohort remain sensitive to the shape cues in the face which are associated with male and female faces. Moreover, DeGutis, Chatterjee, Mercado and Nakayama (2012) demonstrated that individuals with DP may use configural processing for sex classification, in a similar way to neurotypicals. In that study, DeGutis and colleagues (2012) observed a similar decrease in accuracy for categorising the sex of a face when it was inverted, relative to upright, for both DPs and controls. However, it remains unclear whether the ability to classify the sex of an individual from other sensory cues such as the voice, or from biological motion displays, remains intact in this cohort.

As highlighted, in Chapter 5 a number of studies have reported that individuals with DP rely on the voice as a cue to identity. However, only a small number of studies have tested this assertion (e.g. Schall & von Kriegstein, 2014; von Kriegstein et al., 2006). The results from these studies, along with the results from Chapter 5, suggest that speaker recognition and speaker matching for identity may be, to some extent, impaired in this population. However, whether the processing of other paralinguistic cues, for example detecting the sex of a speaker, is also impaired remains unexplored. It has been suggested that the voice contains ‘static’ properties i.e. where there is a change along one dimension such as pitch, as well as ‘dynamic’ properties, for example, changes in the temporal pattern of the speech signal (Belin et al., 2004). As such, it is possible that in line with the preserved ability to use shape cues for sex classification from the face, that the ability to detect the sex of a voice may also be intact, owing to the fact that this sexually dimorphic cue largely arises due to changes along one dimension in the speech signal. This is plausible when we consider that it is the dynamic signature, or the subtle variation across a range of acoustic parameters, which may be poorly represented in some cases of DP, possibly due to the poor ability to process faces at the *individual* level (see results from Chapter 5).

In addition, a number of studies have also reported that developmental prosopagnosics may rely on differences in body shape or gait to recognise familiar others (e.g. Duchaine, Yovel,

Butterworth, & Nakayama, 2006; Grueter et al., 2007; Kennerknecht, Ho, & Wong, 2008; Nunn et al., 2001). This would imply that body perception, and possibly biological motion processing, is intact in this cohort. However, again very few studies have explicitly tested this assumption (Duchaine et al., 2006; Lange et al., 2009; Righart & de Gelder, 2007). Moreover, the studies which have tested this have given rise to conflicting findings. For example, Righart and de Gelder (2007) reported that DP may be associated with poor configural processing of both face and bodies. Specifically, they observed that compared to controls, DP participants exhibited an abnormal N170 ERP response component when viewing static images of faces and bodies. However, Duchaine et al. (2006) reported normal matching of bodies, even across changes in viewpoint in cases of DP. One study which explicitly examined biological motion processing in DP, noted that prosopagnosics were poorer at discriminating the direction and orientation of point-light walkers than controls (Lange et al., 2009). This led the authors to conclude that DP may result from a deficit in a 'general mechanism' which governs the processing of both faces *and* bodies. Indeed, Righart & de Gelder (2007) suggested a similar argument and surmised that a general deficit in configural processing may underpin the face processing deficits observed in DP. Importantly, neither study investigated the ability of individuals with DP to extract higher-order features, such as the identity or the sex of a person, from biological motion cues. As highlighted by DeGutis et al. (2012), DPs appear to process the sex, rather than the identity, of a face in a similar configural manner to neurotypical controls. Thus it remains plausible that the ability to classify sex from biological motion displays may also remain intact in DP. Experiment 15A examined whether UM and PL could classify the sex of an individual in a series of point-light displays using the same paradigm and the same older adult participant data for comparative analysis, to that reported in Experiment 11. In Experiment 15B, UM also participated in the audio-visual version of the experiment completed by the older adult group, i.e. Experiment 14. Again the same older adult participant data was used for comparative performance.

6.9.2. Method

Procedure

The procedure is as described in Experiment 12A.

Stimuli and Apparatus

The stimuli and apparatus are as described in Experiment 13.

6.9.3. Analysis and Results

See Figure 6.6 for a plot illustrating UM and PL's individual slopes, in comparison to the age-matched control mean slope, for judging the sex of the walker as a function of the position of each PLW on the gender continuum (note the control data is the same data as that is plotted in Figure 6.2). Individual slopes were calculated for UM and PLs performance (see Experiment 11 for procedure). This slope was used as a measure of sensitivity in detecting the sex of the walker and was compared to performance from the age-matched control group using a modified *t*-test described by Crawford, Garthwaite and Porter (2010). As highlighted previously, the test is designed to compare whether a single case score is significantly different from a control or normative sample. The analysis revealed that both UM [$t(12)=1.38, p=0.10$] and PL [$t(12)=0.28, p=0.39$] showed the same sensitivity in classifying the sex of the walkers, as their age-matched control group.

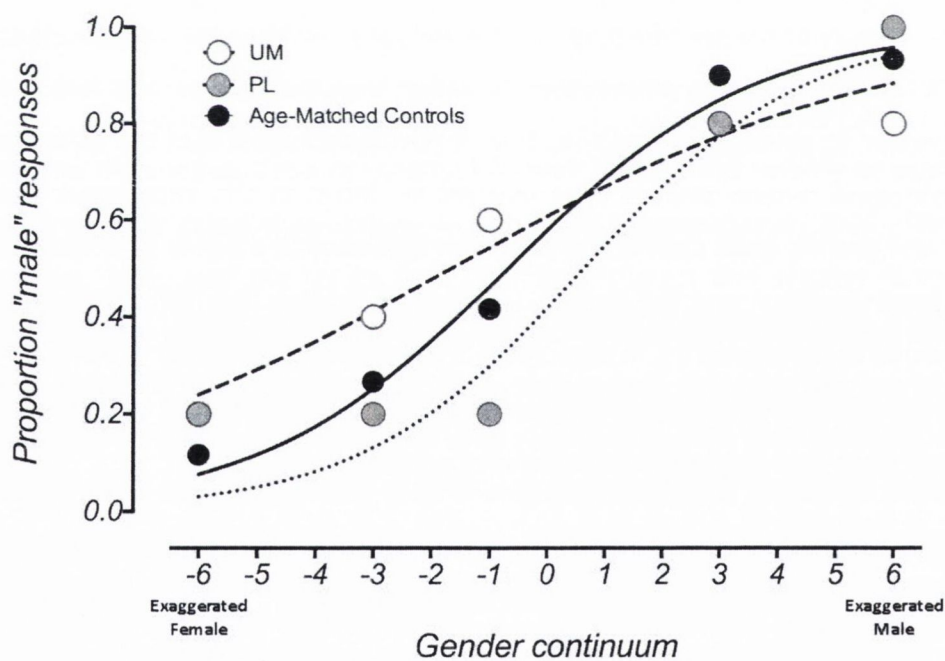


Figure 6.6. Plot showing the individual slopes for UM and PL and the mean slope for age-matched controls as a function of the gender continuum.

6.10. EXPERIMENT 15B

6.10.1. Method

Procedure

The procedure is as described in Experiment 14.

Stimuli and Apparatus

The stimuli and apparatus are as described in Experiment 14.

6.10.2. Analysis and Results

The proportion of 'male' responses reported by UM for gender-ambiguous walkers for vision only, AV female-voice and AV male-voice trials were 0.6, 0.1, 1.0 respectively. As such from looking at these data it can be seen that UM's sex classification for the visual walker performance was strongly influenced by the presence of either a female or a male voice. This pattern was not evident for her older adult counterparts. UM's performance is plotted in Figure 6.7(a) in relation to the data from older adult age-matched controls ($n=16$), taken from Experiment 14. UM's performance on these three conditions was compared to age-matched controls using the modified t -tests described by Crawford, Garthwaite, & Porter (2010). These analyses revealed the proportion of 'male' responses reported by age-matched controls and UM did not differ for the vision only condition [$t(15)=0.54$, $p=0.30$]. However, there was a difference between UM and age-matched controls proportion of 'male' responses to the AV female [$t(15)=-1.82$, $p=0.04$, $Z_{cc}=-1.87$, 95% CI= -2.69 , -1.04] and AV male [$t(15)=2.35$, $p=0.02$, $Z_{cc}=2.42$, 95% CI= 1.42 , 3.40] conditions. As shown in Figure 6.7(a), the proportion of 'male' responses given by UM reduced when the walker was paired with a female voice and increased when the walker was paired with a male voice. This suggests a strong bias in the response to the visual PLW from the presence of the voice information.

As noted in the results section of Experiment 14, five older adults did show an effect of the voice on their sex classification performance, in a similar vein to young adults' performance. Therefore UM's performance was also compared to the five individuals using the Crawford modified t test. The analysis revealed that the proportion of 'male' responses reported for UM to the vision only [$t(4)=0.20$, $p=0.42$] and AV female [$t(4)=-1.45$, $p=0.11$] conditions was similar to these five older adult controls. However, there was a trend for a difference in UM and the five

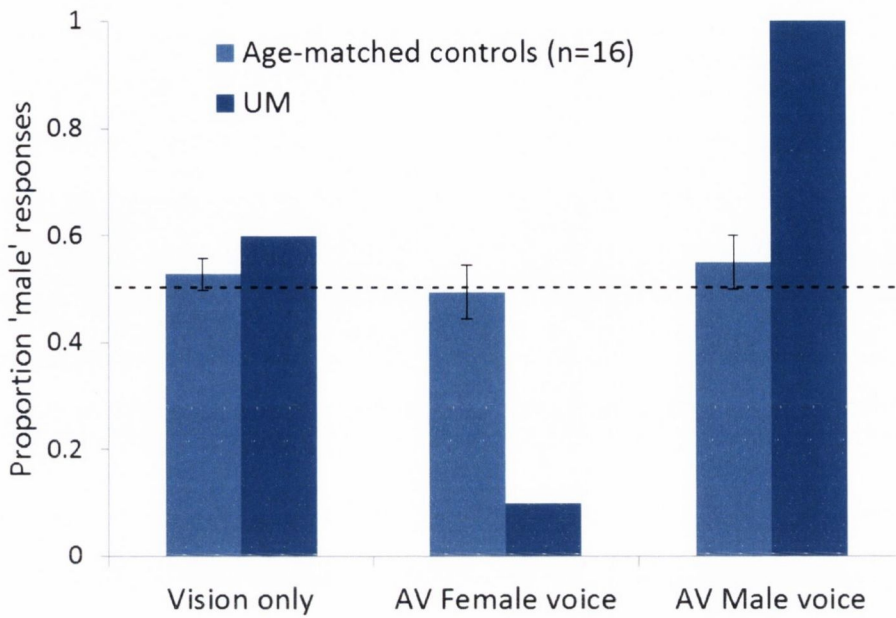
age-matched controls for the proportion of 'male' responses reported when the walker was paired with a male voice [$t(4)=1.63$, $p=0.08$, $Z_{cc}=1.78$, 95% CI=0.28, 3.2]. See Figure 6.7(b), over page.

UM's sex classification performance was not affected by the presence of a female or male voice when the visual displays were of either exaggerated male and exaggerated female walkers. Here she demonstrated perfect performance when classifying the sex of these walkers for vision only and AV conditions (i.e. the proportion of male responses for exaggerated male walkers was 1 and for exaggerated females was 0 for both conditions). The same pattern of performance was also observed for male walkers for the vision-only and AV male voice conditions (proportion male responses=1). However, the voice pairings did affect UM's performance when classifying the sex of female walkers. Here her responses were more accurate when the walker was presented with a female voice (proportion 'male' responses 0), than when it was presented in the vision only condition (proportion 'male' responses= 0.06).

6.10.3. Discussion

The results indicated that both UM and PL showed a preserved ability to detect the sex of a walker from moving point-light displays. This finding is in line with the face processing literature which has demonstrated that individuals with DP can reliably detect the sex of a face, in spite of their poor ability to represent facial identity (Le Grand et al., 2006; Nunn et al., 2001). Moreover, it also argues against a *general* deficit in biological motion processing in this cohort which has been argued by Lange et al. (2009). Rather the results suggest that the extraction of higher-order features from biological motion, such as the sex or gender of an individual, may remain preserved in DP. Moreover, UM's performance profile for the classification of the sex of the walker, appeared to be modulated by the presence of auditory sex cues in the voice. Thus in marked contrast to the results from Chapter 5, UM appears to be able to represent and detect changes in the pitch of the voice to support sex classification when visual information is unreliable. As such, it appears that UM may have a deficit in representing a unique pattern in the voice to support recognition, rather than changes along one dimension, in this case pitch (f_0).

a)



b)

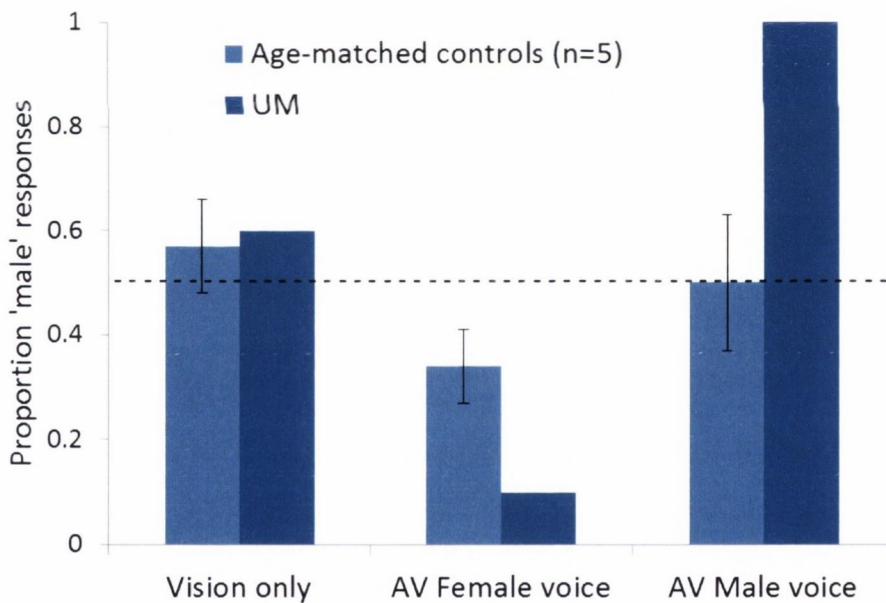


Figure 6.7. Plot showing the proportion of male responses reported for gender-ambiguous walkers, when presented with no voice, a female voice and male voice for UM and the age-matched controls. The results are shown separately for a) UM's performance in relation to *all* older adult controls (n=16) and b) UM's performance in relation to the 5 age-matched older adults who showed an effect of AV presentation. The dashed line indicates chance performance. Error bars represent +1 SEM.

6.11. General Discussion

The aim of the study reported in this Chapter was to investigate whether sex cues present in the voice could modulate the perceived sex of biological motion displays. Such interactions would suggest that information from the voice and the body may *interact* in the processing of higher-order features such as sex or gender. It was predicted that these interactions may be more pronounced when there was some level of ambiguity in the visual signal (i.e. 'gender-ambiguous' walkers), which could be modulated by the presence of an additional sensory cue. This was examined across three cohorts; younger adults, older adults and developmental prosopagnosics. Here participants were presented with gender-ambiguous walkers with male and female voice pairings and participants were required to categorise the sex, or to judge the gender, of the visual walker. It was observed that older adults were overall poorer at classifying the sex of a walker, than their younger adult counterparts (Experiment 11). However, performance for both groups was at chance when classifying the sex of gender-ambiguous walkers. In addition, both UM and PL demonstrated a relatively preserved ability to discriminate the sex of walker, depicted in point-light displays (Experiment 15A). The results for Experiment 12A and 12B suggested that the presence of a voice could modulate the perceived sex of these perceptually ambiguous walkers in the younger adult group and that this effect could not be readily explained due to observers responding to the auditory cue alone (Experiment 12B). Moreover, the results from Experiment 13 suggested that it was likely the paralinguistic cues pertaining to the sex of the voice, which may have modulated the perceived gender of the walker, rather than linguistic aspects of the speech signal. These cross-modal effects were most reliably observed when gender-ambiguous walkers were paired with female voices. However, in marked contrast to the younger adults' performance, the older adults' classification of gender-ambiguous walkers did not appear to be modulated by the presence of a male or female voice (i.e. performance did not differ from chance, Experiment 14). Interestingly, UM's sex classification for gender-ambiguous walkers appeared to be strongly influenced by the presence of both male and female voices, suggesting that she could reliably process sex cues in the voice (Experiment 15B).

The finding that auditory sex cues can modulate younger adults' visual perception of gender-ambiguous walkers is consistent with the findings observed by van der Zwan et al. (2009). In their study, observers judged gender-ambiguous walkers to be more female in appearance when they were paired with a female sounding walking sequence (i.e. female footsteps), compared to when there was no auditory cue present. Interestingly, they observed that this effect did not occur when the auditory sequence was gender-ambiguous, leading the authors to

conclude that the observed effects were likely to relate to interactions between meaningful gender related auditory cues, rather than to the mere presence of an additional auditory signal. The current study demonstrates that these meaningful auditory cues can also be carried by the voice. Moreover, the audio-visual conditions in van der Zwan and colleagues study always contained the pairing of gender-ambiguous with the same female walking sequence. Here, it was observed that voices, particularly female voices, could reliably modulate the perceived sex and gender of the gender-ambiguous walkers across different speakers and across different speech utterances. Moreover, the effects in the current study could not be readily explained by the observer attending to the auditory cue alone. Specifically, incongruent audio-visual pairings for perceptually salient walkers (exaggerated versions of male and female walks) did not affect observed sex judgements in this condition. This suggests that younger adults were not solely attending to the auditory cue when arriving at the decision as to whether the walker was more male or female in visual appearance.

van der Zwan et al. (2009) noted that adaptation to these gender-ambiguous walkers with female footsteps gave rise to significant male after-effects, for subsequently viewed walkers. This provided compelling evidence that this auditory modulation likely reflected perceptual, rather than cognitive effects. However, it is important to note that the audio-visual pairings in that study were always temporally synchronous, such that the sound of the footstep occurred at the same time as the footfall. This temporal synchrony is likely to have played a role in these adaptation effects. In terms of the current findings, it would be speculative to suggest that these effects were purely perceptual and occurring at early stages of processing, an idea which will be returned to later in the discussion.

Previous studies which have examined how the voice may modulate gender processing, have largely focused on cross-modal interactions between the face and the voice (Joassin, Maurage, & Campanella, 2011; Little et al., 2013; Smith et al., 2007). For example, Smith et al. (2007) observed that pure tones in the fundamental frequency range of male or female voices, modulated the perceived gender of androgynous, i.e. gender-ambiguous, face images. Joassin, Maurage, & Campanella (2011) extended these findings using natural speech sounds and face images. In that study, they observed that participants were faster to categorise the sex of a face when both the face and voice were presented, rather than when each gender cue was presented in isolation. The current study extends these findings to demonstrate that these cross-modal effects are not limited to face and voice pairings. Rather the findings from this Chapter suggest that in situations of perceptual uncertainty younger observers may draw on multiple sensory cues to disambiguate the visual percept, which in this case was a gender-ambiguous walker.

In contrast to the performance of the young adult, the presence of the voice did not appear to modulate the perception of gender-ambiguous point-light walkers in the older adult group. There are a number of reasons why this may have been observed. For example, it is possible that categorising the sex of a voice may decline with older age, which may weaken the effect of the auditory sex cues in this task (but see Schwartz & Chatterjee, 2012). For example, it is conceivable that older adults may have less discrete representations for male and female voices, in a similar vein to the broadly tuned or overlapping representations of identity which have been observed in the face perception literature for this cohort (Lee et al., 2011, 2014). Indeed, recent evidence suggests that higher-level auditory features such as the sex or identity of a voice are represented in a similar manner to the face in the brain (Schweinberger et al., 2008; Zäske, Schweinberger, & Kawahara, 2010; Zäske & Schweinberger, 2011). For example, Schweinberger et al. (2008) observed that when participants were adapted to male voices, subsequent voices were perceived to be more feminine, paralleling the well documented after-effects in the face processing literature (Leopold et al., 2005; Leopold, Toole, Vetter, & Blanz, 2001; Rhodes & Jaquet, 2011). These more broadly tuned representations may mean that older adults' perception is less modulated by these additional sensory inputs.

An additional explanation may relate to the change over time information present in the visual and the auditory sensory cues in this study. For example, the results from Chapter 5 demonstrated that older adults may rely more on redundant cross-modal spatio-temporal cues when representing speaker identity. Importantly, in the current study voice information bore no direct temporal correspondence to the dynamic visual display, which may have affected performance. Indeed, other studies which have investigated speech comprehension or speech detection, have demonstrated that temporal information provided by hand movements is tightly bound to the temporal information in the voice, referred to as 'beat gesture' (Hubbard, Wilson, Callan, & Dapretto, 2009; review see Goldin-Meadow & Alibali, 2013). The redundancy of the visual information provided by the 'beat gesture' has been shown to enhance the representation of the speech signal. For example, Hubbard et al. (2009) demonstrated enhanced functional activation in the superior temporal sulcus (STS) when speech was presented with beat gestures displayed by the upper body limb movements, compared to when the speech signal was presented on its own. This finding mirrors results reported by Calvert, Campbell, & Brammer (2000) for facial speech information. It is therefore possible that older adults may have more readily associated the two sensory cues when there was some level of cross-modal redundancy between the signals.

6.11.1. Sex classification in developmental prosopagnosia

The results from this study suggest that biological motion processing may remain, to some extent, intact in DP. Specifically, it was observed that both UM *and* PL could reliably detect the sex of a walker when it was presented without additional vocal cues, demonstrating a preserved ability to discriminate sex cues in moving biological motion displays. While studies investigating deficits in body or biological motion in DP are sparse (Duchaine et al., 2006; Lange et al., 2009; Moro et al., 2012; Righart & de Gelder, 2007), there has been an argument put forward that DP may reflect a general deficit in the ability to represent both faces and bodies. This deficit has been argued to be mediated by a common cause, such as atypical configural processing (Lange et al., 2009; Righart & de Gelder, 2007). However, the results from this study argue that biological motion processing, at least for the classification of sex, may remain intact in this cohort. It is possible that this preserved processing of shape (Duchaine et al., 2006) and motion cues from the body may offer an insight into biological motion processing in the neurotypical brain. Recently, Furl et al. (2014) demonstrated that network interactions between facial form areas, specifically OFA, and motion sensitive area V5, may explain how sensitivity to facial motion may arise in STS (discussed in detail in Chapter 3). The authors also speculated that sensitivity to biological motion in STS might arise due to similar interactions between fusiform and extrastriate body areas and V5, such that form areas may gate the transmission of motion information to STS. With that in mind, it is conceivable that dissociations may arise within DP, such that if the processing of body form is typical, that normal processing of biological motion will also follow (but see e.g. Gilaie-Dotan, Bentin, Harel, Rees, & Saygin, 2011; Vangeneugden, Peelen, Tadin, & Battelli, 2014, for dissociations between body form and body motion processing). However, whether DP is also characterised by a sparing of the ability to detect the identity, rather than the sex, of an individual from point-light walker displays would be an important factor to consider before surmising that biological motion processing may remain unaffected in this cohort.

Interestingly, UM's judgements of the sex of gender-ambiguous walkers were also affected by the presence of the voice. Specifically, it was observed that UM tended to more often judge perceptually ambiguous walkers to be female in appearance, when they were accompanied by a female voice, and vice versa for male voice pairings. Thus in contrast to the findings from Chapter 5, UM appears to be able to discriminate changes in the pitch of the voice to support sex classification. Interestingly, this preserved ability is consistent with many of the findings in the developmental prosopagnosia literature which have largely observed a preserved ability to detect the sex of a face in this cohort (Le Grand et al., 2006; Nunn et al., 2001). Moreover, this preserved

ability to detect the sex of an individual has also been documented in developmental phonagnosia (Garrido et al., 2009). Specifically, Garrido et al., (2009), noted that developmental phonagnosic, KH, could reliably detect the sex of a voice in spite of her poor ability to both discriminate between and to recognise individual speaker identities. As highlighted throughout, determining the sex of a voice is largely achieved by detecting differences in the pitch of the voice. In contrast, representing speaker identity is likely to be mediated by the ability to detect differences across a range of acoustic features for example, pitch, timbre in addition to the dynamic temporal pattern of the voice (Belin et al., 2004; Kamachi, Hill, Lander, & Vatikiotis-Bateson, 2003; Lander et al., 2007; Schweinberger et al., 2014). As such, it appears that the deficits in speaker processing in developmental phonagnosia likely mirror deficits in face processing, lending favour to the view that these cross-modal cues may be represented in a similar manner in the brain (Belin et al., 2004).

It is possible to speculate the neural underpinnings of the effects observed in this Chapter, based on previous work which has investigated how the voice and the face may be integrated for the purpose of detecting gender. For example, Joassin, Maurage, & Campanella (2011) observed that such judgements engaged supramodal parietal and frontal processing areas, as well as visual and auditory areas associated with face and voice processing (Joassin, Maurage, et al., 2011). Importantly, Joassin and colleagues used *static* images of faces and natural voice recordings to establish these gender networks in the brain. Thus similar to the current study there was no direct spatiotemporal relationship between the two cues. They argue that this supramodal processing reflected the cross-modal attentional processes required to integrate the face and the voice for the purpose of detecting gender. It is possible that a similar network, but one which may possibly also engage STS owing to the dynamic nature of the body stimuli (see also Troje et al., 2006), may explain the cross-modal effects in the current study.

Taken together, the findings from this Chapter suggest that information from the voice and the body may interact for the processing of sex and gender. While these findings are promising, further studies, which for example manipulate the temporal synchrony between the voice and the body, may help to understand whether these cues interact at early or late stages of processing in the brain.

Chapter 7

General Discussion

7. General Discussion

7.1. Introduction

To date, a large body of research using static images of faces has informed our understanding of how identity is perceived, processed and represented in the human brain (Bruce & Young, 1986; Kanwisher, McDermott, & Chun, 1997; Valentine, 1991). However, as highlighted throughout the thesis, human interactions are rarely based on encounters with static faces, rather people move and speak and thus our perceptual system is exposed to a range of sensory inputs at any *one* time (e.g. facial motion, body motion and the voice). Indeed, these multiple sensory cues are often tightly bound to one another in terms of space and time. Evidence from younger adult studies suggests that these sensory cues may interact in person perception (O'Toole, Roark, & Abdi, 2002; Schweinberger, Robertson, & Kaufmann, 2007; Thornton & Kourtzi, 2002; van der Zwan et al., 2009; von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005). However, a benefit on perception has not always been consistent for the combination of these sensory cues. This may be due to the fact that younger adults may be able to reliably represent identity based on input from one sensory signal alone, or that these interactions may be more readily observed in this cohort for already established, i.e. familiar, representations. The aim of the thesis was to extend our understanding of how these sensory cues may interact in person perception through the study of three cohorts: younger adults, older adults and developmental prosopagnosics (DP). Importantly, both ageing and DP are characterised (to different extents and with different etiologies) by poor processing of static facial identity cues. This thesis argued that studying these cohorts may offer a unique opportunity to examine how and at what stages these sensory cues may interact to support normal person perception.

This concluding chapter will summarise the key findings from the collection of studies reported in the previous chapters, which focused on three main themes: 1) the interaction of form and motion cues in face perception; 2) the interaction of facial motion and the voice in person perception; and 3) the interaction between dynamic body cues and the voice in person perception. These findings will be discussed and placed within the framework of existing and emerging theoretical models of person perception and recognition, which were introduced in the opening chapter. Throughout this chapter future directions and practical implications arising from the current findings will also be discussed. Finally some of the limitations of the current study will be discussed, with the hope that these limitations may assist in informing future research.

7.1.1. *Summary of the main findings*

In Chapter 2, static face perception and recognition in ageing and developmental prosopagnosia was examined, using a series of standardised face processing tests. The results demonstrated that older age was associated with a decline in both face perception and face recognition. Moreover, consistent with previous findings in the literature, e.g. Habak, Wilkinson, & Wilson, (2008) and Lee, Grady, Habak, Wilson, & Moscovitch (2011), the effect of age was most evident when the task was to match the face across changes in visual appearance. Interestingly, performance across the range of face processing tests was not associated with measures of general cognitive ability, suggesting that age-related changes in face processing are unlikely to be mediated by general changes in cognitive function associated with the ageing process (Hildebrandt et al., 2011). However, measures of face perception performance were strongly associated with measures of face recognition performance in the older adult group. This suggests that poor structural encoding of facial identity at early stages of processing may affect the quality of the representation of the face in memory, affecting recognition performance at later stages of recall. Furthermore, the results from Chapter 2 also illustrated that it was possible to accurately detect and characterise disorders of face processing even into older age, when using age-matched older adults as control norms. Indeed, the series of tests revealed that UM and PL were poorer than age-matched controls in both perceiving *and* recognising facial identity. As such, these results suggest that both UM and PL may have apperceptive developmental prosopagnosia (Németh et al., 2014; Towler & Eimer, 2012; Towler et al., 2012). Moreover, the tests also confirmed that both developmental prosopagnosics recognition deficits were largely restricted to face, compared to object, processing.

Chapter 3 examined the role of motion on unfamiliar facial identity matching performance in younger and older adults. Specifically, the series of experiments described in Chapter 3 were designed to address if learning an unfamiliar face in motion, compared to static presentation, could enhance the representation of facial identity in the older adult group. Indeed, this was observed and the findings demonstrated that the age-related cost in accuracy for matching identities across changes in the visual appearance of the face (as described in Chapter 2), could be reduced when the face had been learned moving in both a rigid (i.e. head rotations) or a non-rigid (i.e. expression) manner. In addition, there was evidence to suggest that younger adults were also sensitive to the change over time information present in the rigid motion sequence. Specifically, disrupting the spatiotemporal alignment of the face images during learning (i.e. jumbling the images so the face did not turn fluidly from the front to the side) negatively affected younger adults' matching performance. This was evidenced by an increase in reaction

times taken to match novel, relative to learned, images of the face in this condition. Interestingly, the motion (non-rigid) advantage evident for the older adult group did not persist when the face stimuli were inverted. Moreover, under inverted viewing conditions both younger and older adults were equally poor at matching novel, relative to learned images of the face. This suggests that the ability to use motion cues for the purpose of face perception is underpinned by a motion encoding system which is tuned to upright faces and is sensitive to configural or holistic face processing. Moreover, the findings suggest that form and motion cues likely interact during the encoding of facial identity and can, as a consequence, enhance the representation of facial identity.

Chapter 4 examined how facial motion may interact with facial form processing in developmental prosopagnosia. To date, studies on face processing in this cohort have been largely constrained to the use of static face images. In line with Chapter 3, both UM and PL learned a series of unfamiliar facial identities moving in both a rigid (i.e. head rotations) or a non-rigid (i.e. expression) manner, or from a sequence of static images extracted from the motion sequences. The results demonstrated that both UM and PL could match facial identity across changes in view when the face was learned in rigid motion. Indeed, their performance profile was similar to age-matched controls in this condition. However, learning a face in non-rigid motion (i.e. expressive movements) significantly interfered with *identity* matching across changes in expression in this cohort. This finding suggests that developmental prosopagnosics may find it difficult to recover facial form cues, when the face has been learned in motion. It is likely that motion interferes with the use of suboptimal identity matching strategies engaged by DP, e.g. direct feature matching. In turn, this suggests that interactions between non-rigid motion and facial form in normal face processing may be guided by the same underlying face processing mechanisms, i.e. holistic or configural processing. In contrast, both UM and PL could reliably match the *expression* of the face across changes in identities, even when the facial expression was learned in motion. These findings suggest that it is the integration of form and motion to support identity, rather than social signal processing, which may be compromised in disorders of face processing. This lends support to the findings from Chapter 3 and argues that interactions between form and motion for identity processing may be mandatory.

Chapter 5 investigated the role of dynamic facial cues on the representation of speaker identity in younger and older adults and DP. Recent evidence suggests that the voice may be represented in a similar manner to the face (Latinus & Belin, 2011; Zäske et al., 2010) and that learning a voice with its dynamic face counterpart can enhance subsequent speaker recognition (von Kriegstein & Giraud, 2006). Using a series of unfamiliar speaker identity matching tasks, the

findings from this Chapter revealed that older adults were poorer at matching speaker identities from vocal cues alone. Interestingly, this pattern was exaggerated in one case of DP. As such, it is possible that atypical face processing may directly impact on how the voice is processed to support identity, leading support for mandatory interactions between these sensory cues. Moreover, in line with previous findings for familiar speaker recognition (Schweinberger et al., 2007), the results also revealed that older adults', and to some extent younger adults', ability to represent an unfamiliar face and the voice as a single identity percept was likely facilitated by the presence of *dynamic* rather than static face cues during learning. As such, interactions between the face and the voice appear to occur at early stages of processing, during the encoding of face and voice identity.

Chapter 6 turned to examine how other socially relevant cues, specifically gender or sex classification, may be supported or mediated by cross-modal interactions. Here it was investigated if sexually dimorphic cues in the voice may influence the perception of a series of 'gender-ambiguous' point-light walker displays. The results demonstrated that auditory sex cues modulated the perceived sex of these gender-ambiguous walkers in the younger adult group. These effects were most robust for female voice pairings, such that gender ambiguous walkers were more often judged to be female when presented with an accompanying female voice. These effects were observed across a change in task and sentence content. Moreover, the effects could not be simply attributed to the observer attending to the auditory cue alone. As such, these findings suggest behaviourally relevant interactions between the voice and the body for the purpose of classifying gender. However, the presence of the voice did not modulate the perceived sex of these walkers in the older adult group. It is possible that older adults did not reliably associate the two sensory cues, owing to the fact that the cross-modal spatiotemporal redundancies between the two signals were altered. Interestingly, both UM and PL showed similar performance to their age-matched controls when judging the sex of a walker from these dynamic point-light displays. Moreover, UM's performance for classifying the sex of a walker appeared to be modulated in favour of the sex of the concurrently presented voice. This argues that biological motion processing, at least for the purpose of discriminating the sex of an individual, may be relatively intact in DP.

7.2. Implications of the findings

7.2.1. *Integrating the face, voice and body into models of person perception*

As highlighted in Chapter 1, seminal models of face processing such as that proposed by Bruce and Young (1986) (also see Burton et al., 1990) proposed that facial form (i.e. the unchangeable aspects of the face) and facial motion provide *discrete* sources of information. Specifically, such models argue that facial identity is represented through facial form cues, while the changeable aspects of the face are processed in parallel to support facial speech and expression analysis. However, as noted in Chapter 1, 3 and 4, a large body of research has demonstrated that facial motion may support familiar face processing, particularly when facial form cues are degraded (e.g. Knight & Johnston, 1997; Lander, Christie, & Bruce, 1999). These findings have led researchers, e.g. O'Toole et al. (2002), to propose that facial motion may act as a 'back up system' for facial identity when visual form cues are unreliable. This view assumes that the observer has over time learned to associate a unique dynamic signature with a facial identity, which is processed independent of the visual form of the face (dorsal based processing).

However, O'Toole et al. (2002) also proposed that these dynamic cues may interact with the processing of facial form, indicating interactions between ventral (e.g. fusiform face area or FFA) and dorsal (i.e. superior temporal sulcus or STS) streams for the purpose of representing facial identity. Yet to date, support for this view has been inconsistent. This thesis argued that one possible reason for this conflicting evidence may be that when perceptual abilities are optimal, such as in young adults, structural cues alone may be sufficient to create a robust representation of facial identity. However, this does not preclude the fact that motion may assist in representing facial form, rather it suggests that these interactions may be more difficult to observe in younger adults' behavioural performance (see also Otsuka et al., 2009 for a related argument). The findings from Chapter 3 support this view. Specifically, they demonstrated that learning a face in motion (rigid and non-rigid), relative to a series of static face images, can enhance the representation of facial identity in older, relative to younger adults. This provides evidence to suggest that facial motion can interact with the processing of facial form during the encoding of new facial identities. These interactions likely render the representation of facial identity more robust.

Moreover, the results from Chapter 4 demonstrated that learning a face in non-rigid motion made it difficult for both UM and PL to recover facial form information, as evidenced by the poor subsequent matching of novel static face images to the learned identity. Again in line

with the results from Chapter 3, this suggests that form and motion cues may influence each other at the encoding stage at least. However, motion likely detracts from encoding facial form in developmental prosopagnosia, while it can enhance, or at least does not detract from, the representation of the face in neurotypical processing. These results suggest that traditional models of face recognition (e.g. Bruce & Young, 1986) must be revised to accommodate a role for interactions between form and motion at early stages of processing. Moreover, the results lend favour to proposed interactions between form and motion processing proposed by O’Toole and colleagues (O’Toole et al., 2014, 2002; O’Toole & Roark, 2011; Roark et al., 2003) and argue against a strict dissociation between static and dynamic face processing (Longmore & Tree, 2013; Pitcher et al., 2014). Moreover, these findings demonstrate that motion is not only relevant for accessing stored face representations, but also in establishing new representations of facial identity (see Figure 7.1).

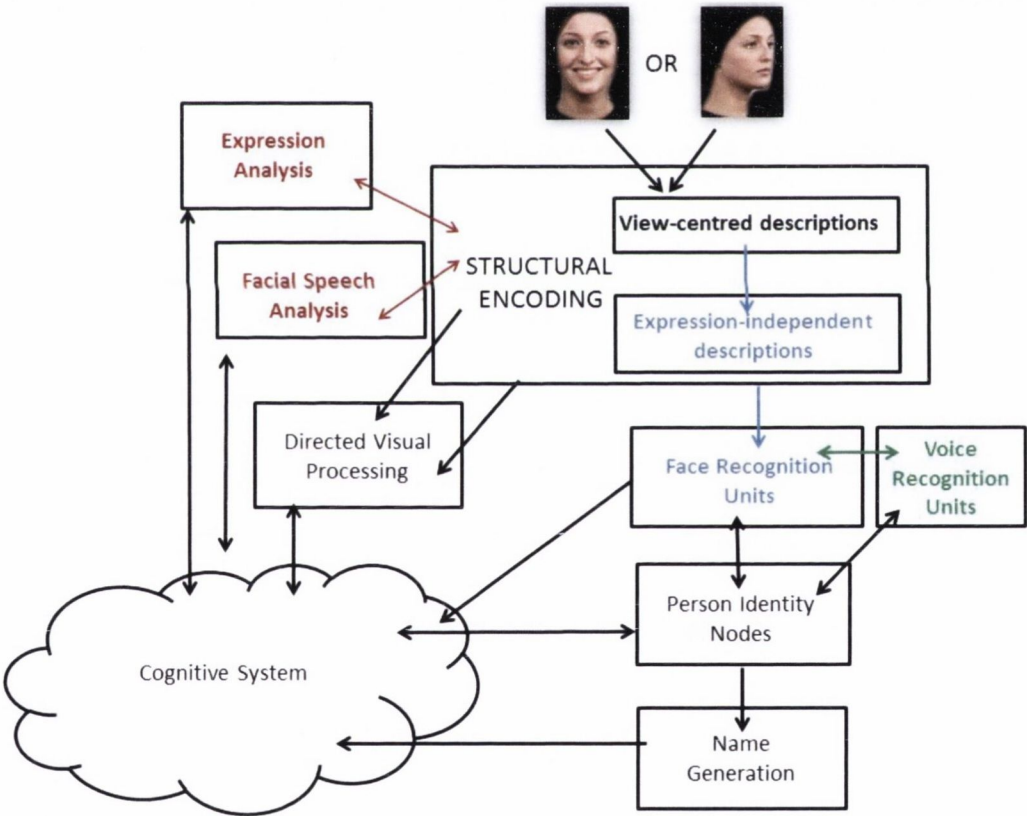


Figure 7.1. A schematic illustration of the Bruce and Young (1986) model, revised in line with the findings from the current thesis. This model is updated to suggest that dynamic facial cues (red text) may interact with the encoding of facial form cues (blue text). In addition, the findings from this thesis propose that the face and voice likely interact at perceptual (green text), rather than exclusively at post-perceptual, stages of processing as previously assumed (see also Burton et al. 1990).

The Bruce and Young (1986) model also proposed that face and the voice were likely to interact only at post-perceptual stages of processing. Indeed, Bruce and Young (1986) stated “a face recognition unit will respond when any view of a person’s *face* is seen, but will not respond at all to his or her voice or name” (p.312). Burton et al. (1990) later followed suit to argue that the face and the voice were likely only to interact at the level of the Person Identity Node (PIN). Here, they argued that units in the PIN could support familiarity decisions across different modalities, such as the voice. Again, in line with the Bruce and Young (1986) model this implies that face and voice undergo extensive *unisensory* processing and then only interact to support recognition at supramodal, i.e. post-perceptual, stages of processing. However, as highlighted in Chapter 1 and 5, recent findings by von Kriegstein et al., (2005) have demonstrated that the fusiform face area (FFA) also responds to familiar voices, in the absence of a facial image. These findings are counter to that predicted by Bruce and Young (1986) or Burton et al. (1990), in fact they suggest that ‘a face recognition unit’ may directly respond to the *voice* of a familiar identity. Furthermore, these effects have been shown to be mediated by direct structural connections between FFA and temporal voice areas (TVA) in the brain (Blank et al., 2011). As such, it appears that interactions between the face and the voice for identity processing may be more pronounced than previously assumed.

Indeed, Schweinberger et al. (2007) observed that familiar voice recognition was enhanced when the voice was presented with its corresponding facial identity, and poorer when the voice was presented with its non-corresponding facial identity. Moreover, these effects were pronounced when the voice was presented with a dynamic, rather than static, face. As such the results from Schweinberger et al. (2007) suggest behaviourally relevant interactions between the dynamic face and the voice. Still, it has been argued that these cross-modal effects are likely to be more pronounced for familiar identities i.e. previously established multisensory representations. Nevertheless, Chapter 5 argued that it is possible that these interactions may be important when establishing representations of previously unfamiliar identities. However, these effects may be more readily observed when the representation of the voice is weaker, such as occurs with the ageing process, or under more challenging ‘listening’ conditions. The results from Chapter 5, lend favour to this argument. Specifically, in the older adult group it was observed that the time taken to match speaker identities across novel speech utterances was slower when the voice had been learned with its static, rather than dynamic, face counterpart. This suggests that older adults could more readily encode the face and the voice as a single identity percept when both identity cues were presented in synchrony. Moreover, it argues that the spatiotemporal redundancies in the visual speech movements play an important role in representing *speaker identity*, i.e. the

paralinguistic aspects of the speech signal. These findings significantly depart from previous assumptions pertaining not only to the idea that the face and voice interact at late stages of processing, but also in terms of what type of information is carried by facial speech movements. For example, Bruce (1988) argued “it is highly unlikely that the information needed to distinguish different mouthed phonemes overlaps with that used to recognise individuals by their faces” (p32). However, other studies have demonstrated that facial identities can be matched from moving speech displays, when no reliable form cues are available (Rosenblum et al., 2002), suggesting that these facial speech cues on their own may be sufficient to support identity processing.

In addition, Chapter 5 reported that younger adults’ face and voice matching performance were equivalent when the face and voice had been learned as a dynamic, rather than static face-voice pairing. In contrast, when the voice had been learned with its static face counterpart, voice matching was poorer than face matching performance. It is thus likely that interactions between the dynamic face and the voice assist in representing the identity of that *person* in memory. Moreover, the observation that UM, a developmental prosopagnosic, was impaired on a task which required matching identity from the voice alone, suggests that poor face processing may impact on voice processing abilities. This implies that throughout development face processing *may* influence and possibly refine voice processing abilities, suggesting mandatory interactions between face and voice processing. Of course, as highlighted in Chapter 5, it is of great importance that UM’s dynamic face and voice matching performance be explicitly tested in the same manner as for the older adult cohort before such a conclusion could be arrived at. Taken together, these findings suggest that spatiotemporal redundancies between the face and the voice may assist in *encoding* new identities and as such these findings support the idea of early, rather than post-perceptual, stage interactions between the face and voice. These findings thus lend favour to emerging models of person perception proposed by von Kriegstein and colleagues and Schweinberger and colleagues, who argue that the face and voice interact to support identity processing at earlier stages of processing than previously assumed (see Figure 7.1).

As highlighted in Chapter 6, emerging research suggests that these face and voice interactions are not only limited to identity processing but that they may also extend to the perception of other higher-order features such as sex or gender (Joassin, Maurage, et al., 2011; Little et al., 2013; Smith et al., 2007). For example, Little et al. (2013) demonstrated that adaptation to a masculinised version of a female voice, made subsequently viewed masculinised images of female faces appear more ‘normal’ and vice versa. The same effects were also observed within each ‘independent’ modality. This led the authors to conclude that representations of

gender are shared across the senses and that auditory gender cues can modulate the perception of visual facial form information and vice versa. Chapter 6 explored whether these audio-visual interactions may also extend to other visual cues which depict the sex of an individual, specifically, biological motion cues displayed by the body. The results from Chapter 6, support this view and revealed that sex cues in the voice affect not only perception of gender in face images, but also gender cues displayed by the body in motion. Here, younger adults judged gender ambiguous walkers to be more female or feminine in appearance when the walkers were presented with a female voice. However, an effect for audio-visual male voice pairings was also observed, although less consistently across the range of tasks reported in Chapter 6. Nevertheless, these findings suggest that paralinguistic information from the voice can interact with the perception of motion cues from the body, such that when visual cues are unreliable the perceptual system may exploit additional sensory cues for the purpose of resolving the visual percept.

However, an issue raised in Chapter 6 pertained to whether these interactions reflected interactions between the senses at early, or later, stages of perceptual processing. Indeed, older adults did not benefit from the accompanying voice in the same way as younger adults, possibly due to the fact that change over time information in the voice bore no direct relationship to the movement of the body. However, it might be possible to resolve some of these reservations through the implementation of adaptation studies. For example, van der Zwan et al. (2009) observed that adaptation to these gender-ambiguous walkers with female footsteps gave rise to significant male after-effects, for subsequently viewed 'gender ambiguous' walkers. In a similar vein it might be possible to adapt observers to female or male vocal stimuli and test the subsequent perception of the gender ambiguous walker stimuli. If it was observed that adaptation to a female voice gave rise to subsequent male after-effects for visual walkers, then it may be concluded that these effects may be governed by interactions within an established 'gender network' which may be involved in perceiving and representing the sex or gender of an individual across sensory modalities (Joassin, Maurage, et al., 2011; Little et al., 2013). Nevertheless, these results lend favour to the idea that sensory cues interact in person perception, and that these interactions are behaviourally relevant. Interestingly, both UM and PL could reliably detect the sex of a walker suggesting that biological motion processing may remain to some extent preserved in DP. One possible explanation put forward in Chapter 6 for this finding was that both prosopagnosics may have intact processing of body form, and as a consequence the perception of biological motion cues from the body may remain unaffected. However, it would be important to examine if other aspects of biological motion perception, such

as the perception of identity also remain intact in these individuals. Moreover, UM could reliably detect whether a speaker was male or female. This was evidenced by the fact that she judged gender ambiguous walkers to be male or female in appearance, when they were accompanied by a male or female voice, respectively. This suggests that her poor processing of voice cues, reported in Chapter 5, may be mediated by a poor ability to represent the voice at an *individual* level, in parallel to her deficit in representing the face.

7.2.2. *Representing faces in the brain: Reforming 'face-space'*

As outlined in the Chapter 1, it is argued that each face we encounter is represented through norm-based coding, or represented in what Valentine (1991) referred to as 'face-space'. Specifically, rather than representing each individual face as an absolute point in face-space, the perceptual system may represent each face we encounter in relation to its deviation from a prototype or typical face. As such, the more 'distinctive' a face is the further it will fall outside of the boundaries of the typical face representation and as a consequence may be more readily recognised. By default, in order to represent a face within face-space the perceptual system must know in what way one face varies in relation to another. However, whether face-space is also concerned with how the face image varies over time, i.e. through changes occurring with facial movement, is relatively unclear. Indeed, to date this multidimensional space has largely been conceptualised in terms of variation in structural form cues, with little consideration for how motion cues may affect this representation.

However, more recent findings highlighting that the recognisability of the face may be modulated by the degree of idiosyncrasy of the *motion* cues displayed (Lander & Chuang, 2005), suggests that facial motion may be represented in a similar norm-based manner to that of facial form (Valentine, 1991). While it is possible that dynamic and static cues may be represented in a similar, yet separate manner, one very recent study by Jeffery, Petrovski, & Rhodes (2014), which revealed identity adaptation after-effects for both moving and static face images, has suggested that there is likely some degree of overlap in how these cues are represented in the brain. While the findings from this thesis do not shed light directly on whether static or dynamic facial cues are represented together in face-space, they do suggest that the interactions between form and motion cues may be relevant for representing each identity within Valentine's originally described face-space model. Specifically, the results from Chapter 3 argue that dynamic cues, particularly non-rigid internal deformations of the face, may enhance (older adults), or at least not detract from (younger adults), the encoding of facial form cues. Thus although the face may still be represented in terms of its static visual appearance in face-space, this does not preclude the idea

that the formation of this representation was enhanced, or at least modulated, by the presence of motion during the learning of that facial identity. See Figure 7.2 for an illustration of a possible integrated face-space. If motion plays an important role in establishing these representations, then it is also possible to speculate that some of the deficits in face processing which are evident in DP, such as impaired norm-based coding, see e.g. Palermo, Riolta, Wilson, & Jeffery (2011), may in part arise from the poor processing of facial motion cues, or at least abnormal interactions *between* form and motion cues. Indeed, O'Toole et al. (2002) suggested the possibility that motion may 'bootstrap' the encoding of invariant facial form cues in typical face processing. Furthermore, it appears that this 'bootstrapping' may be an important part of how normal face processing *develops* (Otsuka et al., 2009). As such, examining if the poor integration of facial form and facial motion is evident in early infancy in DP may assist in refining our understanding of how poor facial form processing may arise in this cohort (see Xiao et al., 2014 for a review on how motion may influence face processing in infancy). This might be possible to observe in families where there is a history of the DP and as such it might be possible to examine face processing in infants born within that family.

In terms of neurotypical processing, it is possible that motion cues may be particularly important when learning faces which are typical in appearance, such that motion may enhance the representation of that face. This idea could be explicitly tested in future studies. For example, it could be examined if learning the identity of faces which are average or typical in appearance could be enhanced when these faces were learned in motion. The learning condition could be manipulated such that the face could be learned through any type of motion profile, or a distinctive facial motion profile and both could be compared to performance for a static learning condition. A benefit for learning the face in motion at later stages of recall (even if the motion is not distinctive) might suggest that motion enhanced the representation of the face in memory. However, if the dynamic enhancement is restricted to a distinctive motion profile it may suggest that face-space may also accommodate other visual dimensions, such as characteristic motion profiles of the face. In addition, distinctive faces could be learned, moving in an average manner, or in a distinctive manner and so forth.

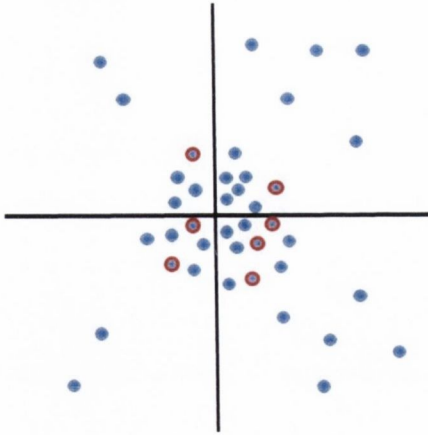


Figure 7.2. An illustration of a possible integrated form and motion face-space. Each blue dot represents a facial identity (form cues), the red outline represents that motion has enhanced the representation of the face. Although speculative, it is possible that motion cues may be particularly beneficial for representing typical faces which are represented in the clustered centre, helping differentiate between identities.

Moreover, it is also possible that these interactions may not be limited to the visual modality. Rather it is possible that, face-space may also accommodate other non-visual dimensions such as the voice of an individual (Bülthoff & Newell, 2015). Indeed, the findings from Chapter 5 argue that there are likely mandatory interactions between the face and the voice when encoding speaker identities. As mentioned in Chapter 1, it has been argued that voice identity is coded in a similar norm-based manner as the face (Zäske et al., 2010). For example, Zäske et al. (2010) noted that adaptation to e.g. voice identity A, significantly biased the perception of a subsequently presented ambiguous speaker identity (a morph between identity A and identity B), away from the adaptor identity. Moreover, they also observed that these effects were evident when the adaptor stimulus was that of a silent moving face, suggesting that cross-modal cues may modulate ‘unisensory’ representations of speaker identity. In a similar vein, to that highlighted above for form and motion interactions in face-space, it is conceivable that, for example, the presence of distinctive dynamic facial cues may affect the representation of the speaker’s identity. In addition, a distinctive voice may affect the representation of facial identity (Bülthoff & Newell, under review). However, it is unlikely that these cross-modal modulatory effects would be entirely equitable. This is largely owing to the fact that the voice is often reported to be a weaker cue to identity than the face (see e.g. Stevenage, Hale, Morgan, & Neil, 2014; Stevenage, Hugill, & Lewis, 2012).

Indeed, it would be of interest to explore these types of interactions between face and voice processing in both developmental prosopagnosia and developmental phonagnosia. The results from Chapter 5, argue that poor face processing may impact on the ability to represent

individual speaker identities in DP, but not the sex of the voice (see Chapter 6). This suggests that interactions between the senses *may* ‘shape’ unisensory processing. For example, dynamic facial speech may bias the observer to attend to particular rhythmic or temporal patterns in the speech signal, which may be relevant for identity processing. While the results from the first documented case report of developmental phonagnosia by Garrido et al. (2009) suggest that face processing is not impaired in this cohort, it is important to consider that only static face images were used in the testing battery. However, it does seem plausible that dynamic face processing may remain unaffected in this cohort. This may arise due to the fact that the face is a salient cue to identity, which may not require as much ‘refinement’ as other sensory modalities. However, it is conceivable that the ability to integrate the face and the voice, even at a perceptual level, may be impaired in this cohort. For example, it might be the case that the presence of a voice (compared to when no voice is presented) during dynamic face presentation, may impair subsequent face matching performance in this cohort.

In addition, these types of interactions could be explored within the normal population. As highlighted in Chapter 1, face processing abilities show individual differences and there is recent evidence reported for ‘super-recognisers’, individuals who exhibit extraordinary face recognition abilities (Russell et al., 2009). Examining speaker perception and recognition in these individuals may provide an insight into whether face processing can modulate voice processing abilities. For example, speaker perception or recognition might also be superior in these individuals, or it may be the case that such persons may rely more heavily on the face, than the voice, for recognition. Moreover, it is possible that future studies may reveal the existence of ‘super-integrators’. Such individuals may show an above average ability to represent the face *and* voice in memory.

7.3. Insights into outstanding issues in the field

7.3.1. *Dynamic enhancements over extended intervals*

The results from Chapter 3 that non-rigid motion *detracted* from the encoding of facial form cues in DP may offer an insight into the findings that facial motion may not facilitate memory for faces in prosopagnosics (Esins et al., 2014; Lander et al., 2004; Longmore & Tree, 2013). For example, while a small number of studies have demonstrated that ‘motion-to-motion’ face matching is relatively intact in this cohort (Longmore & Tree, 2013; Steede, Tree, & Hole, 2007b; see also Lander, Humphreys, & Bruce, 2004 for a similar observation for acquired prosopagnosia), all studies to date (Esins et al., 2014; Lander et al., 2004; Longmore & Tree, 2013),

bar one (Steede et al., 2007a, 2007b), have failed to observe this dynamic enhancement at extended intervals between learning and test. This suggests that individuals with DP may be able to extract and represent idiosyncratic motion cues in working memory. However, these cues are difficult to maintain, across extended intervals between learning and test conditions.

Recently O'Toole & Roark (2011) also argued that the superior temporal sulcus (STS) may possibly represent some 'rough' shape based information about the face, alongside motion cues which can be used to support recognition. Specifically, O'Toole & Roark (2011) proposed that a 'representation enhancement' on face processing may not be *strictly* defined by a subsequent benefit on visual form processing i.e. a motion advantage may not always feedback into ventral based (i.e. visual form) processing. Rather they suggested that there may be a general dynamic advantage which is harboured with STS, such that if a face is learned in motion and subsequently presented in motion at test, an advantage on recognition may be observed. However, the results from Chapter 4 would suggest that establishing these 'rough' shape cues may be impaired when prosopagnosics learn a face in motion, as both UM and PL found it difficult to recover form cues when the face had been learned moving in a non-rigid manner.

Thus it is possible that DPs fail to represent shape cues pertaining to individual facial identity. Although speculative, it is possible that this deficit may underpin some of the conflicting results which have been observed for dynamic face processing in developmental prosopagnosia. Importantly, the studies which have failed to find a benefit on recognition for faces learned in motion in the DP cohort have typically employed the use of unique facial identities, with unique facial form cues, while manipulating the learning condition such that the face can be learned in motion or through static presentation. It is possible that individuals with DP therefore cannot recover the basic shape cues needed to trigger the dynamic representation and as a consequence cannot 'access' the dynamic representation. Indeed, this is conceivable when we consider that the exhibited motion profile, at least for non-rigid motion, is constrained by the shape (i.e. form) of the face itself. If this shape representation is in fact disrupted when the face has been learned in motion, as suggested by the results from Chapter 4, then it is likely that the recognition of these motion cues may be subsequently impaired.

7.3.2. *Expression and identity processing*

To date, there has been substantial debate within the face processing literature as to whether identity and expression are processed by common or separate visual routes (see e.g. Bruce & Young, 1986; Calder & Young, 2005; Kaufmann, Schweinberger, 2004). This debate has

largely centred on the idea that both tasks likely tap into different processes. For example, Bruce (1988) argued that recognising a facial identity requires that the perceptual system represent as common the identity of a face across changes in visual appearance, such as expression. In contrast, the task for the perceptual system when recognising expression is to recognise as common all instances of the same expression across a range of facial identities. This led Bruce (1988) to conclude that it was unlikely that the same information was extracted from the face to support each kind of processing, rather it was argued that each type of processing was likely to be unique (see also Bruce & Young, 1986). Indeed, a number of lesion studies have demonstrated that expression recognition can remain intact despite an impairment in identity processing (Humphreys, Donnelly, & Riddoch, 1993; Tranel, Damasio, & Damasio, 1988) and this has also been observed within the developmental prosopagnosia literature (Duchaine, Parker, & Nakayama, 2003; Humphreys, Avidan, & Behrmann, 2007; Nunn, Postma, & Pearson, 2001). However, recent studies have demonstrated that expression and identity recognition may overlap, such that representations of facial identity may be influenced by facial expression and vice versa (Calder & Young, 2005; de Gelder, Frissen, Barton, & Hadjikhani, 2003; Ellamil, Susskind, & Anderson, 2008; Kaufmann, Schweinberger, 2004). Moreover, it has been argued that behavioural measures which have argued for intact expression recognition, in spite of impaired identity recognition, have rarely been equated for task difficulty (Calder & Young, 2005). For example, when prosopagnosics are asked to identify a famous face image, the face could be any one of a possible endless list of identities. However, when they are asked to identify an expression, the answer is largely restricted to an alternative fixed or forced choice procedure, e.g. is it happy, sad, angry and so forth. As such, Calder & Young (2005) argued that facial expression impairments are likely to be under reported in this context.

This thesis did not set out to establish if identity and expression were processed independently, rather it sought to examine if representations of facial identities were more robust when the face had been learned through dynamic, rather than static, non-rigid or rigid face changes. However, the findings from Chapter 4 still offer some insight. Here, it was observed that both UM and PL could reliably match facial expressions across different facial identities, regardless of whether the face had been learned in motion or through static presentation. Therefore, it is *possible* that both static *and* dynamic expression processing may be to some extent intact in UM and PL. However, even if expression processing is impaired at a more subtle level, the findings at least suggest that UM and PL could recover information to process the expression rather than the identity of the face, when the face had been learned in motion. Indeed, very recent findings argue that even when static expression processing is impaired in

acquired prosopagnosia, this does not rule out the fact that dynamic expression processing may remain intact (Richoz, Jack, Garrod, Schyns, & Caldara, 2014). However, in the current context it is particularly important to state that the expression matching task reported in Chapter 4 was not equated for task difficulty with the identity matching task, a point which has been raised as a possible explanation for the findings of 'preserved' expression, relative to identity processing in this cohort (Calder & Young, 2005). Future studies, may overcome this challenge by using expressions tasks which are more perceptually matched in terms of task difficulty to the identity task. For example, the similarity between the target and the distractor images in the identity and expression matching tasks could be parametrically varied along a critical dimension (e.g. morphing the distractor expression or distractor identity with the target expression or target identity in degrees of e.g. 10-100%). Indeed, the task could also change to one which involved matching for expression while the 'task irrelevant' dimension of identity was altered, and vice versa. Examining how changes in these task irrelevant dimensions affect face processing in DP, particularly in the presence of facial motion, would offer insight into how these two processes may interact in normal face processing.

In addition, it should be noted that all facial identities used in the current thesis were unfamiliar. While it is likely that motion assists in encoding these new facial identities, it is possible that as the face becomes increasingly familiar, facial motion may become more relevant for social signal processing (Bruce & Young, 1986). Although expression and identity may still interact, and possibly be processed by the similar visual mechanism (Calder & Young, 2005), it is possible that the weight given to either cue may change as a consequence of the familiarity of the face.

7.4. Limitations of current studies and future directions

As mentioned in the opening Chapter face perception and recognition is strongly shaped by the faces in our surrounding environment. Indeed, face processing becomes tuned to the 'diet' of faces we are exposed to on a regular basis (McKone et al., 2011; Valentine, 1991). This has a direct consequence on our ability to, for example, discriminate between faces from other races versus the faces of the race by which we are surrounded (Chiroro, Tredoux, Radaelli, & Meissner, 2008; Chiroro & Valentine, 1995). With that in mind, it is worth considering that studies with older adult participants which have used static face images of older adults as stimuli, have reported that age-related decline in face recognition may be less evident for the recognition of older adult faces, relative to younger adult faces (Firestone, Turk-Browne & Ryan 2007; Fulton & Bartlett, 1991; Lamont, Stewart-Williams, & Podd, 2005). This effect is reported as the 'own-age bias' (for

review, see Rhodes & Anastasi, 2011) and is similar to the other-race effect in that it is thought to emerge due to the extensive exposure that an individual has with faces of their own-age group (Wiese, Komes, & Schweinberger, 2012). Therefore, it would be of great interest to examine how facial motion may impact on the processing of faces from the participant's own-age group. The current findings suggest that motion enhances unfamiliar processing in older adults, even when the face image is that of a different age group. It would be possible to directly replicate this paradigm with younger adults, using older adult faces as target identities to establish if this effect may also be evident in younger adults' performance. Moreover, examining if an own-age bias extends to voice and body processing in these cohorts would warrant future research (e.g. see Moyse, Beaufort, & Brédart, 2014).

In addition, it is important to note that the studies in this thesis were strongly focused on face, body and voice *perception* rather than memory, which is typically examined over more extended intervals. Moreover, both cases of DP explored in this thesis were characterised by apperceptive, rather than associative, prosopagnosia. As such, research investigating if these effects extend across different time scales in neurotypicals and associative prosopagnosics may help reveal whether dynamic and multisensory cues confer an advantage in the maintenance of a more robust representation of the identity in long-term memory in neurotypical processing (see e.g. Pilz et al., 2006). Indeed, the results from Chapter 2, which demonstrated that face perception bore a strong relationship with face memory, offers promise to the idea that these effects may extend over more long-term intervals. Furthermore, it would be of benefit to explore the role of motion in the processing of other categories of 'non-social' visual stimuli, such as objects. Although compared to face recognition, the recognition of objects appears to be less affected by the ageing process (Boutet & Faubert, 2006), a recent study has suggested that age is associated with greater viewpoint dependency in object recognition (see Pilz, Konar, Vuong, Bennett, & Sekuler, 2011). It is possible that, in parallel to the findings from Chapter 3, ageing may offer a unique insight into how dynamic cues affect the perceptual encoding of other objects. A number of studies have demonstrated that characteristic motion cues also extend to the learning and subsequent recognition of novel objects (e.g. Bülthoff & Newell, 2006; Chuang, Vuong, & Bülthoff, 2012; Newell, Wallraven, & Huber, 2004; Setti & Newell, 2010). Indeed, examining if developmental prosopagnosics can integrate form and motion cues for the purpose of learning and recognising objects, may offer an insight into how face specific the deficits recorded in Chapter 4 of this thesis are for this cohort.

7.5. Synopsis

The findings from this collection of studies support emerging theories of person perception and demonstrate that dynamic, cross-modal cues such as, facial form, facial motion, body motion and the voice, *interact* in person perception. For the most part, the findings suggest that these interactions occur at earlier stages of processing than previously assumed and that they may be mediated by dynamic network *interactions* in the brain. It is hoped that the findings arising from this thesis may inform future research. Indeed, examining some of the behavioural effects observed in this thesis with neuroimaging techniques and with younger developmental prosopagnosics would appear warranted. It is likely that the next decade will bring many more advances in our understanding of how and when these interactions arise in neurotypical processing, and indeed how they can be disrupted. In time it is possible that this knowledge may also lead to novel training paradigms for disorders of person recognition.

8. References

- Adolphs, R., Damasio, H., & Tranel, D. (2002). Neural systems for recognition of emotional prosody: A 3-D lesion study. *Emotion*, 2(1), 23–51. doi:10.1037//1528-3542.2.1.23
- Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: from physiology to behaviour. *Seeing and Perceiving*, 23(1), 3–38. doi:10.1163/187847510X488603
- Anaki, D., Kaufman, Y., Freedman, M., & Moscovitch, M. (2007). Associative (prosop)agnosia without (apparent) perceptual deficits: a case-study. *Neuropsychologia*, 45(8), 1658–1671. doi:10.1016/j.neuropsychologia.2007.01.003
- Arrighi, R., Burr, D., Nazionale, I., & Inoa, A. (2009). Meaningful auditory information enhances perception of visual biological motion. *Journal of Vision*, 9(4), 1–7. doi:10.1167/9.4.25
- Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and spatial factors in gait perception that influence gender recognition. *Perception And Psychophysics*, 23(2), 145–152. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/643509>
- Barnes, C. S., De L'Aune, W., & Schuchard, R. A. (2011). A test of face discrimination ability in aging and vision loss. *Optometry and Vision Science*, 88(2), 188–199. doi:10.1097/OPX.0b013e318205a17c
- Bartlett, J. C., & Leslie, J. E. (1986). Aging and memory for faces versus single views of faces. *Memory & Cognition*, 14(5), 371–381. doi:10.3758/BF03197012
- Bartlett, J. C., Leslie, J. E., Tubbs, A., & Fulton, A. (1989). Aging and memory for pictures of faces. *Psychology and Aging*, 4(3), 276–283. doi:10.1037//0882-7974.4.3.276
- Barton, J. J. S. (2008). Structure and function in acquired prosopagnosia: lessons from a series of 10 patients with brain damage. *Journal of Neuropsychology*, 2(1), 197–225. doi:10.1348/174866407X214172
- Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nature Neuroscience*, 7(11), 1190–1192. doi:10.1038/nn1333
- Behrmann, M., & Avidan, G. (2005). Congenital prosopagnosia: face-blind from birth. *Trends in Cognitive Sciences*, 9(4), 180–187. doi:10.1016/j.tics.2005.02.011
- Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, 102(4), 711–725. doi:10.1111/j.2044-8295.2011.02041.x
- Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*, 8(3), 129–135. doi:10.1016/j.tics.2004.01.008
- Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, 13(1), 17–26. doi:10.1016/S0926-6410(01)00084-2

- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403(6767), 309–312. doi:10.1038/35002078
- Bennett, P. J., Sekuler, R., & Sekuler, A. B. (2007). The effects of aging on motion detection and direction identification. *Vision Research*, 47(6), 799–809. doi:10.1016/j.visres.2007.01.001
- Bennetts, R., Burke, D., Brooks, K., & Robbins, R. (2010). The use of dynamic cues in self and familiar face recognition. *Proceedings of the 9th Conference of the Australasian Society for Cognitive Science*, 21–27. doi:10.5096/ASCS20094
- Benton, A. L., Hamsher, K., Varney, N. R., & Spreen, O. (1994). *Facial recognition: Stimulus and multiple choice pictures*. New York: Oxford University Press.
- Besle, J., Fort, A., Delpuech, C., & Giard, M. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, 20(8), 2225–2234.
- Beyn, E. S., & Knyazeva, G. R. (1962). The problem of prosopagnosia. *Journal of Neurology, Neurosurgery, and Psychiatry*, 25, 154–158. doi:10.1136/jnnp.25.2.154
- Bilalic, M., Langner, R., Ulrich, R., & Grodd, W. (2011). Many faces of expertise: fusiform face area in chess experts and novices. *Journal of Neuroscience*, 31(28), 10206–10214. doi:10.1523/JNEUROSCI.5727-10.2011
- Billino, J., Bremmer, F., & Gegenfurtner, K. R. (2008). Differential aging of motion processing mechanisms: evidence against general perceptual decline. *Vision Research*, 48(10), 1254–1261. doi:10.1016/j.visres.2008.02.014
- Bindemann, M., & Sandford, A. (2011). Me, myself, and I: Different recognition rates for three photo-IDs of the same person. *Perception*, 40(5), 625–627. doi:10.1068/p7008
- Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, 58, 47–73. doi:10.1146/annurev.psych.57.102904.190152
- Blank, H., Anwender, A., & von Kriegstein, K. (2011). Direct structural connections between voice- and face-recognition areas. *The Journal of Neuroscience*, 31(36), 12906–12915. doi:10.1523/JNEUROSCI.2091-11.2011
- Bodamer, J. (1947). Die Prosop-Agnosie. *Archiv Fuer Psychiatrie Und Nervenkrankheiten*, 179, 6–54.
- Bonda, E., Petrides, M., Ostry, D., & Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *The Journal of Neuroscience*, 16(11), 3737–3744. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8642416>
- Boutet, I., & Faubert, J. (2006). Recognition of faces and complex objects in younger and older adults. *Memory & Cognition*, 34(4), 854–864. doi:10.3758/BF03193432
- Boutsen, L., & Humphreys, G. W. (2002). Face context interferes with local part processing in a prosopagnosic patient. *Neuropsychologia*, 40(13), 2305–13. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12417460>

References

- Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., ... Yovel, G. (2009). Diagnosing prosopagnosia: Effects of ageing, sex, and participant-stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive Neuropsychology*, 26(5), 423–455. doi:10.1080/02643290903343149
- Brooks, A., Schouten, B., Troje, N. F., & Verfaillie, K. (2008). Correlated changes in perceptions of the gender and orientation of ambiguous biological motion figures Tardigrades survive exposure to space in low Earth orbit. *Current Biology*, 18(17), 728–729.
- Brooks, A., van der Zwan, R., Billard, A., Petreska, B., Clarke, S., & Blanke, O. (2007). Auditory motion affects visual biological motion processing. *Neuropsychologia*, 45(3), 523–530. doi:10.1016/j.neuropsychologia.2005.12.012
- Bruce, V. (1982). Changing faces: visual and non-visual coding processes in face recognition. *British Journal of Psychology*, 73(1), 105–116. doi:10.1111/j.2044-8295.1982.tb01795.x
- Bruce, V. (1988). *Recognising Faces*. Essex: UK: Lawrence Erlbaum Associates.
- Bruce, V., Burton, M. a., & Dench, N. (1994). What's distinctive about a distinctive face? *The Quarterly Journal of Experimental Psychology Section A*, 47(1), 119–141. doi:10.1080/14640749408401146
- Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001a). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*, 7, 207–218. doi:10.1037//1076-898X.7.3.207
- Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001b). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology Applied*, 7(3), 207–218. Retrieved from <http://eprints.gla.ac.uk/24953/>
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77(3), 305–327. doi:10.1111/j.2044-8295.1986.tb02199.x
- Bruce, V., & Young, A. (2012). *Face Perception. Face Perception*. Sussex: Psychology Press.
- Bülthoff, H. H., Cunningham, D. W., & Wallraven, C. (2011). Dynamic aspects of face processing in humans. In S. Z. Li & A. K. Jain (Eds.), *Handbook of Face Recognition*. London: Springer. doi:10.1007/978-0-85729-932-1
- Bülthoff, I., & Newell, F. N. (2015). Distinctive voices enhance the visual recognition of unfamiliar faces. *Cognition*, 137, 9-21.
- Bülthoff, I., & Newell, F. N. (2006). The role of familiarity in the recognition of static and dynamic objects. *Progress in Brain Research*, 154(315-325). doi:[http://dx.doi.org/10.1016/S0079-6123\(06\)54017-8](http://dx.doi.org/10.1016/S0079-6123(06)54017-8)
- Burianová, H., Lee, Y., Grady, C. L., & Moscovitch, M. (2013). Age-related dedifferentiation and compensatory changes in the functional network underlying face processing. *Neurobiology of Aging*, 34(12), 2759–2767. doi:10.1016/j.neurobiolaging.2013.06.016

- Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactive model. *British Journal of Psychology*, 81, 361–380. doi:10.1111/j.2044-8295.1990.tb02367.x
- Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, 42(1), 286–291. doi:10.3758/BRM.42.1.286
- Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor quality video: Evidence from security surveillance. *Psychological Science*, 10(3), 243–248.
- Busigny, T., Joubert, S., Felician, O., Ceccaldi, M., & Rossion, B. (2010). Holistic perception of the individual face is specific and necessary : Evidence from an extensive case study of acquired prosopagnosia. *Neuropsychologia*, 48(14), 4057–4092. doi:10.1016/j.neuropsychologia.2010.09.017
- Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, 6, 641–651. doi:10.1038/nrn1724
- Calvert, G. a, & Campbell, R. (2003). Reading speech from still and moving faces: the neural substrates of visible speech. *Journal of Cognitive Neuroscience*, 15(1), 57–70. doi:10.1162/089892903321107828
- Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., Mc Guire, P. K., ... David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, 276(5312), 593–596. doi:10.1126/science.276.5312.593
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, 10(11), 649–657. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10837246>
- Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences*, 11(12), 535–543. doi:10.1016/j.tics.2007.10.001
- Cappelletta, L., & Harte, N. (2011). Viseme definitions comparison for visual-only speech recognition. In *Proceedings of the European Signal Processing Conference* (pp. 2109–2113).
- Cappelletta, L., & Harte, N. (2012). Phoneme-to-viseme mapping for visual speech recognition. In *Proceedings of the International Conference on Patter Recognition Applications and Methods* (pp. 322–329).
- Carr, V. A., Castel, A. D., & Knowlton, B. J. (2014). Age-related differences in memory after attending to distinctiveness or similarity during learning. *Aging, Neuropsychology, and Cognition*, 1–15. doi:10.1080/13825585.2014.898735
- Casey, S. J., & Newell, F. N. (2007). Are representations of unfamiliar faces independent of encoding modality? *Neuropsychologia*, 45(3), 506–513. doi:10.1016/j.neuropsychologia.2006.02.011
- Chaby, L., Narme, P., & George, N. (2011). Older adults' configural processing of faces: role of second-order information. *Psychology and Aging*, 26(1), 71–79. doi:10.1037/a0020873

References

- Chiroro, P. M., Tredoux, C. G., Radaelli, S., & Meissner, C. A. (2008). Recognizing faces across continents: the effect of within-race variations on the own-race bias in face recognition. *Psychonomic Bulletin & Review*, 15(6), 1089–1092. doi:10.3758/PBR.15.6.1089
- Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *The Quarterly Journal of Experimental Psychology Section A : Human Experimental Psychology*, 48a(4), 879–894.
- Christie, F., & Bruce, V. (1998). The role of dynamic information in the recognition of unfamiliar faces. *Memory & Cognition*, 26(4), 780–90. doi:10.3758/BF03211397
- Chuang, L. L., Vuong, Q. C., & Bülthoff, H. H. (2012). Learned non-rigid object motion is a view-invariant cue to recognizing novel objects. *Frontiers in Computational Neuroscience*, 6, 26. doi:10.3389/fncom.2012.00026
- Clifford, B. R. (1980). Voice Identification by Human Listeners : On Earwitness Reliability. *Law and Human Behavior*, 4(4), 373–394.
- Cowey, A., & Vaina, L. M. (2000). Blindness to form from motion despite intact static form perception and motion detection. *Neuropsychologia*, 38(5), 566–578. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10689034>
- Crawford, J. R., & Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia*, 40(8), 1196–208. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11931923>
- Crawford, J. R., & Garthwaite, P. H. (2005). Testing for suspected impairments and dissociations in single-case studies in neuropsychology: evaluation of alternatives using monte carlo simulations and revised tests for dissociations. *Neuropsychology*, 19(3), 318–331. doi:10.1037/0894-4105.19.3.318
- Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology*, 27(3), 245–260. doi:10.1080/02643294.2010.513967
- Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist*, 12(4), 482–486.
- Cronin-Golomb, A., Gilmore, G. C., Nearing, S., Morrison, S. R., & Laudate, T. M. (2007). Enhanced stimulus strength improves visual cognition in aging and Alzheimer's disease. *Cortex*, 43, 952–966. doi:10.1016/S0010-9452(08)70693-2
- Cutting, J. E., & Kozlowski, L. T. (1977). Recognising friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, 9(5), 353–356.
- Cutting, J. E., Proffitt, D. R., & Kozlowski, L. T. (1978). A biomechanical invariant for gait perception. *Journal of Experimental Psychology: Human Perception and Performance*, 4(3), 357–372. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/681885>

- Daniel, S., & Bentin, S. (2012). Age-related changes in processing faces from detection to identification: ERP evidence. *Neurobiology of Aging*, 33(1), 206.e1–28. doi:10.1016/j.neurobiolaging.2010.09.001
- Davis, C., Kislyuk, D., Kim, J., & Sams, M. (2008). The effect of viewing speech on auditory speech processing is different in the left and right hemispheres. *Brain Research*, 1242, 151–61. doi:10.1016/j.brainres.2008.04.077
- De Gelder, B., Frissen, I., Barton, J., & Hadjikhani, N. (2003). A modulatory role for facial expressions in prosopagnosia. *Proceedings of the National Academy of Sciences of the United States of America*, 100(22), 13105–10. doi:10.1073/pnas.1735530100
- De Gelder, B., & Vroomen, J. (1998). Impairment of speech-reading in prosopagnosia. *Speech Communication*, 26, 89–96.
- De Renzi, E., Faglioni, P., Grossi, D., & Nichelli, P. (1991). Apperceptive and associative forms of prosopagnosia. *Cortex*, 27(2), 213–221.
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science*, 208(4448), 1174–1176. doi:10.1126/science.7375928
- DeGutis, J., Chatterjee, G., Mercado, R. J., & Nakayama, K. (2012). Face gender recognition in developmental prosopagnosia : Evidence for holistic processing and use of configural information. *Visual Cognition*, 20(10), 1242–1253.
- Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2012). The Cambridge Car Memory Test: A task matched in format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods*, 44(2), 587–605. doi:10.3758/s13428-011-0160-2
- Dinkelacker, V., Grüter, M., Klaver, P., Grüter, T., Specht, K., Weis, S., ... Fernandez, G. (2011). Congenital prosopagnosia: multistage anatomical and functional deficits in face processing circuitry. *Journal of Neurology*, 258(5), 770–782. doi:10.1007/s00415-010-5828-5
- Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, 293(5539), 2470–2473. doi:10.1126/science.1063414
- Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on “sensory-specific” brain regions, neural responses, and judgments. *Neuron*, 57(1), 11–23. doi:10.1016/j.neuron.2007.12.013
- Driver, J., & Spence, C. (2000). Multisensory perception: beyond modularity and convergence. *Current Biology*, 10(20), R731–R735. doi:10.1016/S0960-9822(00)00740-5
- Duchaine, B. C. (2008). Comment on prevalence of hereditary prosopagnosia (HPA) in Hong Kong Chinese population. *American Journal of Medical Genetics. Part A*, 146A(22), 2860–2. doi:10.1002/ajmg.a.32548

References

- Duchaine, B. C., Dingle, K., Butterworth, E., & Nakayama, K. (2004). Normal greeble learning in a severe case of developmental prosopagnosia. *Neuron*, 43(4), 469–473. doi:10.1016/j.neuron.2004.08.006
- Duchaine, B. C., & Nakayama, K. (2004). Developmental prosopagnosia and the Benton Facial Recognition Test. *Neurology*, 62(7), 1219–1220. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15079032>
- Duchaine, B. C., Parker, H., & Nakayama, K. (2003). Normal recognition of emotion in a prosopagnosic. *Perception*, 32(7), 827–838. doi:10.1068/p5067
- Duchaine, B. C., Yovel, G., Butterworth, E. J., & Nakayama, K. (2006). Prosopagnosia as an impairment to face-specific mechanisms: Elimination of the alternative hypotheses in a developmental case. *Cognitive Neuropsychology*, 23(5), 714–747. doi:10.1080/02643290500441296
- Duchaine, B. C., Germine, L., & Nakayama, K. (2007). Family resemblance: ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology*, 24(4), 419–430. doi:10.1080/02643290701380491
- Duchaine, B., Murray, H., Turner, M., White, S., & Garrido, L. (2009). Normal social cognition in developmental prosopagnosia. *Cognitive Neuropsychology*, 26(7), 620–634. doi:10.1080/02643291003616145
- Duchaine, B., & Nakayama, K. (2005). Dissociations of face and object recognition in developmental prosopagnosia. *Journal of Cognitive Neuroscience*, 17(2), 249–61. doi:10.1162/0898929053124857
- Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, 44(4), 576–585. doi:10.1016/j.neuropsychologia.2005.07.001
- Duchaine, B., & Weidenfeld, A. (2003). An evaluation of two commonly used tests of unfamiliar face recognition. *Neuropsychologia*, 41, 713–720.
- Edmonds, E. C., Glisky, E. L., Bartlett, J. C., & Rapcsak, S. Z. (2012). Cognitive mechanisms of false facial recognition in older adults. *Psychology and Aging*, 27(1), 54–60. doi:10.1037/a0024582
- Eimer, M. (2011). The Face-Sensitive N170 Component of the Event-Related Brain Potential. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford Handbook of Face Perception*. Oxford: Oxford University Press.
- Ellamil, M., Susskind, J. M., & Anderson, A. K. (2008). Examinations of identity in facial expression adaptation. *Cognitive Affective & Behavioral Neuroscience*, 8(3), 273–281.
- Ellis, H. D., & Florence, M. (1990). Bodamer's (1947) paper on prosopagnosia. *Cognitive Neuropsychology*, 7(2), 81–105. doi:10.1080/02643299008253437
- Ellis, H. D., Jones, D. M., & Mosdell, N. (1997). Intra- and inter-modal repetition priming of familiar faces and voices. *British Journal of Developmental Psychology*, 88, 143–156.

- Esins, J., Bülthoff, I., & Schultz, J. (2014). Facial motion does not help face recognition in congenital prosopagnosia. *Journal of Vision*, 14(10), 1436.
- Farah, M. J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision*. Cambridge: MIT Press.
- Farivar, R., Blanke, O., & Chaudhuri, A. (2009). Dorsal-ventral integration in the recognition of motion-defined unfamiliar faces. *The Journal of Neuroscience*, 29(16), 5336–5342. doi:10.1523/JNEUROSCI.4978-08.2009
- Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Perrett, D. I. (2008). The role of femininity and averageness of voice pitch in aesthetic judgments of women's voices. *Perception*, 37(4), 615–623. doi:10.1068/p5514
- Feinberg, D. R., Jones, B. C., DeBruine, L. M., Moore, F. R., Law Smith, M. J., Cornwell, R. E., ... Perrett, D. I. (2005). The voice and face of woman: One ornament that signals quality? *Evolution and Human Behavior*, 26(5), 398–408. doi:10.1016/j.evolhumbehav.2005.04.001
- Feinberg, D. R., Jones, B. C., Little, a. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour*, 69(3), 561–568. doi:10.1016/j.anbehav.2004.06.012
- Fifer, W. P., & Moon, C. M. (1994). The role of mother's voice in the organization of brain function in the newborn. *Acta Paediatrica Supplementum*, 83(s397), 86–93. Retrieved from <http://dx.doi.org/10.1111/j.1651-2227.1994.tb13270.x>
- Firestone, A., Turk-Browne, N. B., & Ryan, J. D. (2007). Age-related deficits in face recognition are related to underlying changes in scanning behavior. *Neuropsychology, Development, and Cognition*, 14(6), 594–607. doi:10.1080/13825580600899717
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189–198.
- Fox, C. J., Hanif, H. M., Iaria, G., Duchaine, B. C., & Barton, J. J. S. (2011). Perceptual and anatomic patterns of selective deficits in facial identity and expression processing. *Neuropsychologia*, 49(12), 3188–200. doi:10.1016/j.neuropsychologia.2011.07.018
- Fox, C. J., Iaria, G., & Barton, J. J. S. (2008). Disconnection in prosopagnosia and face processing. *Cortex*, 44(8), 996–1009. doi:10.1016/j.cortex.2008.04.003
- Fox, C. J., Iaria, G., & Barton, J. J. S. (2009). Defining the face processing network: Optimization of the functional localizer in fMRI. *Human Brain Mapping*, 30(5), 1637–1651. doi:10.1002/hbm.20630
- Fox, C. J., Iaria, G., Duchaine, B. C., & Barton, J. J. S. (2013). Residual fMRI sensitivity for identity changes in acquired prosopagnosia. *Frontiers in Psychology*, 4, 756. doi:10.3389/fpsyg.2013.00756
- Fulton, A., & Bartlett, J. C. (1991). Young and old faces in young and old heads: the factor of age in face recognition. *Psychology and Aging*, 6(4), 623–630. doi:10.1037//0882-7974.6.4.623

References

- Furl, N., Henson, R. N., Friston, K. J., & Calder, A. J. (2014). Network interactions explain sensitivity to dynamic faces in the superior temporal sulcus. *Cerebral Cortex*, epub ahead of print. doi:10.1093/cercor/bhu083
- Gao, L., Xu, J., Zhang, B., Zhao, L., Harel, A., & Bentin, S. (2009). Aging effects on early-stage face perception: an ERP study. *Psychophysiology*, 46(5), 970–983. doi:10.1111/j.1469-8986.2009.00853.x
- Garrido, L., Duchaine, B., & Nakayama, K. (2008). Face detection in normal and prosopagnosic individuals. *Journal of Neuropsychology*, 2(1), 119–140. doi:10.1348/174866407X246843
- Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J. R., ... Duchaine, B. (2009). Developmental phonagnosia: a selective deficit of vocal identity recognition. *Neuropsychologia*, 47(1), 123–131. doi:10.1016/j.neuropsychologia.2008.08.003
- Gauthier, I., & Nelson, C. A. (2001). The development of face expertise. *Current Opinion In Neurobiology*, 11(2), 219–224. doi:10.1016/S0959-4388(00)00200-2
- Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3(2), 191–197. doi:10.1038/72140
- Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform “face area” is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience*, 12(3), 495–504.
- Germine, L., Cashdollar, N., Düzel, E., & Duchaine, B. (2011). A new selective developmental deficit: impaired object recognition with normal face recognition. *Cortex*, 47(5), 598–607. doi:10.1016/j.cortex.2010.04.009
- Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: face learning ability peaks after age 30. *Cognition*, 118(2), 201–210. doi:10.1016/j.cognition.2010.11.002
- Ghazanfar, A. A. (2011). Unity of the Senses for Primate Vocal Communication. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes*. (pp. 653–666). Boca Raton (FL): CRC Press. doi:10.1201/b11092-41
- Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, 10(6), 278–285. doi:10.1016/j.tics.2006.04.008
- Gilaie-Dotan, S., Bentin, S., Harel, M., Rees, G., & Saygin, a P. (2011). Normal form from biological motion despite impaired ventral stream function. *Neuropsychologia*, 49(5), 1033–1043. doi:10.1016/j.neuropsychologia.2011.01.009
- Goffaux, V., & Rossion, B. (2006). Faces are “spatial” - Holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology. Human Perception and Performance*, 32(4), 1023–1039. doi:10.1037/0096-1523.32.4.1023

- Goh, J O Suzuki, A Park, D. C. (2011). Reduced neural selectivity increases fMRI adaptation with age during face discrimination. *Neuroimage*, 51(1), 336–344. doi:10.1016/j.neuroimage.2010.01.107.Reduced
- Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, 64, 257–283. doi:10.1146/annurev-psych-113011-143802
- Gordon, M. S., & Allen, S. (2009). Audiovisual speech in older and younger adults: integrating a distorted visual signal with speech in noise. *Experimental Aging Research*, 35(2), 202–219. doi:10.1080/03610730902720398
- Gougoux, F., Belin, P., Voss, P., Lepore, F., Lassonde, M., & Zatorre, R. J. (2009). Voice perception in blind persons: a functional magnetic resonance imaging study. *Neuropsychologia*, 47(13), 2967–2974. doi:10.1016/j.neuropsychologia.2009.06.027
- Grady, C. L. (2002). Age-related differences in face processing: A meta-analysis of three functional neuroimaging experiments. *Canadian Journal of Experimental Psychology*, 56(3), 208–220. doi:10.1037/h0087398
- Grady, C. L., Maisog, J. M., Horwitz, B., Ungerleider, L. G., Mentis, M. J., Salerno, J. A., ... Haxby, J. V. (1994). Age-related processing changes in cortical blood flow activation of faces and location during visual processing of faces and location. *The Journal of Neuroscience*, 14(3), 1450–1462.
- Grady, C. L., McIntosh, A. R., Bookstein, F., Horwitz, B., Rapoport, S. I., & Haxby, J. V. (1998). Age-related changes in regional cerebral blood flow during working memory for faces. *NeuroImage*, 8(4), 409–25. doi:10.1006/nimg.1998.0376
- Grady, C. L., McIntosh, A. R., Horwitz, B., Maisog, J. M., Ungerleider, L. G., Mentis, M. J., ... Haxby, J. V. (1995). Age-related reductions in human recognition memory due to impaired encoding. *Science*, 269(5221), 218–221. doi:10.1126/science.7618082
- Grady, C. L., McIntosh, A. R., Horwitz, B., & Rapoport, S. I. (2000). Age-related changes in the neural correlates of degraded and nondegraded face processing. *Cognitive Neuropsychology*, 17(1), 165–186. doi:10.1080/026432900380553
- Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, 7(5), 555–562. doi:10.1038/nn1224
- Grossman, E. D., & Blake, R. (2002). Brain Areas Active during Visual Perception of Biological Motion. *Neuron*, 35(6), 1167–1175. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12354405>
- Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience*, 12(5), 711–720. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11054914>
- Grueter, M., Grueter, T., Bell, V., Horst, J., Laskowski, W., Sperling, K., ... Kennerknecht, I. (2007). Hereditary prosopagnosia: the first case series. *Cortex*, 43, 734–749.

References

- Grüter, T., Grüter, M., & Carbon, C.-C. (2008). Neural and genetic foundations of face recognition and prosopagnosia. *Journal of Neuropsychology*, 2(1), 79–97. doi:10.1348/174866407X231001
- Habak, C., Wilkinson, F., & Wilson, H. R. (2008). Aging disrupts the neural transformations that link facial identity across views. *Vision Research*, 48(1), 9–15. doi:10.1016/j.visres.2007.10.007
- Halevina, A., & Troje, N. F. (2007). Sex classification of point light walkers: Viewpoint, structure, kinematics. *Journal of Vision*, 7(9), 483.
- Hancock, P. J. B., Bruce, V., & Burton, A. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, 4(9), 330–337. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10962614>
- Hanley, J. R., Smith, S. T., & Had, J. (1998). I recognise you but I can't place you :An investigation of familiar-only experiences during tests of voice and face recognition. *The Quarterly Journal of Experimental Psychology*, 51A(1), 179–195.
- Hanley, J. R., & Turner, J. M. (2000). Why are familiar-only experiences more frequent for voices than for faces ? *The Quarterly Journal of Experimental Psychology Section A*, 53A(4), 1105–1116.
- Haxby, J., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6), 223–233. doi:10.1016/S1364-6613(00)01482-0
- Haxby, J. V., Hoffman, E. a, & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, 51(1), 59–67. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11801231>
- Hein, G., & Knight, R. T. (2008). Superior temporal sulcus--It's my area: or is it? *Journal of Cognitive Neuroscience*, 20(12), 2125–2136. doi:10.1162/jocn.2008.20148
- Herzmann, G., Kunina, O., Sommer, W., & Wilhelm, O. (2010). Individual differences in face cognition: brain-behavior relationships. *Journal of Cognitive Neuroscience*, 22(3), 571–589. doi:10.1162/jocn.2009.21249
- Hildebrandt, A., Wilhelm, O., Schmiedek, F., Herzmann, G., & Sommer, W. (2011). On the specificity of face cognition compared with general cognitive functioning across adult age. *Psychology and Aging*, 26(3), 701–715. doi:10.1037/a0023056
- Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology*, 11(11), 880–885. doi:10.1016/S0960-9822(01)00243-3
- Hill, H., & Pollick, F. E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light displays. *Psychological Science*, 11(3), 223–238. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11273407>
- Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, 3(1), 80–84. doi:10.1038/71152

- Hubbard, A. L., Wilson, S. M., Callan, D. E., & Dapretto, M. (2009). Giving speech a hand: gesture modulates activity in auditory cortex during speech perception. *Human Brain Mapping, 30*(3), 1028–1037. doi:10.1002/hbm.20565
- Humphreys, G. W., Donnelly, N., & Riddoch, M. J. (1993). Expression is computed separately from facial identity, and it is computed separately for moving and static faces: neuropsychological evidence. *Neuropsychologia, 31*(2), 173–181.
- Humphreys, K., Avidan, G., & Behrmann, M. (2007). A detailed investigation of facial expression processing in congenital prosopagnosia as compared to acquired prosopagnosia. *Experimental Brain Research, 176*(2), 356–373. doi:10.1007/s00221-006-0621-5
- Hutchinson, C. V, Arena, A., Allen, H. A., & Ledgeway, T. (2012). Psychophysical correlates of global motion processing in the aging visual system: a critical review. *Neuroscience and Biobehavioral Reviews, 36*(4), 1266–1272. doi:10.1016/j.neubiorev.2012.02.009
- Jääskeläinen, I. P., Ojanen, V., Ahveninen, J., Auranen, T., Levänen, S., Möttönen, R., ... Sams, M. (2004). Adaptation of neuromagnetic N1 responses to phonetic stimuli by visual speech in humans. *Neuroreport, 15*(18), 2741–2744. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15597045>
- Jeffery, L., Petrovski, S., & Rhodes, G. (2014). Adaptation to dynamic faces produces face identity aftereffects. *Journal of Vision, 14*(10), 554.
- Joassin, F., Maurage, P., & Campanella, S. (2011). The neural network sustaining the crossmodal processing of human gender from faces and voices: an fMRI study. *NeuroImage, 54*(2), 1654–1661. doi:10.1016/j.neuroimage.2010.08.073
- Joassin, F., Pesenti, M., Maurage, P., Verreckt, E., Bruyer, R., & Campanella, S. (2011). Cross-modal interactions between human faces and voices involved in person recognition. *Cortex, 47*(3), 367–376. doi:10.1016/j.cortex.2010.03.003
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*(2), 201–211.
- Johnson, M. H. (2000). Functional brain development in infants: elements of an interactive specialization framework. *Child Development, 71*(1), 75–81. doi:10.1111/1467-8624.00120
- Johnson, M. H. (2001). Functional brain development in humans. *Nature Reviews Neuroscience, 2*(7), 475–483. doi:10.1038/35081509
- Johnson, M. H. (2011). Face processing as a brain adaptation at multiple timescales. *Quarterly Journal of Experimental Psychology, 64*(10), 1873–1888. doi:10.1080/17470218.2011.590596
- Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a cocktail party: voice familiarity aids speech perception in the presence of a competing voice. *Psychological Science, 24*(10), 1995–2004. doi:10.1177/0956797613482467

References

- Jones, R. D., & Tranel, D. (2001). Severe developmental prosopagnosia in a child with superior intellect. *Journal of Clinical and Experimental Neuropsychology*, 23(3), 265–273. doi:10.1076/jcen.23.3.265.1183
- Kamachi, M., Hill, H., Lander, K., & Vatikiotis-Bateson, E. (2003). Putting the face to the voice: matching identity across modality. *Current Biology*, 13(19), 1709–1714. doi:10.1016/j.cub.2003.09.005
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, 17(11), 4302–4311.
- Kaufmann, J M Schweinberger, S. R. (2004). Expression influences the recognition of familiar faces. *Perception*, 33(4), 399–408. doi:10.1068/p5083
- Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Liezhong, G., & Pascalis, O. (2007). The other-race effect develops during infancy evidence of perceptual narrowing. *Psychological Science*, 18(12), 1084–1089. doi:10.1111/j.1467-9280.2007.02029.x.
- Kennerknecht, I., Grueter, T., Welling, B., Wentzek, S., Horst, J., Edwards, S., & Grueter, M. (2006). First report of prevalence of Non-Syndromic Hereditary Prosopagnosia (HPA). *American Journal of Medical Genetics Part A*, 140A, 1617–1622. doi:10.1002/ajmg.a
- Kennerknecht, I., Ho, N. Y., & Wong, V. C. N. (2008). Prevalence of hereditary prosopagnosia (HPA) in Hong Kong Chinese population. *American Journal of Medical Genetics. Part A*, 146A(22), 2863–2870. doi:10.1002/ajmg.a.32552
- Kennerknecht, I., Kischka, C., Stemper, C., Elze, T., & Stollhoff, R. (2011). Heritability of Face Recognition. In T. Barbu (Ed.), *Face Analysis, Modeling and Recognition Systems*. InTech. doi:10.5772/1837
- Kenny Gibson, W., Cronin, H., Kenny, R. A., & Setti, A. (2014). Validation of the self-reported hearing questions in the Irish Longitudinal Study on Ageing against the Whispered Voice Test. *BMC Research Notes*, 7, 361. doi:10.1186/1756-0500-7-361
- Kenny, R. A., Whelan, B., Cronin, H., Kamiya, Y., Kearney, P., O'Regan, C., & Ziegel, M. (2010). *The Design of the Irish Longitudinal Study on Ageing*.
- Knappmeyer, B., Thornton, I. M., & Bülthoff, H. H. (2003). The use of facial motion and facial form during the processing of identity. *Vision Research*, 43(18), 1921–1936. doi:10.1016/S0042-6989(03)00236-0
- Knight, B., & Johnston, A. (1997). The role of movement in face recognition. *Visual Cognition*, 4(3), 265–273. doi:10.1080/713756764
- Köhler, W., & Wallach, H. (1944). Figural aftereffects: an investigation of visual processes. *Proceedings of the American Philosophical Society*, 88, 269–357.
- Konar, Y., Bennett, P. J., & Sekuler, A. B. (2013). Effects of aging on face identification and holistic face processing. *Vision Research*, 88, 38–46. doi:10.1016/j.visres.2013.06.003

- Konar, Y., & Troje, F. (2013). Biological motion perception in the elderly. In *Meeting of the Canadian Association for Neuroscience*.
- Kozlowski, L. T., & Cutting, J. E. (1977). Recognising the sex of the walker from a dynamic point-light display. *Perception & Psychophysics*, 21(6), 575–580.
- Kreiman, J., & Papcun, G. (1991). Comparing discrimination and recognition of unfamiliar voices. *Speech Communication*, 10(3), 265–275. doi:10.1016/0167-6393(91)90016-M
- Lachs, L., & Pisoni, D. B. (2004). Cross-modal source information and spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 30(2), 378–396. doi:10.1037/0096-1523.30.2.378.Cross-Modal
- Lamont, A. C., Stewart-Williams, S., & Podd, J. (2005). Face recognition and aging: effects of target age and memory load. *Memory & Cognition*, 33(6), 1017–1024. doi:10.3758/BF03193209
- Lander, K., & Bruce, V. (2000). Recognizing famous faces: Exploring the benefits of facial motion. *Ecological Psychology*, 12(4), 259–272.
- Lander, K., & Bruce, V. (2003). The role of motion in learning new faces. *Visual Cognition*, 10(8), 897–912. doi:10.1080/13506280344000149
- Lander, K., Christie, F., & Bruce, V. (1999). The role of movement in the recognition of famous faces. *Memory & Cognition*, 27(6), 974–985. doi:10.3758/BF03201228
- Lander, K., & Chuang, L. (2005). Why are moving faces easier to recognize? *Visual Cognition*, 12(3), 429–442. doi:10.1080/13506280444000382
- Lander, K., & Davies, R. (2007). Exploring the role of characteristic motion when learning new faces. *Quarterly Journal of Experimental Psychology*, 60(4), 519–526. doi:10.1080/17470210601117559
- Lander, K., Hill, H., Kamachi, M., & Vatikiotis-Bateson, E. (2007). It's not what you say but the way you say it: matching faces and voices. *Journal of Experimental Psychology: Human Perception and Performance*, 33(4), 905–914. doi:10.1037/0096-1523.33.4.905
- Lander, K., Humphreys, G., & Bruce, V. (2004). Exploring the role of motion in prosopagnosia: Recognizing, learning and matching faces. *Neurocase*, 10(6), 462–470. doi:10.1080/13554790490900761
- Lange, J., de Lussanet, M., Kuhlmann, S., Zimmermann, A., Lappe, M., Zwitserlood, P., & Dobel, C. (2009). Impairments of biological motion perception in congenital prosopagnosia. *PloS ONE*, 4(10), e7414. doi:10.1371/journal.pone.0007414
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, 24(8), 1377–1388. doi:10.1080/02699930903485076
- Latinus, M., & Belin, P. (2011). Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology*, 2(July), 175. doi:10.3389/fpsyg.2011.00175

References

- Latinus, M., & Taylor, M. J. (2012). Discriminating male and female voices: differentiating pitch and gender. *Brain Topography*, 25(2), 194–204. doi:10.1007/s10548-011-0207-9
- Latinus, M., Joassin, F., Watson, R., Charest, I., Love, S., Mcaleer, P., Belin, P. (2011). Conjoint and independent neural coding of bimodal face. *Perception ECVP Abstract Supplement*, 40, 23.
- Lattner, S., Maess, B., Wang, Y., Schauer, M., Alter, K., & Friederici, A. D. (2003). Dissociation of human and computer voices in the brain: evidence for a preattentive gestalt-like perception. *Human Brain Mapping*, 20(1), 13–21. doi:10.1002/hbm.10118
- Lattner, S., Meyer, M. E., & Friederici, A. D. (2005). Voice perception: Sex, pitch, and the right hemisphere. *Human Brain Mapping*, 24(1), 11–20. doi:10.1002/hbm.20065
- Laurienti, P. J., Burdette, J. H., Maldjian, J. a, & Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiology of Aging*, 27(8), 1155–1163. doi:10.1016/j.neurobiolaging.2005.05.024
- Le Grand, R., Cooper, P. A., Mondloch, C. J., Lewis, T. L., Sagiv, N., de Gelder, B., & Maurer, D. (2006). What aspects of face processing are impaired in developmental prosopagnosia? *Brain and Cognition*, 61(2), 139–158. doi:10.1016/j.bandc.2005.11.005
- Lee, L. C., Andrews, T. J., Johnson, S. J., Woods, W., Gouws, A., Green, G. G. R., & Young, A. W. (2010). Neural responses to rigidly moving faces displaying shifts in social attention investigated with fMRI and MEG. *Neuropsychologia*, 48(2), 477–490. doi:10.1016/j.neuropsychologia.2009.10.005
- Lee, Y., Duchaine, B., Wilson, H. R., & Nakayama, K. (2010). Three cases of developmental prosopagnosia from one family: detailed neuropsychological and psychophysical investigation of face processing. *Cortex*, 46(8), 949–964. doi:10.1016/j.cortex.2009.07.012
- Lee, Y., Grady, C. L., Habak, C., Wilson, H. R., & Moscovitch, M. (2011). Face processing changes in normal aging revealed by fMRI adaptation. *Journal of Cognitive Neuroscience*, 23(11), 3433–3447. doi:10.1162/jocn_a_00026
- Lee, Y., Habak, C., & Wilson, H. R. (2010). Seeing an unfamiliar face in rotational motion does not aid identity discrimination across viewpoints. *Vision Research*, 50(9), 854–859. doi:10.1016/j.visres.2010.02.013
- Lee, Y., Smith, C. R., Grady, C. L., Hoang, N., & Moscovitch, M. (2014). Broadly tuned face representation in older adults assessed by categorical perception. *Journal of Experimental Psychology. Human Perception and Performance*, 40(3), 1060–1071. doi:10.1037/a0035710
- Legault, I., Troje, N. F., & Faubert, J. (2012). Healthy older observers cannot use biological-motion point-light information efficiently within 4 m of themselves. *I-Perception*, 3(2), 104–111. doi:10.1068/i0485
- Leopold, D. A., Rhodes, G., Müller, K., Jeffery, L., B, P. R. S., & Mu, K. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society of Biological Sciences*, 272, 897–904. doi:10.1098/rspb.2004.3022

- Leopold, D. A., Toole, A. J. O., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, 4(1), 89–94.
- Little, A. C., Feinberg, D. R., Debruine, L. M., & Jones, B. C. (2013). Adaptation to faces and voices: unimodal, cross-modal, and sex-specific effects. *Psychological Science*, 24(11), 2297–2305. doi:10.1177/0956797613493293
- Liu, R., Pancaroglu, R., Hills, C., Duchaine, B., & Barton, J. (2014). Evaluation of Face and Voice Perception and Recognition Impairments in Prosopagnosia. *Neurology*, 82, S19.005.
- Liu, T. (2007). Learning sequence of views of three-dimensional objects: the effect of temporal coherence on object memory. *Perception*, 36(9), 1320–1333. doi:10.1068/p5778
- Longmore, C. A., Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception and Performance*, 34(1), 77–100. doi:10.1037/0096-1523.34.1.77
- Longmore, C. A., & Tree, J. J. (2013). Motion as a cue to face recognition: Evidence from congenital prosopagnosia. *Neuropsychologia*, 51(5), 864–875. doi:10.1016/j.neuropsychologia.2013.01.022
- Luis, C. A., Keegan, A. P., & Mullan, M. (2009). Cross validation of the Montreal Cognitive Assessment in community dwelling older adults residing in the Southeastern US. *International Journal of Geriatric Psychiatry*, 24, 197–201. doi:10.1002/gps
- Maguinness, C., Setti, A., Burke, K. E., Kenny, R. A., & Newell, F. N. (2011). The effect of combined sensory and semantic components on audio-visual speech perception in older adults. *Frontiers in Aging Neuroscience*, 3, 19. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3244611&tool=pmcentrez&rendertype=abstract>
- Mather, G., & Murdoch, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society B Biological Sciences*, 258(1353), 273–279.
- Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, 6(6), 255–260. doi:10.1016/S1364-6613(02)01903-4
- Mavica, L. W., & Barenholtz, E. (2013). Matching voice and face identity from static images. *Journal of Experimental Psychology: Human Perception and Performance*, 39(2), 307–312. doi:10.1037/a0030945
- Mc Donnell, R., & O' Sullivan, C. (2010). Movements and voices affect perceived sex of virtual conversers. *Applied Perception in Graphics and Visualization*, 1(212), 125–128.
- Mc Kone, E., Martini, P., & Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology: Human Perception and Performance*, 27(3), 573–599.
- McConachie, H. R. (1976). Developmental Prosopagnosia. A Single Case Report. *Cortex*, 12(1), 76–82. doi:10.1016/S0010-9452(76)80033-0

References

- McDonnell, R., Jörg, S., Hodgins, J. K., Newell, F., & O'Sullivan, C. (2009). Evaluating the effect of motion and body shape on the perceived sex of virtual characters. *ACM Transactions on Applied Perception*, 5(4), 1–14. doi:10.1145/1462048.1462051
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 263, 746–748. Retrieved from <http://www.nature.com/nature/journal/v264/n5588/abs/264746a0.html>
- McKone, E., Hall, A., Pidcock, M., Palermo, R., Wilkinson, R. B., Rivolta, D., ... O'Connor, K. B. (2011). Face ethnicity and measurement reliability affect face recognition performance in developmental prosopagnosia: evidence from the Cambridge Face Memory Test-Australian. *Cognitive Neuropsychology*, 28(2), 109–146. doi:10.1080/02643294.2011.616880
- McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences*, 11(1), 8–15. doi:10.1016/j.tics.2006.11.002
- Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: a meta-analytic review. *Psychology, Public Policy, and Law*, 7(1), 3–35. doi:10.1037//1076-8971.7.1.3
- Memon, A., Bartlett, J., Rose, R., & Gray, C. (2003). The aging eyewitness: effects of age on face, delay, and source-memory ability. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 58(6), P338–45. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/14614118>
- Moro, V., Pernigo, S., Avesani, R., Bulgarelli, C., Urgesi, C., Candidi, M., & Aglioti, S. M. (2012). Visual body recognition in a prosopagnosic patient. *Neuropsychologia*, 50(1), 104–117. doi:10.1016/j.neuropsychologia.2011.11.004
- Moro, V., Urgesi, C., Pernigo, S., Lanteri, P., Pazzaglia, M., & Aglioti, S. M. (2008). The neural basis of body form and body action agnosia. *Neuron*, 60(2), 235–346. doi:10.1016/j.neuron.2008.09.022
- Morrison, E. R., Gralewski, L., Campbell, N., & Penton-Voak, I. S. (2007). Facial movement varies by sex and is related to attractiveness. *Evolution and Human Behavior*, 28(3), 186–192. doi:10.1016/j.evolhumbehav.2007.01.001
- Moyse, E., Beaufort, A., & Brédart, S. (2014). Evidence for an own-age bias in age estimation from voices in older persons. *European Journal of Ageing*, 11(3), 241–247. doi:10.1007/s10433-014-0305-0
- Murray, J. E., Halberstadt, J., & Ruffman, T. (2010). The face of aging: sensitivity to facial feature relations changes with age. *Psychology and Aging*, 25(4), 846–850. doi:10.1037/a0019864
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4), 695–699.
- Németh, K., Zimmer, M., Schweinberger, S. R., Vakli, P., & Kovács, G. (2014). The background of reduced face specificity of n170 in congenital prosopagnosia. *PLoS ONE*, 9(7), e101393. doi:10.1371/journal.pone.0101393

- Newell, F. N., Chiroro, P., & Valentine, T. (1999). Recognizing unfamiliar faces: the effects of distinctiveness and view. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 52(2), 509–534. doi:10.1080/713755813
- Newell, F. N., Wallraven, C., & Huber, S. (2004). The role of characteristic motion in object categorization. *Journal of Vision*, 4, 118–129. doi:10.1167/4.2.5
- Norman, J. F., Payton, S. M., Long, J. R., & Hawkes, L. M. (2004). Aging and the perception of biological motion. *Psychology and Aging*, 19(1), 219–225. doi:10.1037/0882-7974.19.1.219
- Norton, D., McBain, R., & Chen, Y. (2009). Reduced ability to detect facial configuration in middle-aged and elderly individuals: Associations with spatiotemporal visual processing. *Journal of Gerontology: Psychological Sciences*, 64B(3), 328–334. doi:10.1093/geronb/gbp008
- Nunn, J. a, Postma, P., & Pearson, R. (2001). Developmental prosopagnosia: should it be taken at face value? *Neurocase*, 7(1), 15–27. doi:10.1093/neucas/7.1.15
- O'Mahony, C., & Newell, F. N. (2012). Integration of faces and voices, but not faces and names, in person recognition. *British Journal of Psychology*, 103(1), 73–82. doi:10.1111/j.2044-8295.2011.02044.x
- O'Toole, A. J., Natu, V., An, X., Rice, A., Ryland, J., & Phillips, P. J. (2014). The neural representation of faces and bodies in motion and at rest. *NeuroImage*, 91, 1–11. doi:10.1016/j.neuroimage.2014.01.038
- O'Toole, A. J., Phillips, P. J., Weimer, S., Roark, D. a, Ayyad, J., Barwick, R., & Dunlop, J. (2011). Recognizing people from dynamic and static faces and bodies: dissecting identity with a fusion approach. *Vision Research*, 51(1), 74–83. doi:10.1016/j.visres.2010.09.035
- O'Toole, A. J., & Roark, D. (2011). Memory for Moving Faces: The Interplay of Two Recognition Systems. In C. Curio, H. Bülthoff, M. A. Giese, & T. Poggio (Eds.), *Dynamic Faces: Insights from Experiments and Computation*. MIT Press.
- O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences*, 6(6), 261–266. doi:10.1016/S1364-6613(02)01908-3
- Otsuka, Y., Konishi, Y., Kanazawa, S., Yamaguchi, M. K., Abdi, H., & O'Toole, A. J. (2009). Recognition of moving and static faces by young infants. *Child Development*, 80(4), 1259–1271. doi:10.1111/j.1467-8624.2009.01330.x
- Owsley, C., Sekuler, R., & Boldt, C. (1981). Aging and low-contrast vision: face perception. *Investigative Ophthalmology & Visual Science*, 21(2), 362–365. Retrieved from <http://www.iovs.org/content/21/2/362.short>
- Palermo, R., Rivolta, D., Wilson, C. E., & Jeffery, L. (2011). Adaptive face space coding in congenital prosopagnosia: typical figural aftereffects but abnormal identity aftereffects. *Neuropsychologia*, 49(14), 3801–3812. doi:10.1016/j.neuropsychologia.2011.09.039
- Papcun, G., Kreiman, J., & Davis, A. (1989). Long-term memory for unfamiliar voices. *The Journal of the Acoustical Society of America*, 85(2), 913–925. doi:10.1121/1.397564

References

- Parvizi, J., Jacques, C., Foster, B. L., Withoft, N., Rangarajan, V., Weiner, K. S., & Grill-Spector, K. (2012). Electrical stimulation of human fusiform face-selective regions distorts face perception. *Journal of Neuroscience*, 32(43), 14915–14920. doi:10.1523/JNEUROSCI.2609-12.2012
- Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology*, 93(1), 603–608. doi:10.1152/jn.00513.2004
- Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience*, 8(8), 636–48. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17643089>
- Peiffer, A. M., Mozolic, J. L., Hugenschmidt, C. E., & Laurienti, P. J. (2007). Age-related multisensory enhancement in a simple audiovisual detection task. *Neuroreport*, 18(10), 1077–81. doi:10.1097/WNR.0b013e3281e72ae7
- Pernet, C. R., & Belin, P. (2012). The role of pitch and timbre in voice gender categorization. *Frontiers in Psychology*, 3(February), 23. doi:10.3389/fpsyg.2012.00023
- Petrini, K., & Pollick, F. (2010). Expertise with multisensory events eliminates the effect of biological motion rotation on audiovisual synchrony perception. *Journal of Vision*, 10(5), 1–14. doi:10.1167/10.5.2.Introduction
- Piepers, D. W., & Robbins, R. A. (2012). A review and clarification of the terms “holistic,” “configural,” and “relational” in the face perception literature. *Frontiers in Psychology*, 3, 559. doi:10.3389/fpsyg.2012.00559
- Pike, G. E., Kemp, R. I., Nicola, A. T., & Phillips, K. C. (1997). Recognizing moving faces : the relative contribution of motion and perspective view information. *Visual Cognition*, 4(4), 409–437. doi:10.1080/713756769
- Pilotti, M., Beyer, T., & Yasunami, M. (2001). Encoding Tasks and the Processing of Perceptual Information in Young and Older Adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 56(2), P119–P128. doi:10.1093/geronb/56.2.P119
- Pilotti, M., Meade, M. L., & Gallo, D. A. (2003). Implicit and explicit measures of memory for perceptual information in young adults, healthy older adults, and patients with alzheimer’s disease. *Experimental Aging Research*, 29, 15–32.
- Pilz K. S., Vuong, Q. C., Bülthoff, H. H., & Thornton, I. M. (2006). Body motion influences the processing of identity. *Perception*, 35 ECVF Abstract Supplement.
- Pilz, K. S., Bülthoff, H., & Vuong, Q. (2009). Learning influences the encoding of static and dynamic faces and their recognition across different spatial frequencies. *Visual Cognition*, 17(5), 716–735. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/13506280802340588>
- Pilz, K. S., Bennett, P. J., & Sekuler, A. B. (2010). Effects of aging on biological motion discrimination. *Vision Research*, 50(2), 211–219. doi:10.1016/j.visres.2009.11.014

- Pilz, K. S., Konar, Y., Vuong, Q. C., Bennett, P. J., & Sekuler, A. B. (2011). Age-related changes in matching novel objects across viewpoints. *Vision Research*, 51(17), 1958–65. doi:10.1016/j.visres.2011.07.009
- Pilz, K. S., Thornton, I. M., & Bülthoff, H. H. (2006). A search advantage for faces learned in motion. *Experimental Brain Research*, 171(4), 436–47. doi:10.1007/s00221-005-0283-8
- Pilz, K. S., Vuong, Q. C., Bülthoff, H. H., & Thornton, I. M. (2011). Walk this way: approaching bodies can influence the processing of faces. *Cognition*, 118(1), 17–31. doi:10.1016/j.cognition.2010.09.004
- Pitcher, D. (2014). Facial expression recognition takes longer in the posterior superior temporal sulcus than in the occipital face area. *The Journal of Neuroscience*, 34(27), 9173–9177. doi:10.1523/JNEUROSCI.5038-13.2014
- Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage*, 56(4), 2356–2363. doi:10.1016/j.neuroimage.2011.03.067
- Pitcher, D., Duchaine, B., & Walsh, V. (2014). Combined TMS and fMRI reveal dissociable cortical pathways for dynamic and static face perception. *Current Biology*, 24(17), 2066–2070. doi:10.1016/j.cub.2014.07.060
- Pollick, F. E., Kay, J. W., Heim, K., & Stringer, R. (2005). Gender recognition from point-light walkers. *Journal of Experimental Psychology: Human Perception and Performance*, 31(6), 1247–1265. doi:10.1037/0096-1523.31.6.1247
- Post, R. B., Haberman, J., Iwaki, L., & Whitney, D. (2012). The frozen face effect: why static photographs may not do you justice. *Frontiers in Psychology*, 3(22). doi:10.3389/fpsyg.2012.00022
- Potamianos, G., Neti, C., Gravier, G., Garg, A., Member, S., & Senior, A. W. (2003). Recent advances in the automatic recognition of audio-visual speech. In *Proceedings of the IEE* (pp. 1306–1326).
- Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *The Journal of Neuroscience*, 18(6), 2188–2199. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9482803>
- Rabe-Hesketh, S., Skrondal, A., & Gjessing, H. K. (2008). Biometrical modeling of twin and family data using standard mixed model software. *Biometrics*, 64(1), 280–288. doi:10.1111/j.1541-0420.2007.00803.x
- Read, D., & Craik, F. I. M. (1995). Earwitness identification: some influences on voice recognition. *Journal of Experimental Psychology: Applied*, 1(1), 6–18. doi:10.1037//1076-898X.1.1.6
- Rhodes, G., & Jaquet, E. (2011). Aftereffects reveal that adaptive face-coding mechanisms are selective for race and sex. In R. B. Adams, N. Ambady, K. Nakayama, & S. Shimojo (Eds.), *The science of social vision*. Oxford, UK: Oxford University Press.

References

- Rhodes, G., Jeffery, L., Boeing, A., & Calder, A. J. (2013). Visual coding of human bodies: perceptual aftereffects reveal norm-based, opponent coding of body identity. *Journal of Experimental Psychology. Human Perception and Performance*, 39(2), 313–317. doi:10.1037/a0031568
- Rhodes, M. G., & Anastasi, J. S. (2011). The own-age bias in face recognition: a meta-analytic and theoretical review. *Psychological Bulletin*, 138(1), 146–174. doi:10.1037/0882-7974.23.2.239
- Rhodes, M. G., Castel, A. D., & Jacoby, L. L. (2008). Associative recognition of face pairs by younger and older adults: The role of familiarity-based processing. *Psychology and Aging*, 23(2), 239–249. doi:10.1037/0882-7974.23.2.239
- Richoz, A., Jack, R. E., Garrod, O. G. B., Schyns, P. G., & Caldara, R. (2014). Reconstructing emotions in motion in prosopagnosia reveals separate representations for identity and emotion. *Perception 43 ECPV Abstract Supplement*, 123.
- Riddoch, M. J., & Humphreys, G. W. (1993). *Birmingham Object Recognition Battery (BORB)*. Hove: UK: Lawrence Erlbaum Associates.
- Righart, R., & de Gelder, B. (2007). Impaired face and body perception in developmental prosopagnosia. *Proceedings of the National Academy of Sciences of the United States of America*, 104(43), 17234–17238. doi:10.1073/pnas.0707753104
- Rivolta, D., Palermo, R., & Schmalzl, L. (2013). What is overt and what is covert in congenital prosopagnosia? *Neuropsychology Review*, 23(2), 111–116. doi:10.1007/s11065-012-9223-0
- Roark, D. A., Barrett, S. E., Spence, M., Abdi, H., & O'Toole, A. J. (2003). Memory for moving faces: Psychological and neural perspectives on the role of motion in face recognition. *Behavioral and Cognitive Neuroscience Reviews*, 2(1), 15–46.
- Robertson, D. M. C., & Schweinberger, S. R. (2010). The role of audiovisual asynchrony in person recognition. *Quarterly Journal of Experimental Psychology (2006)*, 63(1), 23–30. doi:10.1080/17470210903144376
- Rosenblum, L. D., Smith, N. M., Nichols, S. M., Hale, S., & Lee, J. (2006). Hearing a face: cross-modal speaker matching using isolated visible speech. *Perception & Psychophysics*, 68(1), 84–93. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/16617832>
- Rosenblum, L. D., Yakel, D. A., Baseer, N., Panchal, A., Nodarse, B. C., & Niehus, R. P. (2002). Visual speech information for face recognition. *Perception & Psychophysics*, 64(2), 220–229. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12013377>
- Ross, L. a, Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, 17(5), 1147–1153. doi:10.1093/cercor/bhl024
- Rossion, B., Caldara, R., Seghier, M., Schuller, A., Lazeyras, F., & Mayer, E. (2003). A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain*, 126(Pt 11), 2381–2395. doi:10.1093/brain/awg241

- Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., & Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychological Science*, 13(3), 250–257. doi:10.1111/1467-9280.00446
- Roudaia, E., Bennett, P. J., Sekuler, A. B., & Pilz, K. S. (2010). Spatiotemporal properties of apparent motion perception and aging. *Journal of Vision*, 10(14), 1–15. doi:10.1167/10.14.5.
- Russ, J. B., Gur, R. C., & Bilker, W. B. (2008). Validation of affective and neutral sentence content for prosodic testing. *Behavior Research Methods*, 40(4), 935–939. doi:10.3758/BRM.40.4.935
- Russell, R., Chatterjee, G., & Nakayama, K. (2012). Developmental prosopagnosia and super-recognition: no special role for surface reflectance processing. *Neuropsychologia*, 50(2), 334–340. doi:10.1016/j.neuropsychologia.2011.12.004
- Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, 16(2), 252–257. doi:10.3758/PBR.16.2.252
- Saygin, A. P., Driver, J., & de Sa, V. R. (2008). In the footsteps of biological motion and multisensory perception enhanced for upright walkers. *Psychological Science*, 19(5), 469–475.
- Schall, S., & von Kriegstein, K. (2014). Functional connectivity between face-movement and speech-intelligibility areas during auditory-only speech perception. *PloS One*, 9(1), e86325. doi:10.1371/journal.pone.0086325
- Schiltz, C., Sorger, B., Caldara, R., Ahmed, F., Mayer, E., Goebel, R., & Rossion, B. (2006). Impaired face discrimination in acquired prosopagnosia is associated with abnormal response to individual faces in the right middle fusiform gyrus. *Cerebral Cortex*, 16(4), 574–586. doi:10.1093/cercor/bhj005
- Schmalzl, L., Palermo, R., & Coltheart, M. (2008). Cognitive heterogeneity in genetically based prosopagnosia: a family study. *Journal of Neuropsychology*, 2(1), 99–117. doi:10.1348/174866407X256554
- Schmalzl, L., Palermo, R., Harris, I. M., & Coltheart, M. (2009). Face inversion superiority in a case of prosopagnosia following congenital brain abnormalities: what can it tell us about the specificity and origin of face-processing mechanisms? *Cognitive Neuropsychology*, 26(3), 286–306. doi:10.1080/02643290903086904
- Schouten, B., Troje, N. F., Vroomen, J., & Verfaillie, K. (2011). The effect of looming and receding sounds on the perceived in-depth orientation of depth-ambiguous biological motion figures. *PloS ONE*, 6(2), e14725. doi:10.1371/journal.pone.0014725
- Schultz, J., Brockhaus, M., Bülthoff, H. H., & Pilz, K. S. (2013). What the human brain likes about facial motion. *Cerebral Cortex*, 23(5), 1167–1178. doi:10.1093/cercor/bhs106
- Schultz, J., & Pilz, K. S. (2009). Natural facial motion enhances cortical responses to faces. *Experimental Brain Research*, 194(3), 465–475. doi:10.1007/s00221-009-1721-9

References

- Schvartz, K. C., & Chatterjee, M. (2012). Gender identification in younger and older adults: use of spectral and temporal cues in noise-vocoded speech. *Ear and Hearing, 33*(3), 411–420. doi:10.1097/AUD.0b013e31823d78dc
- Schweinberger, S. R., Casper, C., Hauthal, N., Kaufmann, J. M., Kawahara, H., Kloth, N., ... Zäske, R. (2008). Auditory adaptation in voice perception. *Current Biology, 18*(9), 684–688. doi:10.1016/j.cub.2008.04.015
- Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G., & Zäske, R. (2014). Speaker perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 5*(1), 15–25. doi:10.1002/wcs.1261
- Schweinberger, S. R., Kloth, N., & Robertson, D. M. C. (2011). Hearing facial identities: brain correlates of face-voice integration in person identification. *Cortex, 47*(9), 1026–1037. doi:10.1016/j.cortex.2010.11.011
- Schweinberger, S. R., Robertson, D., & Kaufmann, J. M. (2007). Hearing facial identities. *The Quarterly Journal of Experimental Psychology, 60*(10), 1446–1456. doi:10.1080/17470210601063589
- Schweinberger, S. R., Walther, C., Zäske, R., & Kovács, G. (2011). Neural correlates of adaptation to voice identity. *British Journal of Psychology, 102*(4), 748–764. doi:10.1111/j.2044-8295.2011.02048.x
- Setti, A., Burke, K. E., Kenny, R., & Newell, F. N. (2013). Susceptibility to a multisensory speech illusion in older persons is driven by perceptual processes. *Frontiers in Psychology, 4*, 575. doi:10.3389/fpsyg.2013.00575
- Setti, A., & Newell, F. N. (2010). The effect of body and part-based motion on the recognition of unfamiliar objects. *Visual Cognition, 18*(3), 456–480. doi:10.1080/13506280902830561
- Sheffert, S. M., & Olson, E. (2004). Audiovisual speech facilitates voice learning. *Perception & Psychophysics, 66*(2), 352–362. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15129754>
- Smith, E. L., Grabowecky, M., & Suzuki, S. (2007). Auditory-visual crossmodal integration in perception of face gender. *Current Biology, 17*(19), 1680–1685. doi:10.1016/j.cub.2007.08.043
- Snowden, R. J., & Kavanagh, E. (2006). Motion perception in the ageing visual system: Minimum motion, motion coherence, and speed discrimination thresholds. *Perception, 35*(1), 9–24. doi:10.1068/p5399
- Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear and Hearing, 26*(3), 263–275. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15937408>
- Steede, L. L., Tree, J. J., & Hole, G. J. (2007a). Dissociating mechanisms involves in accessing identity by dynamic and static cues. *Visual Cognition, 15*(1), 116–119. doi:10.1080/13506280600975486

- Steede, L. L., Tree, J. J., & Hole, G. J. (2007b). I can't recognize your face but I can recognize its movement. *Cognitive Neuropsychology*, 24(4), 451–466. doi:10.1080/02643290701381879
- Stevenage, S. V., Hale, S., Morgan, Y., & Neil, G. J. (2014). Recognition by association: Within- and cross-modality associative priming with faces and voices. *British Journal of Psychology*, 105(1), 1–16. doi:10.1111/bjop.12011
- Stevenage, S. V., Howland, A., & Tippelt, A. (2011). Interference in eyewitness and earwitness recognition. *Applied Cognitive Psychology*, 25(1), 112–118. doi:10.1002/acp.1649
- Stevenage, S. V., Hugill, A. R., & Lewis, H. G. (2012). Intergrating voice recognition into models of face perception. *Journal of Cognitive Psychology*, 24(4), 409–419.
- Stevenage, S. V., Neil, G. J., Barlow, J., Dyson, A., Eaton-Brown, C., & Parsons, B. (2013). The effect of distraction on face and voice recognition. *Psychological Research*, 77(2), 167–175. doi:10.1007/s00426-012-0450-z
- Stevenage, S. V., Nixon, M. S., & Vince, K. (1999). Visual analysis of gait as a cue to identity. *Applied Cognitive Psychology*, 13, 513–526.
- Sumby, W., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212–215.
- Susilo, T., & Duchaine, B. (2013). Advances in developmental prosopagnosia research. *Current Opinion in Neurobiology*, 23(3), 423–429. doi:10.1016/j.conb.2012.12.011
- Taylor, J. C., Wiggett, A. J., & Downing, P. E. (2007). Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas. *Journal of Neurophysiology*, 98(3), 1626–1633. doi:10.1152/jn.00012.2007
- Thomas, J. P., & Shiffrar, M. (2010). I can see you better if I can hear you coming: action-consistent sounds facilitate the visual detection of human gait. *Journal of Vision*, 10(12), 1–11. doi:10.1167/10.12.14
- Thomas, J. P., & Shiffrar, M. (2013). Meaningful sounds enhance visual sensitivity to human gait regardless of synchrony. *Journal of Vision*, 13, 1–13. doi:10.1167/13.14.8
- Thornton, I. M. (2006). Biological motion: Point-light walkers and beyond. In G. Knoblich, I. M. Thornton, M. Grosjean, & M. Shiffrar (Eds.), *Human Body Perception From The Inside Out*. Oxford, UK: Oxford University Press.
- Thornton, I. M., & Kourtzi, Z. (2002). A matching advantage for dynamic human faces. *Perception*, 31(1), 113–132. doi:10.1068/p3300
- Thornton, I., Mullins, E., & Banahan, K. (2011). Motion can amplify the face-inversion effect. *Psychologia*, 44(1), 5–22. doi:10.2298/PSI1101005T
- Toga, A. W., & Thompson, P. M. (2005). Genetics of brain structure and intelligence. *Annual Review of Neuroscience*, 28, 1–23. doi:10.1146/annurev.neuro.28.061604.135655

References

- Towler, J., & Eimer, M. (2012). Electrophysiological studies of face processing in developmental prosopagnosia: neuropsychological and neurodevelopmental perspectives. *Cognitive Neuropsychology*, 29(5-6), 503–529. doi:10.1080/02643294.2012.716757
- Towler, J., Gosling, A., Duchaine, B., & Eimer, M. (2012). The face-sensitive N170 component in developmental prosopagnosia. *Neuropsychologia*, 50(14), 3588–3599. doi:10.1016/j.neuropsychologia.2012.10.017
- Tranel, D., Damasio, A. R., & Damasio, H. (1988). Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity. *Neurology*, 38(5), 690–696. doi:10.1212/WNL.38.5.690
- Troje, N. F. (2002). Decomposing biological motion: a framework for analysis and synthesis of human gait patterns. *Journal of Vision*, 2(5), 371–387. doi:10.1167/2.5.2
- Troje, N. F. (2008). Retrieving Information from Human Movement Patterns. In T. F. Shipley & J. M. Zacks (Eds.), *Understanding Events: From Perception to Action*. Oxford University Press.
- Troje, N. F., Sadr, J., Geyer, H., & Nakayama, K. (2006). Adaptation aftereffects in the perception of gender from biological motion. *Journal of Vision*, 6(8), 850–857. doi:10.1167/6.8.7
- Troje, N. F., Westhoff, C., & Lavrov, M. (2005). Person identification from biological motion: effects of structural and kinematic cues. *Perception & Psychophysics*, 67(4), 667–675. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/16134460>
- Tye-Murray, N., Spehar, B., Myerson, J., Sommers, M. S., & Hale, S. (2011). Cross-modal enhancement of speech detection in young and older adults: does signal content matter? *Ear and Hearing*, 32(5), 650–655. doi:10.1097/AUD.0b013e31821a4578
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A : Human Experimental Psychology*, 43(2), 161–204.
- Valentine, T., & Ferrara, A. (1991). Typicality in categorization, recognition and identification: evidence from face recognition. *British Journal of Psychology*, 82, 87–102. doi:10.1111/j.2044-8295.1991.tb02384.x
- Van der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: Validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion*, 11(4), 907–920. doi:10.1037/a0023853
- Van der Zwan, R., Machatch, C., Kozlowski, D., Troje, N. F., Blanke, O., & Brooks, A. (2009). Gender bending: auditory cues affect visual judgements of gender in biological motion displays. *Experimental Brain Research*, 198(2-3), 373–382. doi:10.1007/s00221-009-1800-y
- Van Lancker, D. R., & Canter, G. J. (1982). Impairment of voice and face recognition in patients with hemispheric damage. *Brain and Cognition*, 1(2), 185–195. doi:10.1016/0278-2626(82)90016-1

- Vangeneugden, J., Peelen, M. V., Tadin, D., & Battelli, L. (2014). Distinct neural mechanisms for body form and body motion discriminations. *The Journal of Neuroscience*, 34(2), 574–585. doi:10.1523/JNEUROSCI.4032-13.2014
- Vizioli, L., Rousselet, G. A., & Caldara, R. (2010). Neural repetition suppression to identity is abolished by other-race faces. *Proceedings of the National Academy of Sciences of the United States of America*, 107(46), 20081–20086. doi:10.1073/pnas.1005751107
- von Kriegstein, K., Dogan, Ö., Grüter, M., Giraud, A.-L., Kell, C. A., Grüter, T., ... Kiebel, S. J. (2008). Simulation of talking faces in the human brain. *Proceedings of the National Academy of Sciences*, 105(18), 6747–6752.
- von Kriegstein, K., & Giraud, A.-L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology*, 4(10), e326. doi:10.1371/journal.pbio.0040326
- von Kriegstein, K., Kleinschmidt, A., & Giraud, A.-L. (2006). Voice recognition and cross-modal responses to familiar speakers' voices in prosopagnosia. *Cerebral Cortex*, 16(9), 1314–1322. doi:10.1093/cercor/bhj073
- von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A.-L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience*, 17(3), 367–376. doi:10.1162/0898929053279577
- Wallis, G. (1998). Temporal order in human object recognition learning. *Journal of Biological Systems*, 6(3), 299–313. doi:10.1142/S0218339098000200
- Wallis, G., & Bühlhoff, H. (1999). Learning to recognize objects. *Trends in Cognitive Sciences*, 3(1), 22–31. doi:10.1016/S1364-6613(98)01261-3
- Wallis, G., & Bühlhoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences*, 98(8), 4800–4804. doi:10.1073/pnas.071028598
- Westhoff, C., & Troje, N. F. (2007). Kinematic cues for person identification from biological motion. *Perception & Psychophysics*, 69(2), 241–253. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17557594>
- Wickham, L. H., Morris, P. E., & Fritz, C. O. (2000). Facial distinctiveness: its measurement, distribution and influence on immediate and delayed recognition. *British Journal of Psychology*, 91, 99–123. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10717774>
- Wiese, H., Komes, J., & Schweinberger, S. R. (2012). Daily-life contact affects the own-age bias and neural correlates of face memory in elderly participants. *Neuropsychologia*, 50(14), 3496–3508. doi:10.1016/j.neuropsychologia.2012.09.022
- Wilhelm, O., Herzmann, G., Kunina, O., Danthiir, V., Schacht, A., & Sommer, W. (2010). Individual differences in perceiving and recognizing faces-One element of social cognition. *Journal of Personality and Social Psychology*, 99(3), 530–548. doi:10.1037/a0019972

References

- Wilhelm, O., Herzmann, G., Kunina, O., & Sommer, W. (2007). Face cognition: a set of distinct mental abilities. *Nature Precedings*, Available from Nature Precedings at <http://hdl.han.> Retrieved from <http://hdl.handle.net/10101/npre.2007.1385.1>
- Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., ... Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences of the United States of America*, 107(11), 5238–5241. doi:10.1073/pnas.0913053107
- Winston, J. S., Henson, R. N. A., Fine-Goulden, M. R., & Dolan, R. J. (2004). fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *Journal of Neurophysiology*, 92(3), 1830–1839. doi:10.1152/jn.00155.2004
- Wuerger, S. M., Crocker-Buque, A., & Meyer, G. F. (2012). Evidence for auditory-visual processing specific to biological motion. *Seeing and Perceiving*, 25(1), 15–28. doi:10.1163/187847611X620892
- Xiao, N. G., Perrotta, S., Quinn, P. C., Wang, Z., Sun, Y.-H. P., & Lee, K. (2014). On the facilitative effects of face motion on face recognition and its development. *Frontiers in Psychology*, 5, 633. doi:10.3389/fpsyg.2014.00633
- Xiao, N. G., Quinn, P. C., Ge, L., & Lee, K. (2012). Rigid facial motion influences featural, but not holistic, face processing. *Vision Research*, 57, 26–34. doi:10.1016/j.visres.2012.01.015
- Yin, R. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81(1), 141–145. doi:10.1037/h0027474
- Yonan, C. a, & Sommers, M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychology and Aging*, 15(1), 88–99. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10755292>
- Young, A. W., DeHaan, E. H. F., & Bauer, R. M. (2008). Face perception: A very special issue. *Journal of Neuropsychology*, 2(1), 1–14. doi:10.1348/174866407X269848
- Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configural information in face perception. *Perception*, 16(6), 747–759. doi:10.1068/p160747
- Zäske, R., & Schweinberger, S. R. (2011). You are only as old as you sound: auditory aftereffects in vocal age perception. *Hearing Research*, 282(1-2), 283–8. doi:10.1016/j.heares.2011.06.008
- Zäske, R., Schweinberger, S. R., & Kawahara, H. (2010). Voice aftereffects of adaptation to speaker identity. *Hearing Research*, 268(1-2), 38–45. doi:10.1016/j.heares.2010.04.011
- Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., ... Liu, J. (2010). Heritability of the specific cognitive ability of face perception. *Current Biology*, 20(2), 137–142. doi:10.1016/j.cub.2009.11.067

9. Appendices

9.1. Appendix A

General ethics approval letter from the School of Psychology, Trinity College Dublin¹⁴.



¹⁴ Note that this is a general letter of approval issued to the Multisensory Cognition Research Group for a large battery of multisensory tests involving younger and older participants. Although addressed to 'Niamh Merriman' a number of names, including but not limited to, Corrina Maguinness, were listed on the application. Approval was granted for *all* listed names.

9.2. Appendix B

Ethics approval letter from the School of Psychology, Trinity College Dublin, for experiments run in the Science Gallery, Trinity College Dublin.



COLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE DUBLIN
Ollscoil Átha Cliath | The University of Dublin

1st February 2010

F.A.O. Corrina Maguinness

School of Psychology Research Ethics Committee

Dear Corrina,

I am pleased to inform you that your application entitled "Walk right into my life" has been approved by the School of Psychology Research Ethics Committee.

Yours sincerely,

Dr. Tim Trimble
Chair
School of Psychology Research Ethics Committee

SCHOOL OF PSYCHOLOGY
Aras an Phairsaigh
Trinity College
Dublin 2

Scoil na Tríonóide,
Ollscoil Átha Cliath, Baile Átha Cliath
Aras an Phairsaigh, Coláiste na Tríonóide,
Baile Átha Cliath 2, Éire

School of Psychology,
Faculty of Arts, Humanities and Social Sciences,
Aras an Phairsaigh, Trinity College,
Dublin 2, Ireland

T: 353 (0)1 896 1886
F: 353 (0)1 671 2006
psychology@tcd.ie
www.psychology.tcd.ie

'Sensory Cue Interactions in Person Perception: Insights from Ageing and Developmental Prosopagnosia' Corrina Maguinness

A large body of behavioural and neuropsychological research, using static images of faces, has informed our understanding of the mechanisms underlying identity processing. Indeed, pioneering studies revealed dedicated modules for the processing of facial form in the human brain. Yet interactions with people in the real world are accompanied by a greater variety of sensory information beyond visual images, for example, people move and speak and are thus highly dynamic, multisensory stimuli. While these additional cues (e.g. facial motion, body motion and the voice) have received attention from researchers in the field, they have largely been studied in isolation, possibly driven by assumptions that each identity cue was processed independently. However, recent findings that facial form modules in the brain also respond to facial movement and to voice information (in the absence of a facial image) have sparked a re-examination of how these sensory cues may *interact* to support person perception. To date, assertions as to how these cues may interact have been largely based on observations from neurotypical younger adult studies (reviewed in detail in Chapter 1).

The aim of this thesis is to extend our understanding of how multiple sensory cues may interact in person perception through the study of three different cohorts: younger adults, older adults and developmental prosopagnosics (DP). This thesis argues that the unique approach of examining cohorts where there is a decline (older age) or a disorder (DP) in static face processing, may give rise to a more concrete understanding of how static and dynamic cues interact to support normal person perception. For example, it is poorly understood how impaired face processing, as found in cases of DP, may impact on the processing of other perceptual information e.g. perceiving the voice, or whether perception in an older person may benefit from multisensory cue combination, as processing in one modality declines. Such findings would clarify how the interactions between the senses may shape our ability to represent a person in memory. A number of behavioural studies, described in this thesis, examined this issue by centring on three main themes: 1) the interaction of form and motion cues in face perception; 2) the interaction of facial motion and the voice in person perception; and 3) the interaction between dynamic body cues and the voice in person perception.

The first theme involved studies which confirmed and expanded on previous findings of poorer static face processing in ageing and developmental prosopagnosia, using a battery of standardised face processing tests (Chapter 2). The tests revealed that older adults showed poor performance for recognising facial identity across changes in the visual appearance of the face; although this pattern was particularly exaggerated for the two cases of DP. Using a series of matching experiments reported in Chapter 3, the results demonstrated that learning a face in motion can lead to more accurate matching of facial identity across changes in the viewpoint and expression of the face, in older, relative to younger adults. This finding demonstrates that these two sensory cues, form and motion, are readily integrated and can, as a consequence, enhance the representation of the face in memory. The results of the experiments reported in Chapter 4 suggest evidence that certain types of facial motion resulted in poor matching of facial identity in the DP group, suggesting that motion may have distracted from the encoding of static form cues in faces. This finding adds support to the idea that the interactions between form and motion in face processing may be mandatory, as the ability to use motion cues may also be compromised in disorders of face recognition.

The second theme focused on interactions between the face and voice for person perception. Using a series of matching tasks, evidence reported in Chapter 5 revealed that relative to younger adults, older adults were poorer at matching identity from vocal cues alone. Interestingly, this pattern was exaggerated in one case of DP, suggesting that atypical face processing may impact on how the voice is processed to support identity. In addition, the results of Chapter 5 demonstrated that older adults', and to some extent younger adults', ability to represent the face and the voice as a single coherent percept, was likely mediated by the presence of dynamic, relative to static, cross-modal cues. These results support cross-modal sensory interactions for the purpose of representing identity.

The third theme concentrated on interactions between the voice and the body in person perception. Using a sex categorisation task reported in Chapter 6, the results demonstrated that sex cues present in the voice could modulate the perception of 'gender-ambiguous' biological motion displays in younger adults and in DP. However, in general older adults did not appear to benefit from voice information in the same way as younger adults, suggesting that sensory integration for less bound couplings e.g. the body and the voice, compared to the face and the voice, may be less likely to be observed in older age.

Taken together the main findings from this collection of studies support and inform emerging models of person perception and suggest that dynamic, cross-modal cues are integrated for the purpose of person perception at earlier stages of processing than previously assumed. As a consequence person perception can be enhanced or at least modulated by these interactions. The theoretical implications of these results are discussed in detail in Chapter 7.