

KATHOLIEKE UNIVERSITEIT LEUVEN

Faculteit Psychologie en Pedagogische Wetenschappen

Laboratorium voor Experimentele Psychologie

Visual in-depth perception and audiovisual perception of biological motion

Proefschrift aangeboden
tot het verkrijgen van de graad van
Doctor in de Psychologie
door **Ben Schouten**
o.l.v. Prof. Dr. Karl Verfaillie

2010

Ben Schouten *Visual in-depth perception and audiovisual perception of biological motion*

Dissertation submitted to obtain the degree of Doctor in Psychology, 2010
Supervisor: prof. dr. Karl Verfaillie

The process in which the human brain extracts socially and behaviorally relevant information and forms a vivid and meaningful interpretation of another person and his or her actions is a matter of tenths of a second. Remarkably, this amazing perceptual capacity persists, even when the visual information is restricted to a number of point lights representing some of the joints of that person (Johansson, 1973). This observation has lead researchers to employ these so-called 'point-light figures' to investigate human action perception.

The present dissertation consists of two main parts, in both of which the point-light figure (plf), more specifically a point-light walker (plw), is employed as a visual stimulus. In Part I, we study the in-depth perception of point-light walkers (plws). In Part II, we study the audiovisual perception of biological motion.

One of the interesting properties of a plw is that, although its 2D orthographic projection in principle is consistent with an infinite number of 3D configurations, only two interpretations are consistent with an anatomically plausible body configuration. Vanrie, Dekeyser, and Verfaillie (2004) showed that the human perceptual system indeed extracts both veridical interpretations, i.e., that this stimulus is perceptually bistable. More intriguingly, however, was the finding that the perceptual system seems to have a preference for one of the two interpretations. Indeed, whereas a non-lateral 2D orthographic projection of a 3D plw is always equally consistent with a figure that is facing towards the viewer and a figure that is facing away from the viewer, the brain favors the 'facing towards' interpretation in most of the cases. One of the potential causes of this so-called 'facing bias' could be the fact that a person who is oriented towards the viewer, is socially or behaviorally more relevant.

In Part I of this dissertation we present five chapters reporting research on some factors that affect the perceived in-depth orientation of depth-ambiguous plws. In Chapter 1, we start with a general introduction on research on the perception of biological motion and we focus on the relatively young subdomain studying the in-depth perception of point-light figures (plfs). In Chapter 2, we present a tool that we developed to manipulate, with perspective cues, the perceived in-depth orientation and to determine the point of subjective ambiguity of plfs, i.e., the amount of perspective information a particular observer needs in order to perceive the plf as facing towards him or her in 50% of the trials. In Chapter 3, we report a study showing a strong relation between the perceived gender (a variable that may reflect social or behavioral relevance) and the perceived in-depth orientation of a plw. In Chapter 4, on top of the so-called figure gender effect, we report an observer sex effect, that is, an interaction between figure gender and observer sex. The data presented in Chapter 5 show that the relation between perceived gender and perceived in-depth orientation is not causal. Plws that only vary in structural gender information elicit an opposite figure gender effect compared to plws that only vary in kinematic gender information. Moreover, we also show that the perceived in-depth orientation of the upper figure part is opposite to that of the lower figure part and that the lower part is very likely to be responsible for the facing bias. In the closing chapter of Part I, Chapter 6, we refine our perspective technique and show that the observer sex effect, reported in Chapter 4, can also be observed when the changes in perceived in-depth orientation are induced by perspective cues.

Traditionally, perception researchers have studied the senses in isolation. Actually, this is somewhat surprising because in daily life the senses are bombarded with countless simultaneous stimulations. This is especially true for the perception of actions for which often, not only the visual appearance, but also the sound of the action is available. In Part II, we study the audiovisual perception of biological motion. Chapter 7 is an introductory chapter, in which we give a selective overview of research on audiovisual perception in general and, more specifically, of recent neurophysiological and behavioral research on audiovisual perception of biological motion. In the remaining chapters we present empirical research investigating whether sound affects visual biological motion perception. Chapter 8 describes an adjustment technique that we developed to obtain an estimate of the point of subjective simultaneity, i.e., the point at which an auditory and visual stream are perceived in synchrony. In Chapter 9 and Chapter 10, respectively, we explore whether sounds can facilitate the visual detection of a plf and whether auditory and visual information on the speed of walking are integrated in a statistically optimal fashion. In Chapter 11, we present a study showing that the perceived in-depth orientation of a plw can be affected by looming or receding footstep sounds.

To close the dissertation, for Part I and Part II separately, our findings are summarized, conclusions are formulated, and suggestions for future research are provided.

Ben Schouten *Visuele diepteperceptie en audiovisuele perceptie van biologische beweging*

Proefschrift aangeboden tot het verkrijgen van de graad van Doctor in de Psychologie, 2010
Leiding: prof. dr. Karl Verfaillie

Het proces waarin de menselijke hersenen sociale en gedragsmatig relevante informatie onttrekken en een levendige en betekenisvolle interpretatie maken van een andere persoon en zijn of haar acties is een kwestie van tienden van een seconde. Deze merkwaardige perceptuele capaciteit houdt stand, zelfs als de visuele informatie beperkt is tot een aantal puntlichten die enkele gewrichten van die persoon voorstellen (Johansson, 1973). Deze bevinding heeft onderzoekers er toe aangezet om deze zogenaamde 'puntlicht-figuren' te gebruiken voor het onderzoeken van menselijke actieperceptie.

Dit proefschrift bestaat uit twee delen. In beide delen werd de puntlicht-figuur (plf), meer specifiek de puntlichtwandelaar (plw), gebruikt als visuele stimulus. In Deel I bestuderen we de diepte-perceptie van puntlichtwandelaars (plws). In Deel II bestuderen we de audiovisuele perceptie van biologische beweging.

Eén van de interessante eigenschappen van een plw is dat, ondanks het feit dat een 2D orthografische projectie consistent is met een oneindig aantal 3D configuraties, slechts twee interpretaties consistent zijn met een anatomisch plausibele lichaamsconfiguratie. Vanrie, Dekeyser, en Verfaillie (2004) toonden aan dat het perceptueel systeem inderdaad beide interpretaties vormt en dat deze stimulus dus perceptueel bistabiel is. Nog intrigerender was de bevinding dat het perceptueel systeem een voorkeur heeft voor één van de twee interpretaties. Hoewel een niet-laterale 2D orthografische projectie van een 3D plw altijd consistent is met een figuur die naar de observeerder toe en van de observeerder weg kijkt, hebben de hersenen in de meeste gevallen de voorkeur voor de 'naar de observeerder toe' interpretatie. Een mogelijke reden voor deze zogenaamde 'facing bias' zou zijn dat een persoon die naar de observeerder toe georiënteerd is sociaal en gedragsmatig relevanter is.

In Deel I van dit proefschrift presenteren we vijf hoofdstukken die onderzoek rapporteren over enkele factoren die de diepte-interpretatie van plws beïnvloeden. In Hoofdstuk 1 geven we een algemene introductie over onderzoek naar de perceptie van biologische beweging en focussen we op het relatief jonge subdomein die de diepte-perceptie van puntlichtfiguren (plfn) bestudeert. In Hoofdstuk 2 presenteren we een methode die we ontwikkelden om, met perspectief cues, de waargenomen diepte-oriëntatie te manipuleren en het punt van subjectieve ambiguiteit van plfn te bepalen, dat is de hoeveelheid perspectief die een observeerder nodig heeft om een plf als naar zich toe te zien in 50% van de proefbeurten. In Hoofdstuk 3 rapporteren we een studie die een sterk verband aantoont tussen het waargenomen geslacht (een variabele die sociale of gedragsmatige relevantie zou kunnen weerspiegelen) en de waargenomen diepte-oriëntatie van de plw. In Hoofdstuk 4 rapporteren we bovenop het zogenaamde figuur-geslachts-effect een observeerder-geslachts-effect, meer specifiek, een interactie tussen het geslacht van de figuur en het geslacht van de observeerder. De data gerapporteerd in Hoofdstuk 5 tonen aan dat de relatie tussen waargenomen geslacht en waargenomen diepte-oriëntatie niet oorzakelijk is. Plws die enkel variëren in structurele geslachtsinformatie lokken een omgekeerd figuur-geslachts-effect uit vergeleken met plws die enkel variëren in kinematische geslachtsinformatie. Bovendien tonen we aan dat de waargenomen diepte-oriëntatie van de bovenste figuurhelft omgekeerd is in vergelijking met die van de onderste figuurhelft en dat de onderste figuurhelft hoogst waarschijnlijk verantwoordelijk is voor de facing bias. In het laatste hoofdstuk van Deel I, Hoofdstuk 6, verfijnen we onze perspectief-techniek en tonen we dat het observeerder-geslachts-effect, gerapporteerd in Hoofdstuk 4 ook kan optreden als de veranderingen in waargenomen diepte-oriëntatie gedreven worden door perspectief cues.

Traditioneel hebben perceptie-onderzoekers de zintuigen afzonderlijk bestudeerd. Eigenlijk is dit verwonderlijk omdat in het dagelijkse leven de zintuigen geconfronteerd worden met ontelbare gelijktijdige stimulaties. Dit geldt in het bijzonder voor de perceptie van acties waarbij, vaak niet alleen visuele informatie, maar ook geluid perceptueel beschikbaar is. In Deel II bestuderen we de audiovisuele perceptie van biologische beweging. Hoofdstuk 7 is een inleidend hoofdstuk waarin we een selectief overzicht geven van audiovisueel perceptie-onderzoek in het algemeen en, meer specifiek, van recent neurofysiologisch en gedragsmatig onderzoek over de audiovisuele perceptie van biologische beweging. Verder presenteren we empirische studies die onderzoeken of geluid de visuele perceptie van biologische beweging beïnvloedt. Hoofdstuk 8 beschrijft een adjustment-techniek die we ontwikkelden om een schatting te bekomen van het punt van subjectieve simultaniteit, het punt waarop auditieve en visuele stimulatie als synchroon worden ervaren. In Hoofdstuk 9 en Hoofdstuk 10 exploreren we respectievelijk of geluid de visuele detectie van een plf kan faciliteren en of de wandelsnelheid van een plf geïntegreerd wordt op een statistisch optimale manier. In Hoofdstuk 11 presenteren we een studie die aantoont dat de waargenomen diepte-oriëntatie van een plw beïnvloed kan worden door naderende en weggaande voetstapgeluiden.

Om het proefschrift te beëindigen worden voor Deel I en Deel II afzonderlijk onze bevindingen samengevat, conclusies geformuleerd, en suggesties voor toekomstig onderzoek gedaan.

Additional information

Additional information (movies, links, pdf) available at:
https://perswww.kuleuven.be/~u0050370/PhD

Thanks

Here, I want to thank the people that were involved in guiding me during my PhD: Karl Verfaillie, Jean Vroomen, Anna Brooks, Rick van der Zwan, and Niko Troje. On the next page, I want to thank some other people in Dutch.

Karl, you guided my Master Thesis and you in the first place made it possible for me to start a PhD. You have always given me the freedom to go my own way, in terms of geographical destinations but in terms of research focus as well. I have always appreciated your integrity, honesty, kindness, and open mind. Despite your busy schedule, you invited me for dinner or coffee at your home, so we could proceed our work there. The number of times that you quickly hopped into my office or that you called or emailed me to check how I was doing are countless. If there was something to discuss or to decide - not always but very often :) - we shared the same opinion or we came up with exactly the same remarks. Your explanation for this: “Great minds think alike”. How could I not want to agree on that? :) Thanks Karl!

Jean, I doubt whether it is typical for Belgians but I feel very comfortable when Dutch people are around :). Thanks Jean, for your critical eye on my work and for inviting me several times in Tilburg. Beer and chocolates always seem to work :) In the end the research line that we planned to collaborate on was put on hold. Who knows, maybe we can pick up this work later on? Keep up the good work with your lovely group in Tilburg! I hope to see you all again soon!

Rick, Anna, I had the time of my life in Australia! Thanks so much for inviting me into your lab and into your lives. Thanks for the amazing opportunities you offered me! After the visit, it wasn’t easy to discuss scientific issues only by email. However, I think we can be proud of the scientific output that resulted from our collaboration. I would love to see you both again one day, in Leuven (always welcome of course :)), or in Lausanne, Coffs Harbour, or anywhere else. Meanwhile I’ll keep my memories alive on the cozy times that we shared. “Oh man, chockablock”-“No worries, mate!” :)

Niko, you also invited me into your lab and into your live. There is one particular word that describes your Biomotion Lab pretty well: ‘Cool!’. The trips with your family and Mila were lovely! No wonder you prefer to settle in Canada :). It was an amazing

experience for me to discover that we share so many interests. I truly appreciated the time you spent with me inside and outside your lab. We would love to have you over in Leuven!

Karl, Jean, Rick, Anna, Niko, thanks for the wonderful time, your great support, and for believing in me!

I also thank the people that kept in touch after a conference or helpfully responded to my questions by email, Cora MacHatch, Russell Reid, Rob van Eijk, Stuart Jackson, Waka Fujisaki, David Burr, Thomas Koelewijn, and Roberto Arrighi.

Finally, I would like to thank the members of the jury for the time and energy they invested in reading this dissertation, Prof. Ian Thornton, thanks for coming to Leuven, Prof. Johan Wagemans, Prof. Ruffin Vogels, as well as the members of the supervisory committee, Prof. Stephan Swinnen, Prof. Ruffin Vogels, Prof. Ralf Krampe, Prof. Johan Wagemans, and Prof. Jean Vroomen who provided helpful feedback and comments on my research project.

Ben

October 2010

Bedankt

Naast het clubje professoren mocht ik voor raad en suggesties ook beroep doen op een aantal andere krakken in het vak. Peter De Graef, Filip Germeys, Jan Vanrie, Peter Zaenen, en Robbe Goris. Bedankt mannen voor jullie ideeën, glashelder bij een tas koffie in het labo of al wat troebeler, maar daarom zeker niet minder spitsvondig, aan de toog.

Gelukkig waren er mijn bureaugenootjes. Eerst Sara en Kristien in 00.68 (ha je nie haan sluiten? – brrrocoli met krrruidenkaas) en later ook Karolien en Goedele in een vernuftig rotatiesysteem in 01.87. Aline en Marc, jullie tellen eigenlijk ook als bureaugenoten. Aan allen, ik ben blij dat jullie in mijn team zaten :)

Alex, Valeria, Iwona, Elke, Ilse, en Sarah, thanks for your precious help and enthusiasm! Joa, expert in het inkorten en herformuleren, Tom W., Wilfried, bedankt voor de statistische support!

Ik was eigenlijk van plan iedereen van het labo van experimentele psychologie bij naam te vernoemen maar aangezien de labopopulatie de laatste jaren sneller gegroeid is dan de muizenpopulatie in de kelder (en die kunnen er wat van!), is dat er misschien wat over. We moeten toegeven, het is wel gezellig zo'n grote groep. Wie geeft het volgende feestje? :). Wat mijn dag in het labo dikwijls goed maakte waren de 'yoows' in de wandelgangen, de babbels aan de koffiemachine, de soepjes of broodjes onder de middag, de pauzes aan de buitendeur (in de kou) bij Marleen en co, de squashpartijtjes (ja Sara, ook al werd ik telkens afgedroogd), de muffe geur van de dove kamer, de uuuuuren trial-plezier (zeker als ze je achteraf vertellen dat er iets is misgelopen bij het opslaan van de data, è Tom P. en Robbe :)), de feestjes in de akoestisch uitgebalanceerde en sfeervol verlichte psi-cafetaria en de etentjes en kleine doorzakkingen achteraf…

Niet alleen de mensen van ons labo maar ook andere psi-gang(st)ers fleurden dikwijls mijn werkdag op of maakten het leven simpelweg gemakkelijker door maar een telefoontje of een paar deuren, gangen of verdiepingen verwijderd te zijn, de psi-Theo's, de Cleppers, de mensen van het secretariaat en decanaat, Noël en de Johannen van DICT, Rob Stroobants, en Marcel van het onderhoud. Bedankt ook aan de honderden proefpersonen die mij van data voorzagen, hetzij tegen betaling, hetzij tegen de waardevolle beloning van 'course credit'.

En dan zijn er nog de mensen buiten de werkcontext, die mij ijverig assisteerden in de discipline ‘ontspanning’. De Neerpeltse CBA-delegatie (Commissie Bourgondische Activiteiten, hilarische in-de-gloria-aflevering mannen :)), de hele Theo-bende (laat dat volgende trouwfeest maar komen!), de Bronlaan-residenten (Johan, Wilfried en mijn broer Tom), en Aster en Bram.

Ma, Pa, Tom, Melissa, Patrica en Dirk, wat jullie al voor mij gedaan hebben is niet te vatten in een doctoraat, laat staan in een dankwoord. Super dikke merci! Bedankt ook aan mijne peter Ome Leo, tante Peterina en Nick. Maai en Hans, we wachten nu nog vol spanning af. Nonkel Ben en Tante Innie zijn klaar voor haar :). Bedankt voor jullie hulp en alvast een dikke proficiat! Nannie, Montje, Guido, Ann, Flinn en Fem bedankt om zo’n geweldige familie te zijn!

En tenslotte, mijn vriend-Inneke, mijn vrouw. Schatje, tijdens de opleiding Psychologie deelden we (met nog een paar honderd anderen) Auditorium Michotte. In het laatste jaar deelden we (met een bende Theo’s) hetzelfde stagelokaal. Vanaf dan zijn we meer en meer met elkaar beginnen delen, de doctoraatsopleiding, onze gedachten, meningen, onze slaapplek (gelukkig niet met honderden anderen of de Theo’s, sorry mannen :)), onze vrienden, ouders, onze collega’s, ons huisje in de Rondestraat, onze Gonzo, het leed dat we er mee hebben en het fortuin dat hij kost :),…, ons uur van opstaan en onze wandelweg naar het werk, een huwelijkscontract, een stukske wereld in Testelt, geweldige herinneringen aan een super deugddoend trouwfeest, en ons hoews in opbouw…Het enige wat we nog niet deelden was de doctoraatstitel, maar daar komt hopelijk met dit boekje verandering in :-)

Mijn doctoraatsjaren waren misschien minder ‘rollercoaster jaren’ dan bij jou het geval was en de portie drama die er bij mij aan te pas kwam iets minder, hoewel …



Bedankt liefje, om mee te gaan in mijn enthousiasme als de data er weer veelbelovend uitzagen en mij weer op te krikken als dat 's anderdaags niet meer het geval was, om mij toch te 'accepten' (met uitgebreide accept e-mail en al) als ik weer een reject-mail gekregen had, en om mij er terug bovenop te helpen toen ik 'gescooped' werd[1].



Thanks pumpkin, you're the one and only!

Ben

Oktober 2010

[1] Figure credits: "Piled Higher and Deeper" by Jorge Cham. www.phdcomics.com

Contents

PART I

SOME OBSERVATIONS ON THE FACING BIAS IN BIOLOGICAL MOTION PERCEPTION 1

CHAPTER 1

INTRODUCTION TO PART I 3

The perception of biological motion 5

The perception of point-light figures 5

Neural correlates of the perception of biological motion 13

The in-depth perception of point-light figures 14

What follows 22

CHAPTER 2

DETERMINING THE POINT OF SUBJECTIVE AMBIGUITY OF AMBIGUOUS BIOLOGICAL MOTION FIGURES WITH PERSPECTIVE CUES 23

Abstract 25

Introduction 25

Technique 28

Experiment 29

Conclusion 35

Appendix 36

CHAPTER 3

CORRELATED CHANGES IN PERCEPTIONS OF THE GENDER AND ORIENTATION OF AMBIGUOUS BIOLOGICAL MOTION FIGURES 37

Abstract 39

Experiments 39

Conclusion 42

CHAPTER 4

THE FACING BIAS IN BIOLOGICAL MOTION PERCEPTION: EFFECTS OF STIMULUS GENDER, AND OBSERVER SEX 45

Abstract 47

Introduction 47

Method 49
Results 51
Discussion 52
CHAPTER 5
THE FACING BIAS IN BIOLOGICAL MOTION PERCEPTION: STRUCTURE, KINEMATICS, AND BODY PARTS 55
Abstract 57
Introduction 57
Experiment 1 59
Experiment 2 66
Experiment 3 71
General Discussion 77
CHAPTER 6
THE FACING BIAS IN BIOLOGICAL MOTION PERCEPTION: SOME ADDITIONAL OBSERVATIONS ON OBSERVER SEX DIFFERENCES, PERSPECTIVE, AND RESPONSE TIMES 83
Introduction 85
Experiment 86
PART II
AUDIOVISUAL PERCEPTION OF BIOLOGICAL MOTION 95
CHAPTER 7
INTRODUCTION TO PART II 97
Audiovisual perception 99
Audiovisual perception of biological motion 112
CHAPTER 8
FINDING THE POINT OF SUBJECTIVE SIMULTANEITY IN A PHASE ADJUSTMENT TASK 117
Introduction and Goal 119
Method 119
Results and discussion 120
Conclusion 122
CHAPTER 9
DOES SYNCHRONIZED SOUND ENHANCE THE VISUAL DETECTION OF A POINT-LIGHT FIGURE? 123
Experiment 1 125

Experiment 2 127
General Discussion 134
CHAPTER 10
ARE AUDITORY AND VISUAL INFORMATION ON THE SPEED OF WALKING INTEGRATED IN A STATISTICALLY OPTIMAL WAY ? 139
Introduction 141
Experiment 141
CHAPTER 11
DO LOOMING AND RECEDING SOUNDS AFFECT THE PERCEIVED IN-DEPTH ORIENTATION OF DEPTH-AMBIGUOUS BIOLOGICAL MOTION FIGURES? 145
Abstract 147
Introduction 147
Experiment 148
CONCLUDING REMARKS AND FUTURE RESEARCH 155
Part I 155
Part II 159
REFERENCES 167
MOVIES 189

Part I

Some observations on the facing bias in biological motion perception

Chapter 1

Introduction to Part I

THE PERCEPTION OF BIOLOGICAL MOTION

During our daily-life activities, we experience it as quite evident that we recognize other people and what precisely it is that they are doing. Indeed, seldom we are aware of the complexity of this process, because it happens so fast and without any noticeable effort. When we start realizing, however, it is quite remarkable that the process in which the brain extracts socially and behaviorally relevant information, and forms a vivid and meaningful interpretation of another person and his or her actions, doesn't take more than some tenths of a second. Even more remarkable, is the fact that this perceptual capacity persists when the visual information is restricted to a number of point lights representing some of the joints of that person.

In sharp contrast to the brain's processing speed of motions generated by animate beings, it has been a matter of more than three decades, so far, that researchers are trying to understand and describe this process, also called '*perception of biological motion*'. The observation that such a rich percept is preserved under point-light conditions (Johansson, 1973) has lead researchers to employ these so-called 'point-light figures' as a tool to investigate human action perception.

THE PERCEPTION OF POINT-LIGHT FIGURES

Johansson (1973)[2] recorded an actor wearing tight fitting dark clothes with light bulbs or retro reflective material, attached to the main joints (see Movie 1 for a contemporary example), in a sparsely lit studio. After adjusting the contrast, so that only some 'point lights' were visible on a uniform black background, it became evident that, despite the impoverished visual information, humans easily perceive what action the actor is performing. Johansson's (1973) paper became a highly cited classic. Recently, the paper reached over one thousand citations (at the time of this writing 1005 citations; ISI Web of Knowledge).

Obviously, it is not evident, and beyond the scope of the present dissertation, to give an extensive overview of the research on biological motion perception. Several good reviews are available (see Blake & Shiffrar, 2007; Knoblich, Thornton, Grosjean, & Shiffrar, 2006; Troje & Chang, in press; Vangeneugden, 2010; Vanrie, 2005; Verfaillie, 2000).

[2] For more historical background on precursors of this technique, see Verfaillie (2000).

Here, we will restrict ourselves to mentioning the most relevant stimulus manipulations and findings on the perception of biological motion and its neural correlates. Thereafter, we focus on research investigating the in-depth perception of point-light figures.

THE PERCEPTUAL RICHNESS OF THE STIMULUS

Ever since Johansson's (1973) demonstration, point-light figures have been widely used and several techniques for creating them have been developed and refined. For an overview of techniques, see Dekeyser, Verfaillie, and Vanrie (2002). During the past decades, the research on biological motion has confirmed the richness of information that is conveyed by these figures. Indeed, despite the strong degradation of visual information, observers are able to effortlessly extract a variety of behaviorally relevant information such as the type of action (Dittrich, 1993; Vanrie & Verfaillie, 2004), the identity of the actor (Cutting & Kozlowski, 1977; Loula, Prasad, Harber, & Shiffrar, 2005; Troje, Westhoff, & Lavrov, 2005), their gender (Barclay, Cutting, & Kozlowski, 1978; Cutting, 1978a; Hiris, & Ewing, 2010; Jordan, Fallah, & Stoner, 2006; Kozlowski & Cutting, 1977; Mather & Murdoch, 1994; Pollick, Kay, Heim, & Stringer, 2005; Troje, 2002; Troje, Sadr, Geyer, & Nakayama, 2006; van der Zwan et al., 2009), emotional state (Dittrich, Troscianko, Lea, & Morgan, 1996; Montpare, Goldstein, & Clausen, 1987; Pollick, Paterson Bruderlin, & Sanford, 2001), age (Montepare & Zebrowitz-McArthur, 1988), and intentions (Manera, Schouten, Becchio, Bara, & Verfaillie, 2010). Observers are, for example, also able to infer the weight of a lifted object under point-light conditions (Runeson & Frykholm, 1981).

EXPERIMENTAL CONTROL

The relative ease[3] with which the figure can be experimentally controlled, has also contributed to its widespread use. Both spatial and temporal characteristics of the figure

[3] The ease with which the point-light figure can be experimentally controlled strongly depends on the type of information that the researcher wants to manipulate. The richness of information that is contained in point-light figures makes it very hard to isolate and manipulate only a single variable at a time. Indeed, many variables are somehow related. Take, for instance, the structural and motion information (also see below). Moreover, often crucial information is contained in the characteristics of (relative) motions of the dots. For example, the asymmetries in the motions of the dots of a point-light walker could convey specific information about the identity of the walker (e.g., the actor is slightly limping). Hence, when trying to control for spatial or temporal asymmetries in the stimulus (see, e.g., the point-light walker in Vanrie and Verfaillie, 2004), crucial biological motion information might be discarded. There seems thus to be a tradeoff between the controllability of the stimulus (how easily it could be mathematically described) and the informational

can be manipulated. Without giving an exhaustive list of all references available, spatial manipulations include for instance, low-level aspects of the dots such as dot luminance (Grossman & Blake, 1999), contrast polarity or variations in the dot's spatial frequency (Ahlström, Blake, & Ahlström, 1997), the number of dots of which the point-light figure consists (Mather, Radford, & West, 1992), the locations where on the body the dots are positioned (Dittrich, 1993), the eccentricity of the figure (Ikeda, Blake, & Watanabe, 2005; Thompson, Hansen, Hess, & Troje, 2007) the upper figure part rotated 180° relative to the lower figure part (Verfaillie, 1993, 2000), or in-plane (Pavlova & Sokolov, 2000; Sumi, 1984) and in-depth (Jackson, Cummins, & Brady, 2008; Vanrie, Dekeyser, & Verfaillie, 2004; Verfaillie, 1993, 2000; Verfaillie & De Graef, 2000; Verfaillie, de Troy, & Van Rensbergen, 1994) rotations of the point-light figure.

Examples of manipulations in the temporal domain are, varying the overall duration or the duration of the delay between frames (Thornton, Pinto, & Shiffrar, 1998), reversing the presentation order of the frames (Pavlova, Krägeloh-Mann, Birbaumer, Sokolov, 2002; Vanrie et al., 2004; Verfaillie, 1993, 2000), restricting the lifetime of the dots (Beintema & Lappe, 2002; Neri, Luu, & Levi, 2006; Neri, Morrone, & Burr, 1998; also see Chapter 9), exaggerating temporal differences (Hill & Pollick, 2000), or temporally scrambling the dots (Arrighi, Marini, & Burr, 2009; Bertenthal & Pinto, 1994; Hiris, Humphrey, & Stout, 2005; Neri et al., 2006; also see Experiment 2 of Chapter 9).

An often applied manipulation of task performance is adding dots to the display that resemble the dots of the point-light figure, also called 'masking' (Bertenthal & Pinto, 1994; Cutting, Moore, & Morrison, 1988). In random dot masks, the masking dots appear on a new position on each frame. This type of masking has proven to be not really effective for traditional point-light figures because the figures easily 'pop out' (Thornton et al., 1998). Therefore, often scrambled motion masks are used (Cutting et al., 1988; Thornton, Rensink, & Shiffrar, 2002), which consist of dots featuring the same motion trajectories as the motion trajectories of the dots of the point-light figure (Arrighi et al., 2009; Hiris et al., 2005; Neri et al., 2006). Note, however, that random dot masks are more effective when also the dots of the point-light figure are repositioned on each frame so that they do not contain local motion (Beintema & Lappe, 2002; also see Experiment 1 of Chapter 9).

---

richness of the stimulus (how well the characteristics, that may convey important information, are maintained).

ESTABLISHED EFFECTS AND POINTS OF DISCUSSION

Inversion effects

An inversion effect for a certain type of stimulus is usually defined as a decrease in recognition performance when the stimulus is inverted, i.e., presented upside down. Inversion effects are typically found to be larger for faces than for objects. A number of studies demonstrate that inversion effects in faces can be attributed to the processing of global, configural information, i.e., to the processing of the spatial layout of facial features rather than to the identification and classification of the particular features themselves (Diamond & Carey, 1986; Leder, Candrian, Huber, & Bruce, 2001; Maurer, Le Grand, & Mondloch, 2002; Van Belle, De Smet, De Graef, Van Gool, & Verfaillie, 2009; Van Belle, De Graef, Verfaillie, Rossion, & Lefevre, 2010).

Strong inversion effects have also been reported for the perception of point-light figures (Sumi, 1984). Accuracy for detecting a coherent point-light target in a mask of scrambled walkers drops to chance level when the target is inverted (Bertenthal & Pinto, 1994). Observers' abilities to judge the type of the action (Dittrich, 1993) and the gender of the actor (Barclay et al., 1978) are also disrupted upon inverting point-light displays. In analogy to the face inversion effect, the inversion effect for point-light figures has also been attributed to impaired 'global processing' (e.g., Bertenthal & Pinto, 1994). However, Troje and colleagues (Chang & Troje, 2008, 2009a, 2009b; Saunders, Suchan & Troje, 2009; Troje & Westhoff, 2006) recently showed that the local inversion effect in point-light figures is carried in large part by information in the feet. Their studies suggests the existence of a so-called 'life detector', a visual filter that can quickly and automatically detect the presence of a moving living organism and direct attention to it. Evidence for such a filter was found in tasks in which observers had to detect the left-right orientation of a spatially scrambled point-light figure, containing no translational motion. Because the visual filter is specifically sensitive to gravitational acceleration, the left-right orientation of a point-light figure can easily be detected, even under spatially scrambled conditions. However, when the scrambled point-light display or only the trajectories of the foot dots are inverted, this ability disappears.

It should be noted though, that inversion effects have also been documented for static body postures (Reed, Stone, Bozova, & Tanaka, 2003; Reed, Stone, Grubb, & McGoldrick, 2006). Rather than arguing against a role of configural processing, Troje and

Chang (in press) suggest the existence of two entirely different sources for the inversion effect in biological motion perception. The first source may very well be related to the inversion effect observed with faces: Diagnostic features inherent to the subtle changes in the configuration of the stimulus can be derived better when the stimulus is seen in a familiar orientation. The second source, on the contrary, has nothing to do with configural organization but is related to local motion information, i.e., the accelerations in the feet.

Bottom up vs. top down and the role of attention

Whereas some studies have suggested that a global biological motion percept can result from attention-free bottom-up processes (Giese & Poggio, 2003; Johansson, 1973; Mather et al., 1992; Thornton & Vuong, 2004), other studies have shown convincing effects of top-down processes (Bertenthal & Pinto, 1994; Bulthoff, Bulthoff, & Sinha, 1998; Thornton et al., 1998; Thornton, Vuong, & Bulthoff, 2003) or strongly suggest that attention is required for the processing of point-light figures (Battelli, Cavanagh & Thornton, 2003; Cavanagh, Labianca, & Thornton, 2001; Safford, Hussey, Parasuraman, & Thompson, 2010; Thornton et al., 2002). Recently, Chandrasekaran, Turner, Bulthoff, and Thornton (2010) found that performance on biological motion tasks highly correlate with the ability to control selective attention. Observers who were better at selectively attending, were faster and more accurate at processing point-light displays. However, Chandrasekaran et al. (2010) do not dispute the fact that the processing of biological motion can proceed in a fast, efficient, bottom-up manner. Indeed, certain display manipulations or task demands (e.g., left-right orientation discrimination) may not require the addition of top-down, active processing strategies, while others (e.g., identification of the action or the actor) clearly do.

Note that the above mentioned hypothesized existence of a ‘life detector’ would predict that perceiving a point-light figure does not require explicit attention, but rather attracts it (Troje & Chang, in press). More specifically, local invariants contained in the kinematics of the feet might attract attention. Obviously, a prerequisite for this to happen is that these natural local invariants are present in the stimuli. As suggested by Troje and Chang (in press) this was not the case in some studies, especially in studies in which the synthetic Cutting algorithm (Cutting, 1978b) was employed to generate point-light walkers (Thornton et al., 2002; Thornton & Vuong, 2004). Cutting's algorithm meant a breakthrough for research on biological motion perception. It allowed to easily generate point-light displays by a computer program and provided considerable control over the

spatial and temporal characteristics of the display. However, walkers generated by the algorithm do not contain the local invariants that are present in motion-captured point-light figures (Saunders et al., 2009).

Once attention is on the point-light figure, possibly attracted by a pre-attentive local mechanism like the life detector, further global processing mechanisms can retrieve additional information (Troje & Chang, in press). Generally, it is often concluded that both bottom-up and top-down processes are involved in the perceptual analysis of point-light animations (Blake & Shiffrar, 2007; Vanrie, 2005; Vanrie & Verfaillie, 2006, 2008).

Form and/or motion

One of the current hot topics in the literature on the perception of biological motion is whether form or rather motion is most crucial in perceiving point-light figures. The discussion is sometimes heated and quite complex. Here, we restrict ourselves to mentioning some of the most relevant findings.

The initial observation of Johansson (1973) suggested that the motions of the dots are crucial to reach a rich percept of the performed action. Mather et al. (1992), found that omitting the wrist and ankles dots of Johansson's point-light walker strongly degraded performance, compared to omitting more proximal shoulder and hip dots. Mather et al. (1992) suggested a special role for the local motion of the distal parts of the limbs. However, they never tested the role of wrist and ankle dots separately. Later it became clear, as already mentioned, that crucial biological motion information is contained in the local motions of the feet (Chang & Troje, 2008, 2009a, 2009b; Saunders et al., 2009; Troje & Westhoff, 2006). In addition, opponent motion, that is, motion of two (or more) dots in an opposite direction, has also been suggested to be a crucial aspect of biological motion (Casile & Giese, 2005; Chang & Troje, 2009a; Thurman & Grossman, 2008).

Others have emphasized the contribution of form to the perception of point-light biological motion. Beintema and Lappe (2002) created a limited-lifetime stimulus in which eight light points were positioned, not on the joints, but rather on the limbs at a random position between joints (Movie 2). In every frame of the animation, each point was reallocated to another randomly selected position on the limb, with every position on the line segment between the two joints having equal probability. Beintema and Lappe (2002) found that observers spontaneously recognized a moving human figure in these displays without local image motion and that their performance was very similar to that for a classic

point-light display. The authors concluded that image motion may generally aid processes such as segmenting figure from background, but that it is not the basis for the percept of biological motion. Rather, they suggest that biological motion is derived from dynamic form information on body posture evolving over time (also see Lange, Georg, & Lappe, 2006).

As is often the case for other discussions (as for example the discussion about whether attention is necessary or not to perceive biological motion), whether form or motion information is relevant to perceive point-light biological motion strongly depends on the task and the specific point-light displays that are used (Thirkettle, Benton, Scott-Samuel, 2009). Indeed, controversies can only be resolved if it is acknowledged that the perception of biological motion is not a single phenomenon but consists of a number of different aspects that have to be distinguished carefully, both conceptually and experimentally. For example, detecting a masked point-light figure or deriving its facing direction might address very different processing stages (Troje, 2008a).

Computational models on biological motion perception (e.g., Giese & Poggio, 2003; Lange & Lappe, 2006) can account for a large range of results. However, they still have considerable limitations. For example, top-down processes are usually neglected and the models primarily deal with 2D sagittal representations. The models thus contain no explicit means for disambiguating the 3D orientation of a perceived human figure, while experimental results indicate that the neural signals supporting biological-motion perception integrate information on the form, motion, and 3D depth orientation. (Jackson & Blake, 2010).

Effects of action execution on action perception

Converging lines of evidence indicate that one's own actions can affect one's perception of the actions of others. Sensory representations used during action perception are believed to overlap with the motor representations used during action planning, also referred to as 'common coding' (Prinz, 1997; also see the Theory of Event Coding, Hommel, Musseler, Aschersleben, & Prinz, 2001, and Wilson & Knoblich, 2005).

For example, an observer is better able to notice a change in the limb position of an actor when the observer too is moving the corresponding limb (Reed & Farah 1995). Jacobs and Shiffrar (2005) let observers judge the speed of a point-light walker while the observers were walking themselves (on a treadmill), cycling (on a stationary bike), or

standing. Judging the speed of the walkers was worst when observers were walking themselves, suggesting that perception and performance of the same action interfere.

If motor experience influences visual action analysis, then observers should be most sensitive to their own movements. If view-dependent visual experience determines visual sensitivity to human movement, then observers should be most sensitive to the movements of their friends. To test these predictions, Loula et al. (2005) presented to observers sagittal displays of point-light depictions of themselves, their friends, and strangers performing various actions. In actor identification and discrimination tasks, sensitivity to one's own motion was highest while visual sensitivity to the actions of friends, but not to the actions of strangers, was above chance, suggesting that both motor and visual experience define visual sensitivity to human action.

If motor programs directly influence visual action perception then the acquisition of novel motor behaviors should improve their visual recognition, even in the absence of visual learning. Casile and Giese (2006) tested this prediction by measuring the visual recognition of gait patterns from point-light stimuli before and after non-visual motor training. Participants were blindfolded and learned a novel coordinated upper-body movement based only on verbal and haptic feedback. The learned movement matched one of the visual test patterns. Despite the absence of visual stimulation during training, Casile and Giese (2006) observed a selective improvement of the visual recognition performance for the learned movement. Visual recognition performance after training correlated strongly with the accuracy of the execution of the learned motor pattern showing that motor learning has a direct and highly selective influence on visual action recognition. Recently, Bidet-Ildei, Chauvin, and Coello (2010) showed that the execution of an action can prime the subsequent perception of a similar action even when the motor prime is very short. Indeed, producing a running activity as short as 30 seconds facilitated the perception of the direction of a point-light runner embedded in a dense dynamical mask.

Note that the causal arrow can also point in the opposite direction: Action perception can influence action execution (Blake & Shiffrar, 2007). For instance, Kilner, Paulignan, and Blakemore (2003) let participants make arm movements while observing either a robot or another human making the same or qualitatively different arm movements. Variance in the executed movement was measured as an index of interference to the movement. Results demonstrated that observing another human making incongruent movements interfered with the execution of the movement. However, interference did not occur when the observed action was executed by a robotic arm making incongruent

movements. The simultaneous activation of overlapping neural networks that process movement observation and execution thus infers not only a measurable cost to the perception of actions (e.g., Loula et al., 2005) but also to motor control (Kilner et al., 2003).

NEURAL CORRELATES OF THE PERCEPTION OF BIOLOGICAL MOTION

Although the specific neural correlates that subserve the visual perception of actions or biological motion remain largely unclear, a general consensus seems to exist on the involvement of two neural structures (Vanrie, 2005): the Superior Temporal Sulcus (STS) and mirror neurons.

The first evidence for the involvement of the STS in the perception of actions came from Perrett et al. (1985). Using single cell recording in macaque cortex, these researchers discovered neurons in the STS selectively responsive to human forms and motions (Oram & Perrett 1994). Ever since, the STS has been reported in almost all neurophysiological studies on the perception of biological motion (for some examples, see Allison, Puce, & McCarthy, 2000; Barraclough, Xiao, Baker, Oram, & Perrett, 2005; Beauchamp, Lee, Argall, & Martin, 2004; Beauchamp, Lee, Haxby, & Martin, 2002; Grossman, Battelli, & Pascual-Leone, 2005; Grossman & Blake, 2002; Jastorff & Orban, 2009; Puce & Perrett, 2003; Safford et al., 2010; Saygin, 2007; Thompson, Clarke, Stewart, & Puce, 2005). Peuskens, Vanrie, Verfaillie, and Orban (2005) found that it was primarily the STSp (posterior part of STS; also see Bonda, Petrides, Ostry, & Evans, 1996) in the right hemisphere that responds strongly to human motion.

Mirror neurons were discovered in the premotor cortex (area F5) in monkeys (e.g., Buccino, Binkofski, & Riggio, 2004; Hari, Avikainen, Kirveskari, Salenius, & Rizzolatti, 1998; Molenberghs, Brander, Mattingley, & Cunnington, 2010; Nelissen, Luppino, Vanduffel, Rizzolatti, & Orban, 2005; Rizzolatti & Craighero, 2004; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). These neurons are not only active during action observation, but also when the monkey is performing the action. In humans, mirror neurons have been localized in the inferior frontal gyrus (Broca's area; e.g., Iacoboni, Woods, Brass, Bekkering, Mazziotta, & Rizzolatti, 1999). The human mirror neuron system in premotor cortex has also been shown to respond to point-light biological motion (Saygin, 2007; Saygin, Wilson, Hagler, Bates, & Sereno, 2004).

For other regions responsive to the sight of biological motion such as, for example, the extrastriate body area (EBA), a region that is activated when one views human bodies or isolated body parts (Downing, Jiang, Shuman, & Kanwisher, 2001; also see Peelen, Wiggett, & Downing, 2006), we refer to works that provide a more detailed discussion on the neural structures and brain networks involved in the perception of actions (e.g., Blake & Shiffrar, 2007; Knoblich et al., 2006; Vangeneugden, 2010).

Finally, we mention that, as suggested by Troje and Chang (in press), the Superior Colliculus, known to drive orientation responses in many animals and responsible for directing eye movements in humans (also see Chapter 7), might be a likely candidate for the implementation of the life detector mechanism. This subcortical structure could then activate higher cortical areas, preparing them to further extract relevant information (Vangeneugden, 2010).

THE IN-DEPTH PERCEPTION OF POINT-LIGHT FIGURES

The research presented in Part I of this dissertation follows up on research on the perception of biological motion that has recently been done in our lab, in particular, the research on the in-depth perception of point-light figures.

Dekeyser et al. (2002) described a method for creating biological-motion stimuli on the basis of motion capture data and animation software. With this method they created a database of 22 point-light actions each in five orientations: Frontal/back view, left and right ¾ view, and left and right lateral view (Vanrie & Verfaillie, 2004). Using the *point-light walking* action in five orientations, Vanrie et al. (2004) then started to investigate the in-depth perception of point-light figures.

One of the interesting properties of 2D orthographical projections of point-light figures is that while, in fact, they are consistent with an infinite number of 3D configurations, they are consistent with only two 3D configurations that respect the anatomical constraints of the human body. Both interpretations are equally plausible in principle. For instance, the orthographic projection of a 3D walker who is facing the viewer in frontal view (0°) is identical to the projection of a walker who is facing away from the viewer (180°). The same holds for, for example, a ¾ view. The 2D projection of a 3D walker who is facing away 135° to the right, or facing the viewer 45° to the right, results in the very same stimulus, namely a right ¾ view (see Figure 1A). Note that for a lateral view (90° or 270°), the 2D projection does not make the in-depth orientation

ambiguous in the sense that the walker could be either oriented towards the viewer or away from the viewer. However, it does make the in-depth organisation ambiguous because from the 2D projection it is not clear which arm or leg is closest to the viewer. For instance, a right lateral view (90°) can be interpreted as front leg closest or front leg furthest away (see Figure 1B).

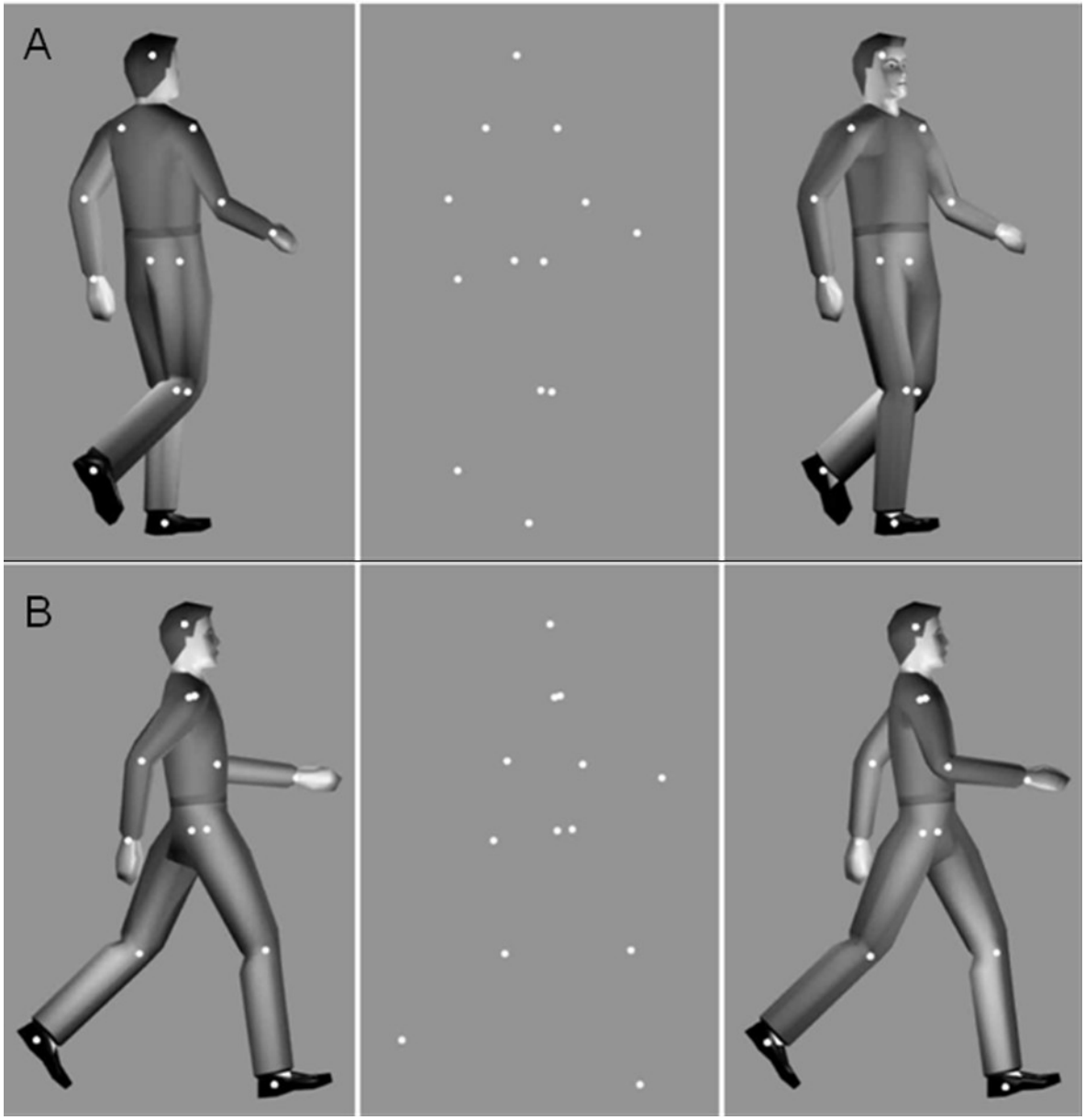


Figure 1. Single frame of the point-light walker (middle), and the two veridical interpretations with the point lights superimposed: A. A right ¾ view can be interpreted as either a 45° or 135° orientation. B. A right lateral view (90°) can be interpreted as front leg closest or front leg furthest away. Reproduced with kind permission from Vanrie, Dekeyser, and Verfaillie (2004).

VANRIE, DEKEYSER, AND VERFAILLIE (2004)

In their first experiment Vanrie et al. (2004) on each trial presented to observers a completely depth-ambiguous point-light walker, walking as if on a treadmill, in one of five in-depth orientations and asked observers to indicate the perceived in-depth orientation. On a A4 cartboard sheet with a movable arrow, observers indicated one of the eight possible orientations. Vanrie et al. (2004) aimed, first, to explore whether observers actually perceive this stimulus in different ways (i.e., as a human walker that is either oriented towards them or oriented away from them), and second, to check whether, in the absence of relevant cues to depth, one interpretation may be preferred and thus reported more frequently. To control for biasing factors that were not specifically related to the processing of biological motion, they included a control condition in which the point-light walker was inverted (i.e., presented upside down). Indeed, an inverted walker retains much of the low-level features and is equivalent in terms of depth cues (i.e., no explicit depth cues).

In this experiment, Vanrie et al. (2004) did the following observations. First, they found that for both the upright and inverted walkers, the proportion of "facing the viewer" responses never reached 1 or 0, indicating that observers reported both of the veridical interpretations and implying that the point-light figure can indeed be considered a perceptually multistable figure. Second, despite the absence of differentiating visual cues, for all non-lateral views, observers showed a strong preference to interpret the upright walker as facing towards them (in about 80% of the cases). Third, this preference was not present for the inverted walkers. Vanrie et al. (2004) interpreted these findings as evidence for a perceptual bias in the processing of bistable biological motion stimuli.

One could argue, however, that observers are biased to interpret the action of walking as approaching (rather than receding) regardless of the orientation of the figure in depth. Indeed, the global motion implied by the articulatory pattern was systematically confounded with the global orientation in depth. That is, a point-light walker was seen as "approaching" when it was seen as "facing the viewer" and vice versa. To distinguish these two alternatives, Vanrie et al. (2004) repeated the first experiment but now reversed the walking motion of the point-light figure so that it was walking backwards (still as if on a treadmill). Hence, if it was the implied global motion to which observers were responding then, now, an "approaching" walker should be perceived as facing away from the viewer. The results of the second experiment showed, again, that point-light figures were perceived

in both orientations. More importantly, observers were still biased to interpret the upright point-light figure as facing towards them, even when this implied that the walker was consistent with moving away from them.

In a third experiment, Vanrie et al. (2004) examined whether the interpretation of the point-light walker would be affected by the introduction of a depth cue signaling a single veridical orientation. More specifically, they investigated whether occlusion or perspective cues could disambiguate the orientation of the point-light walker. The occlusion cue was a locally defined cue that consisted of the disappearance of a point light when another limb was positioned in between the point light and the observer. The perspective cue was a more global cue because the locations of the dots and the relative distances between them changed depending on their location in depth. In this experiment Vanrie et al. (2004) presented point-light walkers in two conditions (occlusion or perspective) and in eight different orientations (frontal, back, 45° left and right, 90° left and right, and 135° left and right). The results revealed that the addition of either occlusion or perspective information effectively disambiguated the point-light walker. However, of both cues the perspective cue was most efficient in disambiguating the in-depth orientation. In some orientations the bias to perceive the point-light walker as facing towards the viewer was strong enough to counteract the occlusion cue.

In summary, the study of Vanrie et al. (2004) showed that a point-light walker containing no cues to depth is perceived as either oriented towards the viewer or away from the viewer. The most remarkable finding by Vanrie et al. (2004) was the tendency of observers to perceive the point-light walker as oriented towards them. According to Vanrie et al. (2004) there are several factors that could be responsible for this perceptual bias. First, high-level cognitive factors could play an important role (Klopfer, 1999) or even stored motion patterns or ‘sprites’ (Cavanagh et al., 2001). This would imply that familiarity is a key factor. However, as argued by Vanrie et al. (2004), familiarity does not seem to be a valid explanation for their results. Indeed, there is no principled reason for assuming that observers are confronted visually more often with frontal views than with back views of their conspecifics. Rather, biological or social relevance might be at the heart of the phenomenon (Vanrie et al., 2004).

VANRIE AND VERFAILLIE (2006)

Vanrie and Verfaillie (2006) studied the generality of the bistability and the perceptual bias and coined the term '*facing bias*' to refer to the bias to perceive the point-light figure as facing towards the viewer. More specifically, they investigated whether the process of depth assignment depends on the spatiotemporal characteristics of the point-light figure, or in other words, on the action that is performed by the point-light figure. They formulated two opposite hypotheses, a "general tendency hypothesis" and an "action dependent hypothesis", and verified in three experiments which of the two hypotheses could be supported empirically.

It could be the case, for example, that resolving the depth ambiguity of point-light figures might just be a stochastic process in which the a priori probability of a facing interpretation is simply much higher (see e.g., Hogervorst & Eagle, 1998; Mamassian & Landy, 1998). That is, for some reason, for example, the higher social importance of a facing orientation, there might exist a general preference to perceive human bodies as being oriented toward the viewer. As a result of this, any ambiguity would generally be resolved by interpreting all stimulus features to be consistent with a facing interpretation, regardless of the specific movement characteristics of the body. In such case, the action that the point-light figure performs does not contribute to the resolution of the ambiguity concerning the depth orientation and will thus not significantly affect the interpretation. Contrary to this "general tendency" hypothesis, both the bistability and the perceptual bias might be consequences of a more stimulus-driven process, making these phenomena specific for the kinematics of a walking movement (action dependent hypothesis).

In the first experiment Vanrie and Verfaillie (2006) asked observers to indicate the in-depth orientation of four different point-light actions from the action database (Vanrie & Verfaillie, 2004): An aerobic-style jumping movement, a cycling action, a rowing motion, and a one-handed saluting movement. Each action was presented in one of five orientations (a frontal/back view, two lateral views, and two ¾ views). As in Vanrie et al. (2004), on each trial, observers indicated one of the eight possible orientations. Results showed perceptual ambiguity for all four actions, confirming the status of the point-light figure as being an ambiguous or bistable stimulus. A facing bias, however, was not found for all movement patterns and even for the patterns for which it was found, the strength of the bias varied. Vanrie and Verfaillie (2006) therefore concluded that the facing bias is not an

essential aspect of the perception of ambiguous point-light figures. A general preference is thus not the sole determining factor of the facing bias.

In the second and third experiment Vanrie and Verfaillie (2006) further explored the factors that mediate the process of depth assignment. In the second experiment they tried to exclude semantically related effects as much as possible by including unfamiliar actions. That is, by using movements that can be assumed to be only rarely encountered, and that do not have an explicit goal or obvious verbal label, the authors attempted to minimize possible effects of stored knowledge of actions (cf. "sprites"; Cavanagh et al., 2001). The movement pattern of the actions in the first experiment for which a high facing bias was previously found revealed at least two common features, namely, the plane in which the upper limbs were located, and the plane in which the lower limbs were located, vis-à-vis the torso. More specifically, for walking, rowing, and cycling, both the arms and legs moved in front of the torso, that is, in a plane parallel to the midsagittal plane (i.e., movement perpendicular to the torso). For saluting and jumping, in contrast, the limbs remain mostly in the same plane as the torso, that is, in the midfrontal plane (lateral movement). In the second experiment Vanrie and Verfaillie therefore presented unfamiliar actions in which the plane of movement of upper and lower limbs was systematically manipulated to be either lateral, vis-à-vis the torso, or perpendicular to it. The results can be summarized in three general findings. First, the plane of movement of the lower limbs appeared not to be the most relevant factor, because in itself it did not have a significant effect on the interpretation. Second, the kinematic pattern clearly had a strong effect on the depth assignment process. A preferred interpretation became apparent only when the upper limbs moved perpendicular, vis-à-vis the torso. Third, the previous points did not hold in every instance, but seemed to depend in part on the viewpoint from which the stimulus was presented. When the figure was presented from a frontal viewpoint, the preference to interpret the stimulus as facing the viewer, as observed for a ¾ view, was either diminished or completely absent.

The third experiment, in which actions were presented in a random order instead of blocked, and in which the stimulus duration was reduced to a single step or arm movement instead of two steps or arm movements, basically confirmed the results of the first and the second experiment (for more details on this experiment, see General Discussion of Chapter 5). Vanrie and Verfaillie (2006) concluded that point-light figures without ordinal depth cues are generally bistable, regardless of their kinematic pattern. The perception of the

global orientation, and thus the facing bias, is strongly affected by aspects of the movement pattern.

VANRIE AND VERFAILLIE (2010)

In another study, also consisting of three experiments, Vanrie and Verfaillie (2010) exploited the phenomenon of reversibility to search for stimulus features relevant in the process of depth assignment. In this study the authors showed that perceptual switches indeed occur during prolonged viewing of a point-light action. However, switches occur less frequently as compared to, for example, the number of switches that are observed for a Necker cube (Experiment 1; also see Jackson, Cummins, & Brady, 2008).

The reversibility of point-light actions was confirmed in their second experiment, in which the perceptual ambiguity of the point-light action was manipulated (by manipulating the type of action; see Vanrie & Verfaillie, 2006) as well as the intention of observers to perceive a particular alternative. Moreover, the pattern of eye movements revealed stimulus features specifically associated with the perception of the different alternatives. Whereas eye movement recordings of observers reporting switches in the perceived orientation of point-light *walkers* delivered somewhat inconsistent data, eye movement data of observers reporting switches in the in-depth interpretation of a *jumping* point-light figure suggested that fixations in the lower part of the figure resulted in relatively more 'facing the viewer' interpretations and fixations in the upper part of the figure in relatively more 'facing away' interpretations (also see Experiment 2 of Chapter 5).

In the third experiment, the importance of these features as determining factors of the initial interpretation was investigated by manipulating the location of the first fixation on the stimulus. For a jumping figure, but not for a walking figure, an initial fixation on the lower (upper) region of the jumping PLF heightened the probability that the figure will be perceived as being oriented towards (away from) the viewer. The fact that this not only happened for internally determined initial fixations (resulting from an observer generated saccade), but also for externally determined fixations (experimentally controlled), suggests that processing of these local features can, in fact, influence the process of depth assignment. Thus, when the visual system is confronted with a dynamic, perceptually highly instable stimulus, the initial fixation of certain discriminating features could tip the balance in one or the other direction.

VANRIE AND VERFAILLIE (2008)

Finally, we summarize the findings of Vanrie and Verfaillie (2008) exploiting the phenomenon of perceptual coupling to investigate the in-depth perception of bistable point-light figures. In this study observers were asked to indicate the perceived in-depth orientation of a pair of simultaneously presented point-light figures. The similarity between both figures was manipulated. In the first experiment, Vanrie and Verfaillie (2008) found a higher occurrence of coupled percepts for identical point-light figures. That is, observers tended to perceive two simultaneously presented point-light figures in the same orientation (both as facing the viewer or both as facing away) when they were similar. Perceptual coupling broke down when either the movement pattern or both the viewpoint and the phase-relation were changed. A second experiment confirmed these results but also demonstrated that point-light figures with different movement patterns can still be subject to perceptual coupling, as long as they share the same viewpoint and exhibit equivalent degrees of ambiguity.

RESEARCH BY OTHER GROUPS

Recently, other research groups have started to investigate the in-depth perception of point-light figures. To isolate and characterize the neural mechanisms mediating perception of biological motion, Jackson and Blake (2010) used an adaptation paradigm together with bistable point-light figures and showed that, after exposure to a point-light walker with a particular stereoscopically defined facing direction, observers experienced a consistent after-effect: A new presentation of a point-light walker was perceived predominantly reversed in depth. When adapting and test walkers differed in size or appeared on opposite sides of fixation, aftereffects did also occur, implying higher-level origins of neural adaptation. After-effects did not occur when the adaptor was phase-scrambled indicating that the after-effect did not result from local, disparity-specific motion adaptation. The authors also showed that disparity is not essential for producing adaptation: When suspended on top of a stereoscopically defined, rotating globe, a context-disambiguated "globetrotter" was sufficient to elicit an after effect, as were full-body adaptors. In summary, the results of Jackson and Blake (2010) imply that the neural signals supporting biological-motion perception integrate information on the form, motion, and three-dimensional depth orientation of the moving human figure.

Jackson et al. (2008) investigated the temporal dynamics of perceptual reversal with point-light biological motion. They demonstrated rapid and repeated perceptual switching when a rotational component was added to a standard point-light walker. In addition to the results of Vanrie and Verfaillie (2010), their results provide a compelling demonstration of the robustness of perceptual reversals with point-light biological motion.

Finally, using a frontally presented stick walker whose limbs consisted of segments reversible in depth, Oh (2009) found that in the upright orientation observers were more likely to see the arms bend forwards and the legs bend backwards with respect to the body. When these figures were inverted, however, observers were more likely to see all limbs bend in the same direction rather than in random directions. According to Oh (2009) this codirectional perception of limb structure in the inverted orientation suggests that inverted body parts are organised on the basis of their perceptual similarities, rather than stored representations of the human body form as is the case for upright walkers.

WHAT FOLLOWS

Part I of this dissertation consists of five empirical chapters in which we explore some factors that influence the perceived in-depth orientation of point-light figures. First, in Chapter 2, we present a technique to manipulate the perceived in-depth orientation and to determine the point of subjective ambiguity. Next, in Chapter 3 we report a study that resulted from our collaboration with Prof. Rick van der Zwan, Dr. Anna Brooks (Southern Cross University, Coffs Harbour, Australia), and Prof. Niko Troje (Queens University, Kingston Ontario, Canada). Together we started to explore the role of the gender of the point-light figure on its perceived in-depth orientation. Our experiments revealed a strong effect of figure gender on the perceived in-depth orientation. In Chapter 4 we further examine this figure gender effect. Using a set of 20 male and 20 female observers we demonstrate an interaction between figure gender and observer sex. In Chapter 5 we test whether the figure gender effect is truly caused by perceived gender. In this chapter, we also investigate the role of structural and kinematic stimulus properties and the role of different figure parts. Finally, in Chapter 6, we revisit our perspective technique (Chapter 2). We present a slightly modified version of the technique and test whether the observer gender effect, as observed in Chapter 4, can also be observed when the perceived in-depth orientation is manipulated with perspective cues.

Chapter 2

Determining the point of subjective ambiguity of ambiguous biological motion figures with perspective cues

This chapter is based on:

Schouten, B. & Verfaillie, K. (2010). Determining the point of subjective ambiguity of ambiguous biological motion figures with perspective cues. *Behavior Research Methods, 42,* 161-167.

ABSTRACT[4]

Orthographic frontal/back projections of biological motion figures are bistable: The point-light figure in principle can be perceived either as facing towards the viewer or as facing away from the viewer. Some point-light actions, e.g., walking, elicit a strong 'facing bias': Despite the absence of objective cues to depth, observers tend to interpret the figure as facing towards the viewer in most of the cases. In this study we present and experimentally validate a technique that affords full experimental control on the perceived in-depth orientation of point-light figures. We demonstrate that by parametrically manipulating the amount of perspective information in the stimulus, it is possible to obtain any desired level of subjective ambiguity. Directions for future research, in which this technique can be fruitfully implemented, are suggested. Program code of a demo is provided that can be modified easily for program code of new experiments.

INTRODUCTION

Gunnar Johansson (1973) was the first to introduce the point-light method to investigate the perception of biological motion. He attached light-emitting bulbs or light-reflecting patches to the bodies of moving actors and filmed them under such conditions that only the lights were perceptible for observers. Ever since, this point-light technique has been widely used and refined by vision scientists studying human action perception. During the past decades, the research on biological motion has demonstrated the richness of information contained in these so-called *point-light figures*. Despite the strong degradation of visual information, observers are able to effortlessly extract a variety of behaviorally relevant information such as the type of action (Dittrich, 1993; Vanrie & Verfaillie, 2004), the identity of the actor (Cutting & Kozlowski, 1977; Loula, Prasad, Harber, & Shiffrar, 2005), their gender (Kozlowski & Cutting, 1977; Pollick, Kay, Heim, & Stringer, 2005; Troje, 2002), emotional state (Dittrich, Troscianko, Lea, & Morgan, 1996; Montpare, Goldstein, & Clausen, 1987; Pollick, Paterson Bruderlin, & Sanford, 2001), and age (Montepare & Zebrowitz-McArthur, 1988). See Verfaillie (2000) for some notes on the early history of this line of research.

Only recently, however, the in-depth perception of point-light figures has been systematically investigated. Vanrie, Dekeyser, and Verfaillie (2004) presented observers

[4] Thanks to Sarah Moens for her help in this study.

with an orthographic 2D projection of a symmetrical point-light walker derived from the captured 3D coordinates of a walking human (for details see Dekeyser, Verfaillie, & Vanrie, 2002). The stimulus is bistable, in the sense that it has two veridical and anatomically plausible interpretations of a human figure: The figure is either perceived to be 'facing the viewer' (FTV) or it is perceived as 'facing away' (FA) from the viewer (see Figure 2). More importantly, Vanrie et al. (2004) demonstrated that, despite the theoretically ambiguous nature of this stimulus, on average people tend to interpret the walker as facing towards the observer in more than 80% of the cases.

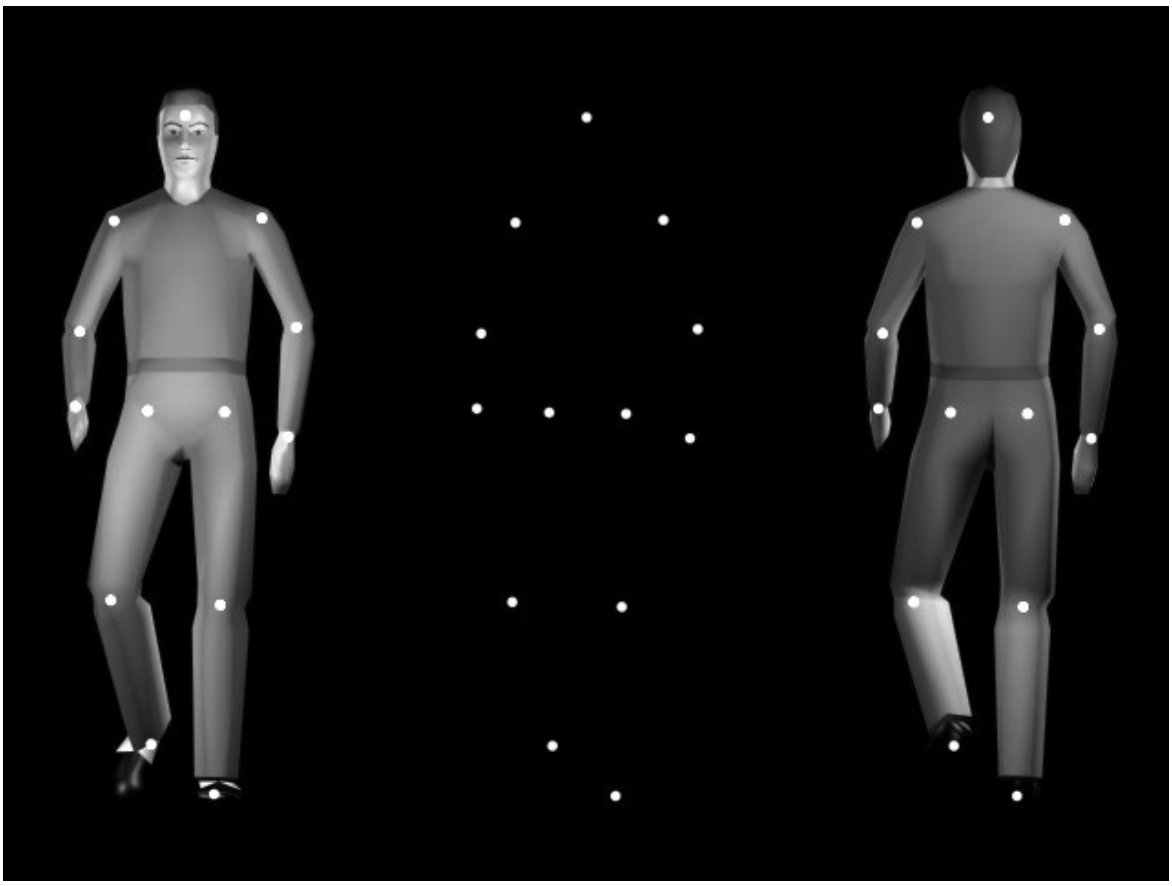


Figure 2. Illustration of a single frame of the ambiguous point-light walker (middle), and the two veridical interpretations with the point-lights superimposed. Whilst both veridical interpretations are equally likely, observers prefer the 'facing the viewer' interpretation (left). Adapted from Vanrie et al. (2004).

Subsequent research on the in-depth perception of point-light figures (Vanrie & Verfaillie, 2006) demonstrated perceptual bistability for other point-light actions (Vanrie & Verfaillie, 2004) but also revealed that the presence of a preferred interpretation strongly

depends on the performed action, suggesting that the process of depth assignment takes into account the movements of the point-light figure.

To date, it is still not clear why observers exhibit such a strong tendency to interpret some point-light actions as facing towards the viewer. A growing group of researchers is now studying contextual or stimulus specific factors such as the gender of the point-light figures (Brooks, Schouten, Troje, Verfaillie, Blanke, & van der Zwan, 2008; see Chapter 3), looming and receding sounds (Menodonça & Santos, 2008; Schouten, Moyens, Brooks, van der Zwan, & Verfaillie, 2008; also see Chapter 11), or the presence of more than one point-light figure (Vanrie & Verfaillie, 2008), that seem to affect the perceived in-depth orientation of point-light figures.

Studying factors that affect the in-depth orientation perception of point-light figures, however, is often complicated by the relatively large interindividual variation in the average perceived in-depth orientation. E.g., when repeatedly confronted with the same point-light figure, some observers might judge the in-depth orientation of the point-light figure on average as rather ambiguous, while for other observers the figure is on average judged as convincingly facing the viewer. The relatively large variation between observers can cause a decrease in power of the test when studying the effect of a contextual variable across observers, because for at least a part of the observers the effect of the contextual variable will result in a ceiling or floor effect. Hence, to more meaningfully compare the effect of a contextual variable on the perceived in-depth orientation across observers, it would be convenient if the subjective experiences could be aligned across observers to serve as a baseline or control condition. The ideal baseline to prevent ceiling and floor effects would be the point where the figure in half of the cases is perceived as oriented towards the observer and in the other half of the cases is perceived as oriented away from the observer. At this point, the in-depth orientation of the point-light figure is, for a particular observer, subjectively ambiguous. Analogous to other measures of subjective equality (PSE, e.g. Ernst & Banks, 2002) or subjective simultaneity (PSS, e.g., Arrighi, Alais, & Burr, 2006) we refer to this point as the *point of subjective ambiguity (PSA)*.

To psychophysically determine the PSA for a particular observer, the first prerequisite is to identify a physical variable that reliably affects the perceived in-depth orientation of a point-light figure. One such variable is the amount of perspective information that is conveyed by the figure. Indeed, Vanrie et al. (2004) showed that the perceived in-depth orientation can be altered by introducing perspective cues. However, in their experiment, the perspective manipulation was only categorical. That is, Vanrie et al

(2004) showed that by using a (strong) perspective projection instead of an orthographic projection the facing bias can be countered. The perspective projection of a point-light figure who is oriented away from the viewer elicited the in-depth orientation interpretation consistent with the figure's veridical orientation. The second prerequisite to psychophysically determine the PSA, however, is that the physical variable can be parametrically controlled and that parametrical variations of the variable lead to gradual variations in the perceived in-depth orientation. The goal of the current study is to verify whether indeed parametrically manipulating the amount of perspective information in a point-light figure can provide a way (1) to gradually manipulate the perceived in-depth orientation and (2) to determine the PSA for each observer individually.

In this study, we first present a technique that offers complete experimental control on the perceived in-depth orientation of a point-light stimulus by parametrically manipulating the amount of available perspective information. Second, we experimentally validate the technique. Finally, we discuss potential future research and provide the program code of a demo that can serve as a helpful starting point in programming future experiments.

TECHNIQUE

Generally, observers are unaware of the fact that what they visually perceive is actually the result of a reconstruction of the 3D world projected on the 2D retinal surface. Whilst there are various ways to graphically project 3D objects to a 2D surface, two types of projections are specifically relevant to the current work: orthographic projection and perspective projection.

Consider a human 3D figure in a 3D space that is only represented by the mere collection of small spheres positioned in the middle of the major joints (shoulders, elbows, wrists, hips, knees, and ankles) and the head. Consider also the projection of this 3D figure on a 2D surface (e.g., the computer screen or a piece of paper), parallel to its coronal plane. In an *orthographic projection*, in which all projection lines are parallel, any information regarding the distance between the 3D figure and the projection plane is effectively removed. Hence, changing the location of the spheres in depth will not change their location on the projection plane in any way. This projection effectively removes all perspective cues to depth (see Movie 3). In a *perspective projection,* on the contrary, projection lines are converging at one point (vanishing point). Hence, changing the

distance between the collection of spheres and the projection plane will change the size, shape, and location of the 2-D projection of the spheres on the projection plane. This projection does contain perspective cues to depth (see Movie 4).

In their study, Vanrie et al. (2004) applied the perspective projection to a 3D point-light figure as a categorical manipulation. That is, the perspective projection consisted of only one particular observer-source distance with a fixed field of view angle. As their main goal was to examine whether a strong global cue would render the true orientation of the figure more perceptible, they deliberately chose a projection distance that yielded a relatively strong perturbation of the point-light configuration. By doing so, they were able to succesfully disambiguate the figure. Notably, Vanrie et al. (2004) refer to the perspective projection as an important tool to render the true orientation of the point-light figure more visible. The main rationale of the present study was to examine whether it would be possible to *gradually* manipulate the amount of perspective information present in a point-light walker to further exploit the usefulness of this tool.

To gradually vary the available perspective information in a point-light walker, in a perspective projection the distance between the 3D figure and the projection plane was parametrically manipulated. When the distance between the 3D figure and the projection plane is decreased, the position, size, and shape distortions of the projection of the spheres on the projection plane increase. These distortions are in fact perspective cues. So, gradually boosting these distortions by further decreasing the distance between the 3D figure and the projection plane gradually exaggerates the perspective cues.

As a result of decreasing the distance (D) between the 3D figure and the projection plane, however, the size of the projected image (S) and hence also the overall size of the projected point-light figure, increases. To overcome this side effect, it is sufficient to proportionally increase the field of view angle (V) of the perspective projection according to the following formula: V = 2*arctan[S/(2*D)]. This technique to incorporate perspective cues to depth is, among visual artists, also known as *foreshortening*. Movie 5 illustrates how the 2D projection of the point-light figure varies as a result of the manipulation explained above.

EXPERIMENT

Crucial for the applicability of the technique is that the parametric stimulus manipulations also lead to gradual and meaningful changes in the perceived orientation of

a biological-motion figure. To test this, we conducted an experiment in which we parametrically increased the perspective information of a point-light walker who was oriented *away* from the observer. By measuring the proportion of 'facing the viewer' (FTV) responses to a walker who was actually walking away, as a function of the amount of perspective information, we were able to verify whether the tendency to perceive the walker as facing the viewer (the facing bias under orthographic projection) could be gradually decreased and altered to the interpretation that matches its veridical in-depth orientation.

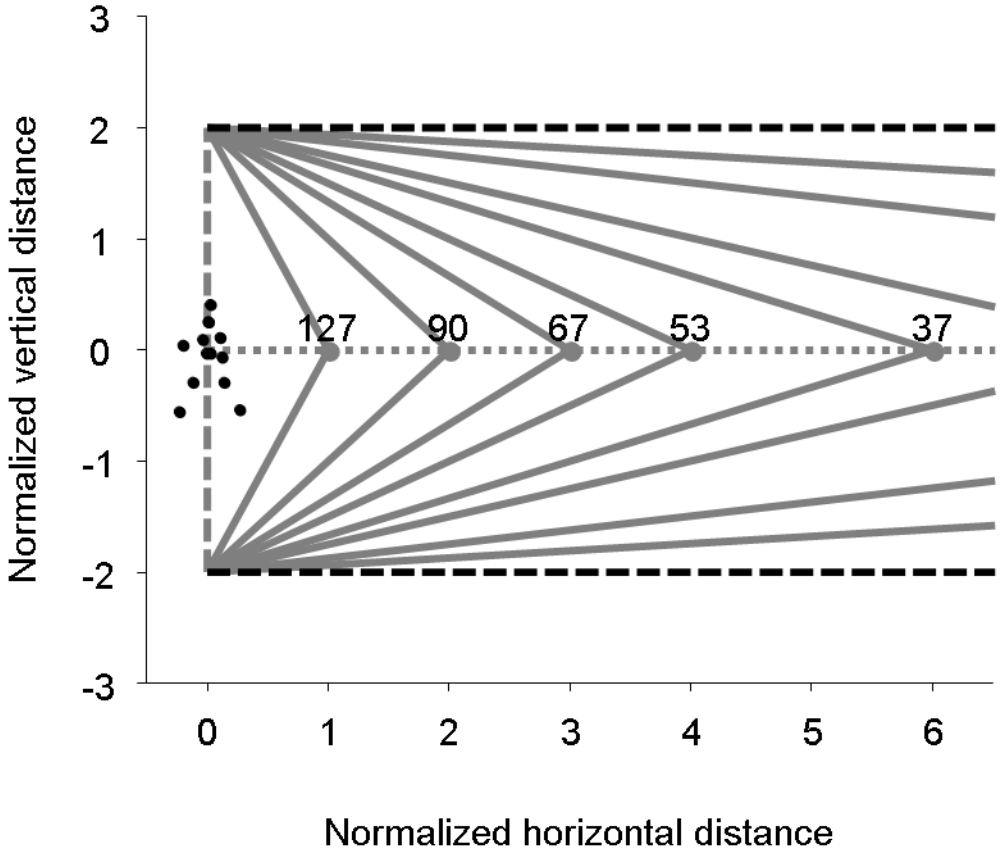


Figure 3. Schematic illustration of the lateral 2D view of the 3D projection space. The horizontal and vertical distances are normalized to the height of the point-light walker. The horizontal black dashed lines represent the projection lines of the orthographic projection. The grey solid lines represent the projection lines of the perspective projections. The black dots represent the point-light walker. The vertical grey dashed line indicates the size of the projected image (S). The grey dots illustrate the distance (D) between the projection plane and the point-light figure (only 6, 4, 3, 2, and 1 are shown). The angles that are formed between the grey lines at each grey dot illustrate the field of view angles (only 37, 53, 67, 90, and 127 degrees are shown). The shorter the distance D, the larger the field of view angle is.

METHODS

Participants. Nine psychology students of the University of Leuven participated for course credit. All observers had normal or corrected to normal vision and were naïve to the aims of the experiment.

Stimuli. All point-light figures were derived from the same 3D-coordinates of a human walker of which the translational component was removed (Dekeyser et al., 2002). As a result, the walker moved as if walking on a treadmill, in the center of the screen. The walker was facing away from the viewer. The perspective cues carrying the information concerning the veridical orientation of the point-light figure were gradually manipulated (nine levels) from absent (orthographic projection) to very strong. The nine levels of perspective information were composed of one orthographic projection (green dashed line in Movie 5A and horizontal grey dashed line in Figure 3) and eight distance (D) manipulations (blue dots in Movie 5A and grey dots in Figure 3) of the convergence point, corresponding to 32, 16, 8, 6, 4, 3, 2, and 1 times the height of the walker. This resulted respectively in field of view angles of 7, 14, 28, 37, 53, 67, 90, and 127 degrees (See Movie 5A and Figure 3, angle between solid lines). The height of the point-light figures (adjusted according to the viewing distance, as discussed above) subtended about seven degrees of visual angle. Hence, the height of the projected image (screen height, four times the height of the walker) subtended about 28 degrees. Each dot subtended about 15 arc mins. Each stimulus presentation lasted about 4.5 s consisting of three walking cycles, each consisting of two steps. The start position of the animation cycle was randomized across trials.

Procedure. Observers were seated in a dimly lit and sound attenuated room. The head was immobilized by means of a chin rest. Viewing distance was 57cm from a CRT monitor (refresh rate = 85 Hz). On each trial, observers had to indicate by a key-press (arrow down for FTV, arrow up for FA) whether the visually presented stimulus was perceived as oriented towards or away from them. They were instructed to respond according to their own subjective experience and it was stressed that an equal distribution of both response alternatives was not necessary. In total, each observer completed 360 (9 levels * 40 repetitions) trials.

RESULTS

Figure 4 plots the proportions of FTV responses for the orthographic projection, the proportions FTV responses as a function of the field of view angle for the perspective projection, and the PSA for each of the nine observers. Inspection of the figure points to two important observations. First, when the point-light walker is projected orthographically, that is, when no perspective cues are present, all observers are biased to interpret the figure as facing towards the viewer (horizontal dashed line), in line with the effect reported by Vanrie et al. (2004). This is confirmed by the high proportion of FTV responses for each observer (mean = 0.87; min = 0.68; max = 1.0) in the orthographic condition. Second, proportions of FTV responses gradually drop as more perspective cues are added to the figure walking away. Moreover, proportions of FTV responses as a function of field of view angle are well fit with a cumulative Gaussian function, as confirmed by the goodness of fit test on each individual dataset: Deviance of the observations to the cumulative Gaussian fit (Wichmann & Hill, 2001a, 2001b) was not significantly higher than simulated deviance (10000 Monte Carlo simulations) for eight out of nine observers (all p values > 0.05, except for observer S.M., p = 0.01). PSA values of the nine observers, expressed as the field of view angle, vary between 29 degrees (R.S.) and 51 degrees (K.N.) with a mean of 40 degrees (SD = 6.64). Slopes vary between -0.011 and -0.080 with a mean of -0.023 (SD = 0.022). Note that the choice to express the ordinate values as the field of view angle of the perspective projection is somewhat arbitrary. Proportions FTV responses can be just as well plotted and fitted as a function of e.g., the distance D or the logarithm of the distance D.

DISCUSSION

First, earlier findings demonstrating that orthographically projected point-light figures elicit a strong facing bias (Vanrie et al., 2004) are confirmed. Second, the results indicate that the technique presented here is well suited to gradually manipulate the perceived in-depth orientation of a bistable biological motion figure: The perceived in-depth orientation of a point-light walker who is oriented away from the viewer gradually changes from the ‘facing the viewer’ interpretation (facing bias) to the veridical ‘facing away’ interpretation when perspective cues are introduced parametrically. Proportions of FTV responses as a function of added perspective cues are, at individual observer level,

well fitted with a cumulative Gaussian psychometric function. The single significant deviance for the fit to the data of observer S.M indicates overdispersion (unexplained variability), presumably provoked by unstable behavior in the task.

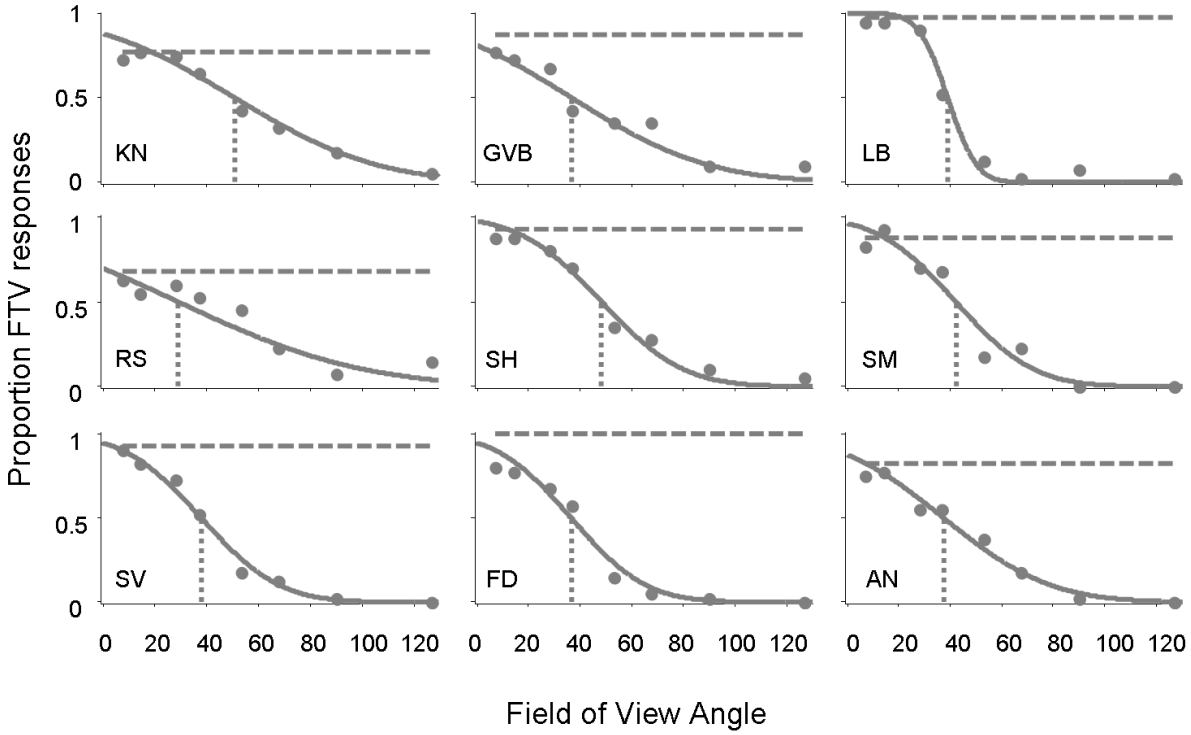


Figure 4. Proportions of FTV responses as a function of field of view angle for nine observers. The horizontal dashed line represents the proportion of FTV responses when no cues to depth are present (orthographic projection). Filled dots represent the proportions of FTV responses for all perspective projections and the solid curved line represents their cumulative Gaussian fit. The vertical dotted line represents the point of subjective ambiguity (PSA), corresponding to 50% FTV responses.

In our experiment, the mean of the cumulative Gaussian function indicates the PSA, the stimulus of which the perceived in-depth orientation is subjectively perfectly ambiguous. This 50% threshold gives an estimation of the amount of perspective that a particular observer, or a group of observers, needs to counter the facing bias. Moreover, the PSA can be a good reference point in studying factors affecting the perceived in-depth orientation. By comparing the proportions of FTV responses of the subjectively ambiguous point-light walker with proportions of FTV responses to the same ambiguous walker but with added contextual cues, the disambiguating effect of the contextual cues can be directly measured as an increase or decrease in the proportion of FTV responses. Taking

the individual PSA as a baseline for each observer gives the opportunity to standardize subjective experiences across observers and to prevent ceiling and floor effects.

To further illustrate the potential usefulness of the experimental tool presented here, we briefly discuss two lines of research we are currently pursuing. First, Brooks et al. (2008; see Chapter 3) recently reported a striking correlation between the depicted gender and the perceived in-depth orientation of a bistable biological-motion walker: Male walkers generally were perceived as approaching the observer, whereas female walkers were seen as retreating. To examine the effect in more detail, a technique similar to the tool presented here was used: Perspective cues disambiguating the walker's facing direction were gradually added to the figure judged as most female, to the figure judged as most male, and to a gender-neutral walker. As expected, perspective cues affected facing judgments, but much stronger cues signalling a receding conspecific had to be added to the neutral and male walker than to the female walker, in order to change the 'facing bias' for the male and neutral walker into a 'retreating bias'. Second, the technique offers opportunities to study crossmodal perception of actions. There is indeed evidence that mammalian observers represent actions in an audiovisual format (e.g., Barraclough, Xiao, Baker, Oram, & Perrett, 2005; Bidet-Caulet, Voisin, Bertrand, & Fonlupt, 2005; Keysers, Kohler, Umilta, Nanetti, Fogassi, & Gallese, 2003; Kohler, Keysers, Umilta, Fogassi, Gallese, & Rizzolatti, 2002; also see Chapter 7), rather than in separate visual and auditory codes that are merged only in a relatively late processing stage. Research on audiovisual integration has suggested that auditory effects on visual perception are strongest when the visual information is weak or ambiguous – also known as the rule of inverse effectiveness (Meredith, & Stein, 1983; but see Holmes, 2007). Hence, when the visual information concerning the in-depth orientation of a point-light walker is maximally ambiguous, unambiguous auditory looming or receding information might affect the perceived in-depth orientation most strongly. Our technique could be a useful tool to examine this issue. Pilot psychophysical experiments (Menodonça & Santos, 2008; Schouten, Moyens, Brooks, van der Zwan, & Verfaillie, 2008) indeed suggest that looming and receding sounds can affect the perceived in-depth orientation of a point-light walker (also see Chapter 11). More research is necessary to test at what processing level looming and receding sounds can affect the perceived in-depth orientation of a multistable point-light figure.

Besides pointing to the advantages and the possibilities offered by this technique we also want to be clear on its limitations. First, because the technique is based on an exaggeration of the distortions related to changes in depth, it probably will be less efficient

for point-light actions in which all joints/dots remain in the same depth plane (orthogonal to the observer's line of sight). For example, a frontal/back view of a point-light jumper in which the arms and legs only move sideways in the figure's coronal plane, would not be an appropriate stimulus for our technique because it contains only a small amount of potential perspective information. Interestingly, the presence or absence of a facing bias to a frontal/back view indeed seems to be related to the amount of motion in the sagittal plane: Actions that do not contain substantial motion in the sagittal plane generally do not elicit a facing bias (Vanrie & Verfaillie, 2006). Second, the distortions at the projection plane include distortions of the shape, size, and position of the dots that may function as several local cues instead of one global cue. The magnitude of the distortion increases proportionally with the amplitude of the horizontal, vertical and depth displacement of the dot relative to the center of the figure. It is not clear yet to what extent these individual distortions contribute to the overall change in perceived orientation. Therefore, it would be interesting to test to what extent isolated shape, size or position changes lead to changes in the perceived in-depth orientation.

To support future research we provide the code of a demo (see Appendix below) that can serve as an easily modifiable starting point for new experiment code.

CONCLUSION

The technique presented here enables the study of factors that affect the perceived in-depth orientation over the entire range of possible depth interpretations (from convincingly facing the viewer to convincingly facing away from the viewer). More specifically, the technique makes it possible to express factors affecting the in-depth interpretation quantitatively, as changes in the mean (PSA) and slope of a psychometric curve. We are convinced that a systematic experimental study of factors affecting the perceived in-depth orientation will shed light on the processes of depth assignment in point-light figures and will help to identify the causal factors in the intriguing and robust bias to interpret some point-light figures as facing toward the observer. We hope that the tool provided here will contribute to this research.

APPENDIX

To support future research we provide the code of a demo that can serve as an easily modifiable starting point for new experiment code. See http://brm.psychonomic-journals.org/content/42/1/161/suppl/DC1.

Presentation (Neurobehavioral Systems) is a stimulus delivery and experimental control program. If you don't have Presentation installed on your computer, to run the code provided here, a free trial of the software with full functionality for 30 days can be downloaded from http://www.neurobs.com/nbs_online/presentation/trial_procedure.

The demo presents a continuously walking point-light walker in a frontal/back view. The demo starts with showing an orthographic projection of a point-light walker. Pressing the left arrow repeatedly, gradually adds more perspective information that the figure is facing away. Pressing the right arrow gradually adds more perspective information that the figure is facing the viewer. Below the figure, the angle of the perspective projection is shown (only for perspective projections).

To view the demo:

1. Download, install, and activate Presentation: follow the instructions provided on http://www.neurobs.com/nbs_online/presentation/trial_procedure.
2. Download the following files into the same folder on your computer:

Demo_Schouten_Verfaillie_2009.exp (Experiment file containing the settings of the demo)

Demo_Schouten_Verfaillie_2009.sce (Scenario file containing the code of the demo)

Walking.txt (Coordinate file containing the 3D coordinates of the point-light figure)

3. Double click the .exp file in your folder. This opens Presentation with the correct settings.
4. Double click the .sce file in your folder. This opens the scenario file in the editor window of Presentation.
5. Press F7 to compile and then F5 to run the script.
6. Press left and right arrow while the demo is running to change the perspective information.
7. Press ESC to stop the demo.

Chapter 3

Correlated changes in perceptions of the gender and orientation of ambiguous biological motion figures

This chapter is based on:

Brooks, A., Schouten, B., Troje, N. F., Verfaillie, K., Blanke, O., & van der Zwan, R. (2008). Correlated changes in perceptions of the gender and orientation of ambiguous biological motion figures. *Current Biology*, *18*(17), 728-729.

ABSTRACT

The sensitivity of the mammalian visual system to biological motion cues has been shown to be general and acute (Blake, 1993; Herman, Morrel-Samuels, & Pack, 1990; Johansson, 1973). Human observers, in particular, can deduce higher-order information, such as the orientation of a figure (which way it is facing), its gender, emotional state, and even personality traits, on the basis only of sparse motion cues. Even when the stimulus information is confined to point lights attached to the major joints of an actor (so-called pointlight figures), observers can use information about the way the actor is moving to tell what they are doing, whether they are a male or female, and how they are feeling (Dittrich, 1993; Dittrich, Troscianko, Lea, & Morgan, 1996; Kozlowski & Cutting, 1977). Here we report the novel finding that stimulus manipulations that made such walkers appear more female also had the effect of making the walkers appear more often as if they were walking away from rather than towards observers. Using frontal-view (or rear-view) point-light displays of human walkers, we asked observers to judge whether they seemed to be walking towards or away from the viewing position. Independent of their own gender, observers reliably reported those figures they perceived to be male as looking like they were approaching (as reported in Vanrie, Dekeyser, & Verfaillie, 2004), but those they perceived to be female as walking away. Furthermore, figures perceived to be gender-neutral also appeared more often, although not exclusively, to be walking towards observers.

EXPERIMENTS

In Experiment 1a, observers judged the gender, male or female, of each walker. Mean ratings across observers are illustrated in Figure 5 (right ordinate, solid circles). A cumulative Gaussian was fitted to the ratings of each figure. Female walkers uniformly were judged as being female, and male walkers as male. The point of subjective gender ambiguity corresponded to the middle walker (indicated in Figure 5 as 0) of the 13 stimuli used. In Experiment 1b, the same observers were asked, when viewing the same stimuli, to judge whether or not each figure was walking towards them. Means were calculated, for each walker, as a proportion of the number of times observers judged that walker to be facing (walking towards) them and a cumulative Gaussian was fitted to the function (Figure 5, solid triangles, left ordinate). The point of subjective in-depth orientation

ambiguity — that is, the point at which observers were equally likely to judge the figure as walking towards them as away from them — corresponded to a female walker. In other words, all walkers perceived to be male were judged by observers to be facing towards them. Only when the walkers had characteristics consistent with being female did observers begin to perceive them more often as facing away. While not unequivocal these data do suggest that perceived direction-of-facing is related to the perceived gender of a point-light walker.

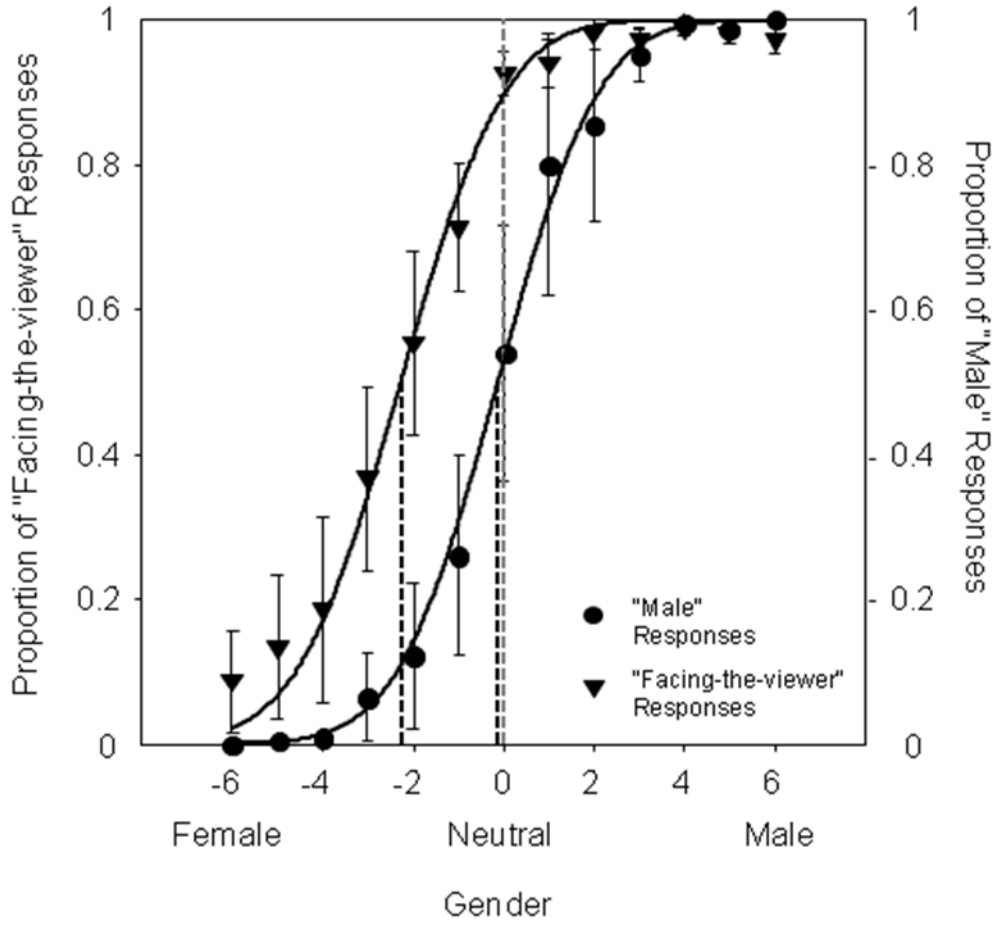


Figure 5. Mean proportions of 'male' answers (solid circles) and 'facing the viewer' answers (solid triangles), both as a function of the 13 point-light walkers used in Experiment 1. The gender dimension (measured as Z-scores) extends from extreme female (–6) to extreme male (+6); The gender-neutral point (0) is indicated by a broken vertical grey line. The mean of each function is indicated with a broken vertical black line. As with earlier studies (Troje, 2002) the morphed figures give rise to gender perceptions that range from female to male with the point of subjective neutrality around 0. These data show that only the point-light figures perceived as walking away from observers are those perceived as most female. Error bars indicate ±1 SE.

In Experiment 2, we examined the effect of perspective information on perceptions of orientation (cfr. Chapter 2). Three point-light walkers used in Experiment 1, those judged as most female (–6), most male (+6), and gender-neutral (0), were presented again on a treadmill but this time with perspective information. The three walkers were presented to observers, this time incorporating visual cues consistent with the walker's approach or retreat. The addition of perspective cues was expected to standardise observers' perceptions across the three walkers if perspective affects perceived orientation more strongly than gender. Conversely, if perceived orientation is affected more strongly by the gender than the perspective of the walker the gender differences observed in Experiment 1b will persist. The gender effects reported for Experiment 1 were confirmed. Although perspective cues modified facing judgements for each figure (female, neutral, male) from convincingly looming to convincingly receding, Figure 6 shows orientation biases that were consistent with the findings from Experiment 1b: male walkers were reported as facing towards observers on more than 50% of trials for all perspective cues except for the two walkers with perspective cues most strongly suggesting retreat. A similar pattern of results was observed for the neutral walkers. This was different for female walker stimuli which frequently were judged as facing away from observers even when perspective cues unambiguously signalled looming. Figure 6 (insert) illustrates those points by plotting the means of the fitted functions against the perspective cues. Veridical (non gender-biased) use of the perspective cues would place the mean for each of the walker stimuli at the perspective-cue neutral point (broken horizontal grey line). Instead, the means of the functions arising from judgements of the male and gender-neutral walkers are above the neutral point, consistent with those walkers appearing to approach even when there were unambiguous perspective cues to retreat. Conversely, the mean of the function fitted to judgements of the female walker is below the neutral point, consistent with female walkers appearing to retreat even when there were unambiguous perspective cues to approach. These effects were independent of observer gender (also see Chapter 4). The results demonstrate that the perceived direction-of-facing of an otherwise ambiguous point-light figure is related to the perceived gender of the figure (also see supplementary information[5]).

[5] Supplementary information available at http://www.current-biology.com/cgi/content/full/18/17/R728/DC1

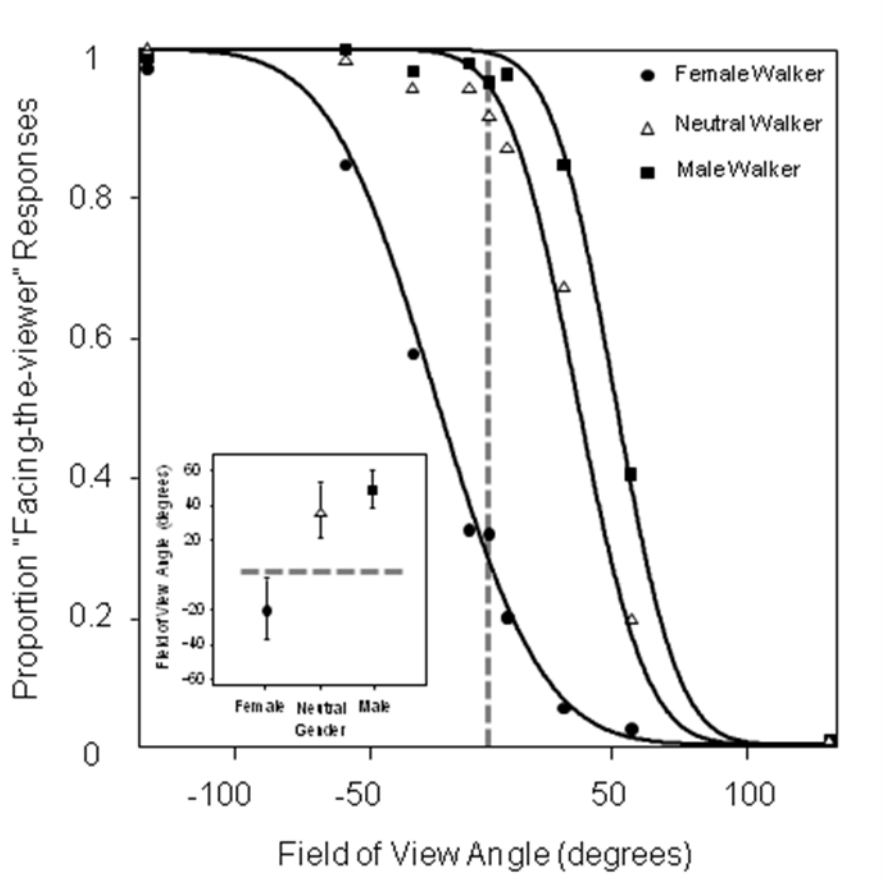


Figure 6. Proportions 'facing the viewer' answers as a function of perspective information, represented by field-of-view angle. Negative scores denote perspective cues to looming and positive scores perspective cues to retreat. The dashed grey line indicates the perspective-cue neutral point (orthographic projection). Female walkers (filled circles) elicited mainly perceptions consistent with retreat. Neutral (open triangles) and male walkers (filled squares) elicited mainly perceptions consistent with approach. The means of the Gaussian functions fitted to each set of data (inset, bottom left) show that judgements of the orientation of the walkers could not be predicted on the basis of the perspective cues incorporated into each stimulus. Bars indicate 95% confidence intervals about each mean.

CONCLUSION

Humans, like many other species, are fundamentally social animals and have evolved mechanisms allowing them successfully to work in large social groups (Allison, Puce, & McCarthy, 2000). Our data suggest that biological motion is an important cue for social organisms trying to operate in environments where other cues as to the actions or intentions of other organisms may be ambiguous. Whilst the precise role of local cues in mediating these effects requires further explication, it is tempting to speculate that the

orientation biases reported here reflect the development of perceptual mechanisms that weigh in the probable cost of *misinterpreting* the actions and intentions of others. For example, a male figure that is otherwise ambiguous might best be perceived as approaching to allow the observer to prepare to flee or fight. Similarly, for observers (especially infants) the departure of females might signal also a need to act, but for different reasons.

Chapter 4

The facing bias in biological motion perception: Effects of stimulus gender and observer sex

This chapter is based on:

Schouten, B., Troje, N. F., Brooks, A., van der Zwan, R., & Verfaillie, K. (2010). The facing bias in biological motion perception: Effects of stimulus gender and observer sex. *Attention, Perception & Psychophysics,72*(5), 1256-1260.

ABSTRACT

Under orthographic projection, biological motion point-light walkers offer no cues as to the order of the dots in depth: Views from the front and from the back result in the very same stimulus. Yet, observers show a bias towards seeing a walker facing the viewer (Vanrie, Dekeyser, & Verfaillie, 2004). Recently, we reported that this facing bias strongly depends on the gender of the walker (Brooks, Schouten, Troje, Verfaillie, Blanke, & van der Zwan, 2008). The goal of the current study was, first, to examine the robustness of the effect by testing a much larger subject sample and, second, to investigate whether the effect depends on observer sex. Despite the fact that we found a significant effect of figure gender, we clearly failed to replicate the strong effect observed in the original study. We did, however, observe a significant interaction between figure gender and observer sex.

INTRODUCTION

In 1973, Johansson demonstrated that despite degradation of visual information to just a few dots (point lights) attached to the major joints of an otherwise invisible actor, observers are able to effortlessly recognize human actions. Since then it has been reported that humans are able to extract a variety of behaviorally relevant characteristics including the type of action (Dittrich, 1993; Vanrie & Verfaillie, 2004), the identity of the actor (Cutting & Kozlowski, 1977; Loula, Prasad, Harber, & Shiffrar, 2005), their gender (Kozlowski & Cutting, 1977; Pollick, Kay, Heim, & Stringer, 2005; Troje, 2002), emotional state (Dittrich, Troscianko, Lea, & Morgan, 1996; Montepare, Goldstein, & Clausen, 1987), and age (Montepare & Zebrowitz-McArthur, 1988). Only more recently, however, has the question of how observers interpret the in-depth orientation of point-light figures systematically been addressed (Jackson, Cummins, & Brady, 2008; Vanrie, Dekeyser, & Verfaillie, 2004; Vanrie & Verfaillie, 2006, 2008).

A point-light walker without explicit depth cues constitutes a perfectly ambiguous stimulus. Any non-lateral view of the walker could in principle be the result of a 2D projection of two different veridical 3D orientations. For instance, the orthographic projection of a 3D walker facing the viewer in frontal view is identical to the projection of a walker who faces away from the viewer. Both interpretations are equally plausible in principle. Yet, observers exhibit a perceptual bias to interpret the figure as being oriented

towards the viewer (Vanrie et al., 2004). An orthographically projected frontal/back view point-light walker generally elicits more than 80% 'facing the viewer' (FTV) responses.

To date, the causal factors of this 'facing bias' are unclear. As argued by Vanrie et al. (2004) there is in principle no reason to assume that observers are confronted visually more often with frontal views than with back views of their conspecifics. Consequently, an explanation based on familiarity seems inadequate. An account based on social and biological relevance offers an alternative possibility. One could argue that, in most cases, another person facing the observer probably is socially more relevant than a person facing away. From this perspective, the perceptual system might take into account the potential cost of misinterpreting the actions and intentions of others: Expecting someone to approach who is actually retreating might be less costly than misinterpreting an approach for a retreat. Such an account is consistent with existing explanations of the well documented perceptual bias for rising tones, which signal a looming object or entity in the auditory modality (Bach et al., 2008; Neuhoff, 1998), or of the preferential responsiveness to visual and audiovisual looming signals (Maier, Neuhoff, Logothetis, & Ghazanfar, 2004; Schiff, Caviness, & Gibson, 1962).

Following this line of reasoning, we (Brooks et al., 2008; Chapter 3; henceforth referred to as 'initial study') recently investigated the effect of the perceived gender of an ambiguous point-light figure on its perceived in-depth orientation. In the study of Vanrie et al. (2004), only male walkers were shown. However, if social and biological relevance indeed affect perceptual in-depth organisation then, we reasoned, the cost of misinterpreting actions could be different for the perception of male conspecifics than for the perception of female conspecifics. For instance, an approaching male might be more threatening than an approaching female (see Gunns, Johnston, & Hudson, 2002, for related research on the opposite effect, namely how movement information can specify how vulnerable another person seems to be). Therefore, male and female ambiguous point-light figures might elicit different degrees of the facing bias. Using depth ambiguous point-light figures in frontal/back view, and varying gender on a continuum from extremely male to extremely female (see Troje, 2002), we collected both gender ratings and in-depth orientation ratings. A strong correlation between perceived gender and perceived orientation of ambiguous point-light walkers was observed. Consistent with the facing bias found by Vanrie et al. (2004), figures perceived as male generally were perceived as facing the viewer. The remarkable finding was that, the more the point-light figures were perceived as female, the more the figures were perceived as facing *away* from the viewer.

The purpose of the present study was twofold. First, the number of participants tested in the initial study was relatively small (n=5). We examined the robustness of the effect by testing a larger subject sample. Second, in the initial study, no observer sex differences were found. In fact, since it was not the goal of that study, the sample size was too small (two males and three females) to reliably test for an effect of observer sex. To this end, in the present study 20 females and 20 males completed the same task as in Experiment 1 of Brooks et al. (2008).

METHOD

Participants. Forty students (20 females and 20 males) from the University of Leuven participated for course credit. All observers had normal or corrected to normal vision and were naïve to the purpose of the experiment.

Stimuli and apparatus. Stimuli and procedure were designed to be as similar as possible to Experiment 1 of Brooks et al. (2008). Each point-light walker was presented in frontal view in the centre of the screen (subtending a size of 8º of VA height) and each consisted of 15 black dots (radius = 15 arc mins of VA) on a grey background. The gender of the point-light figures was gradually manipulated from extremely female to extremely male. As in the initial study, the two-dimensional positions of the point lights were derived from the gender continuum described in detail by Troje (2002). Seven sets of two-dimensional positions were chosen so as to be equally spaced between extremely female and extremely male (z-scores: -6, -4, -2, 0, 2, 4, 6 SD, based on the distribution of 100 individual walkers which span the walker space, for details see Troje, 2008b). Each stimulus presentation consisted of three walking cycles (1.5 s per cycle), each comprised of two steps (128 frames per cycle at 85 fps). The start position of the animation cycle was randomized across trials.

Procedure. All observers were tested individually. They were seated in a dimly lit and sound attenuated room in front of a 21 inch CRT monitor. Viewing distance was 57 cm. The experiment consisted of two phases: A ‘gender judgement’ phase and a ‘facing judgement’ phase. To control for the effect of order, the order of the gender judgement task and the facing judgement task was counterbalanced across observers. Before the start of the experiment, observers were only informed about the contents of the first task (gender or facing judgement) and were unaware about the contents of the second task (facing or gender judgement, respectively). Observers were instructed to indicate on each trial by a

key-press whether the presented stimulus was perceived as facing towards them or facing away (FA) from them (arrow down for FTV, arrow up for FA) in the 'facing judgement' phase, or whether the presented stimulus was perceived as female or male (left arrow, right arrow) in the 'gender judgement' phase. They were instructed to respond according to their own subjective experience and it was stressed that an equal distribution of both response alternatives was not necessary. After instructions were given, observers completed a practice block (random selection of all possible conditions) in which they were familiarized with the stimuli and the task. Then it was checked again whether they had understood the task and the experiment commenced. Within each phase, trials were randomized and divided in blocks of 35 between which participants could take a break. In total each participant completed 420 trials (7 genders * 2 phases * 30 repetitions).

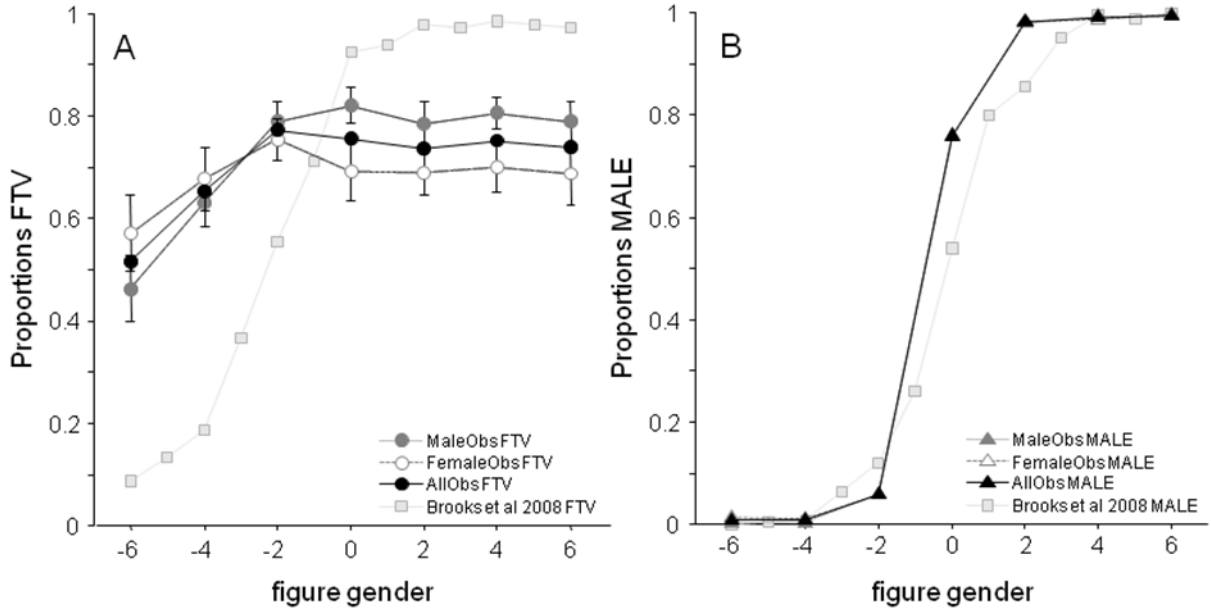


Figure 7. Mean proportions of FTV (A) and MALE (B) responses as a function of Figure gender for all observers (black), female observers (white) and male observers (grey) in the current experiment. Note, that MALE responses from male and females observers were so similar that the three symbols overlap and cannot be distinguished in panel B. Data of Experiment 1 in Brooks et al. (2008) are also depicted (light grey squares). Error bars indicate +- 1 standard error around the mean proportions of FTV responses for male and female observers.

RESULTS

In Figure 7A, we plot the proportions of FTV responses as a function of figure gender, averaged across all observers (black), female observers only (white), and male observers only (grey). For comparison, we also show proportions of FTV responses of the initial study (light grey squares). Figure 7A clearly shows that the pattern of proportions of FTV responses as a function of figure gender in the current experiment (black circles) is highly different from the pattern reported in Brooks et al. 2008 (light grey squares). In the initial study, we found that the facing bias was strongly modified by figure gender. While we observed close to 100% FTV responses for walkers that appeared extremely male, this percentage dropped to 10% for extremely female walkers.

The current data look quite different. While the overall facing bias was about the same as in the initial study (70% of FTV responses in the current study and 68% in the initial study), the effect of figure gender was much weaker. For gender-neutral walkers and within the male part of the stimulus range, the facing bias was about constant with values around 75% FTV responses. Only in the strongly female part of the range (z-scores smaller than -2 SD), these values dropped and reached a value of 52% for the most extreme female walker at -6 SD. Even though the effect of figure gender was weaker than reported in the initial study, it is still highly significant (two-way repeated measurements ANOVA on probit transformed proportions, $F(6,228) = 7.10$, $p < 0.001$). The order of the phases (gender/facing vs. facing/gender) did not have an effect on the facing bias ($F < 1$) nor did it interact with the figure gender effect ($F < 1$).

Figure 7A also depicts the effect of observer sex on the relation between figure gender and facing bias. While observer sex itself did not show a main effect on facing bias, we found a significant interaction between figure gender and observer sex ($F(6,228) = 2.24$, $p < 0.05$). Figure gender seems to have a stronger effect on male observers than on female observers. Compared to female observers, male observers showed a generally higher facing bias for male walkers.

Figure 7B plots the proportions of MALE responses as a function of figure gender, averaged across all observers (black), female observers only (white), and male observers only (grey). For comparison, the proportions of MALE responses of the initial study are also depicted (light grey squares). The data show, that the walkers at the two ends of the figure gender spectrum were unanimously rated either female or male by every single observer and for any single trial. Ambiguous ratings were only found in the range between

-2 and +2 SD. In contrast to the data obtained in the initial study (mean = -0.0970, CI_{95} = [-0.27, 0.08]; proportions fitted with a cumulative Gaussian; confidence intervals based on 10000 Monte Carlo simulations; see Wichmann & Hill, 2001a, 2001b), the curve was shifted a little to the left (mean = -0.61, CI_{95} = [-.67, -0.55]), indicating a bias towards responding MALE more often than FEMALE. This so-called 'male bias' has been described before (Troje, Sadr, Geyer, & Nakayama, 2006; Troje & Szabo, 2006; van der Zwan et al., 2009). Recovering it in the present data makes us confident that they are more representative than the data obtained from the smaller number of subjects used in the initial study (Brooks et al., 2008; Chapter 3).

Figure 7B also plots the proportions of MALE responses as a function of figure gender for female (white triangles) and male observers (grey triangles), separately. However, the two curves are so similar that they cannot even be distinguished in Figure 7B.

DISCUSSION

In the current experiment we did not replicate the results of Brooks et al. (2008; Chapter 3). Indeed, the pattern of proportions of FTV responses as a function of figure gender observed here differed quite strongly from the pattern previously observed. We did observe an effect of figure gender, but it was much weaker than in the initial study. Interestingly, we found an interaction between stimulus gender and observer sex. Additionally, in accordance with previous research in the current experiment we observed two types of biases. First, we confirmed the facing bias that was previously reported in studies on the in-depth perception of point-light figures (Schouten & Verfaillie, 2010; Vanrie et al., 2004; Vanrie & Verfaillie 2006, 2008). Second, we observed a "male bias", similar to observations in other studies (Troje et al., 2006; Troje & Szabo, 2006; van der Zwan, et al., 2009) that report gender ratings on the point-light walkers along the gender continuum created by Troje (2002, also see Troje, 2008b).

While the current study constitutes an important rectification to Brooks et al. (2008; Chapter 3), several questions remain. First, it remains puzzling as to why the data in the current experiment are so different from the data reported previously. In the present experiment, 40 observers who were naïve to the goal of the experiment participated, compared to only five observers in the initial study, four of which were experienced psychophysical observers. It could be the case that experienced observers in the initial

study attended to different properties of the stimuli compared to naïve observers. Alternatively, experimenter effects may have influenced the observers' responses in the initial study. We believe that the current data give a more refined and reliable reflection of the perceived in-depth orientation of the point-light figures along the gender continuum.

Next, the question remains as to why the perceived in-depth orientation to some degree depends on the figure gender and why this dependency is modulated by the sex of the observer. In line with what was suggested in Brooks et al. (2008; Chapter 3), one explanation for the effect of figure gender could be that, if the facing bias reflects an adaptation to the higher cost of misinterpreting an approaching person, misinterpreting the approach of a female figure might be less of a risk compared to misjudging the approach of a more dangerous male figure. This view would also provide an hypothesis about the observed interaction between stimulus gender and observer sex: Males are probably more likely to become involved in fights than females. Therefore, the difference in the costs involved with the two kinds of possible errors that an observer can make might be higher for men than for women.

An alternative to this evolutionarily motivated explanation is that simple stimulus properties might be involved that trigger a particular in-depth perception independently of the perceived gender. That is, particular configural or kinematic information correlated with stimulus gender might be responsible for the figure gender effect. For instance, some of the geometric features that distinguish male and female point-light walkers may be misinterpreted as perspective distortions (see Schouten & Verfaillie, 2010; Chapter 2). In that case, the interaction between figure gender and observer sex could be the result of the fact that, especially when it concerns figures that signal a particular gender (Johnson & Tassinary, 2005; Hewig, Trippe, Hecht, Straube, & Miltner, 2008), male observers attend to different parts of the figure than female observers. Future research should investigate whether or not it is possible to identify stimulus properties that trigger a particular in-depth orientation perception in depth-ambiguous point-light figures.

Chapter 5

The facing bias in biological motion perception: Structure, kinematics, and body parts

This chapter is based on:

Schouten, B., Troje, N. F., & Verfaillie, K. (in press). The facing bias in biological motion perception: Structure, kinematics, and body parts. *Attention, Perception & Psychophysics*.

ABSTRACT[6]

Depth-ambiguous point-light walkers elicit a facing bias: Observers perceive a point-light walker as facing towards them more often than as facing away (Vanrie, Dekeyser, & Verfaillie, 2004). While the facing bias correlates with the point-light walker's perceived gender (Brooks, Schouten, Troje, Verfaillie, Blanke, & van der Zwan, 2008; Schouten, Troje, Brooks, van der Zwan, & Verfaillie, 2010), it remains unclear whether the change in perceived in-depth orientation is *caused* by a change in perceived gender. In Experiment 1, we show that structural and kinematic stimulus properties that lead to the same changes in perceived gender elicit opposite changes in perceived in-depth orientation, indicating that the relation between perceived gender and in-depth orientation is not causal. The results of Experiments 2 and 3 further suggest that the perceived in-depth orientation of point-light walkers is strongly affected by locally acting stimulus properties. The facing bias seems to be induced by stimulus properties in the lower part of the walker.

INTRODUCTION

Adequately perceiving and understanding the actions of conspecifics for humans is a prerequisite for normal social development and interaction (Gallese, Keysers, & Rizzolatti, 2004). Since the findings of Johansson (1973), point-light figures have proven to be very useful in studying the visual perception of actions or biological motion. Despite the limitation of the visual information to a small number of moving point lights, observers are able to effortlessly judge a variety of behaviorally relevant properties. Verfaillie (2000), Blake and Shiffrar (2007), and Troje (2008b) provide reviews.

Orthographically projected point-light figures are inherently ambiguous with respect to their in-depth orientation. Based on the stimulus-properties, observers are expected to perceive the figure as facing towards them in half of the cases and as facing away from them in the other half. Vanrie, Dekeyser, and Verfaillie (2004; see Jackson & Blake, 2010, and Jackson, Cummins, & Brady, 2008, for related research on the perception of bistable biological motion stimuli), however, demonstrated that this is not what is empirically found when observers have to indicate the perceived in-depth orientation for a set of depth-ambiguous point-light actions. Several point-light stimuli, especially point-light walkers, elicit a strong facing bias. That is, observers perceive the walker as facing

[6] Thanks to Ilse Van Dromme for her help in data collection.

towards them in most of the cases (also see Vanrie & Verfaillie, 2008). To date, it is not clear what causes this perceptual bias. One could assume, for instance, that in daily life, observers are more often confronted with a person facing towards themselves than with a person that is oriented away. There are, however, no natural image statistics available that show that this is indeed the case (Vanrie et al., 2004). Otherwise, it could be hypothesized that a person who is facing towards an observer is more socially or behaviourally relevant for the observer than a person who is oriented away. The facing bias might thus reflect an evolved perceptual bias to err at the side of caution. In other words, perceiving someone who is walking away from you as walking towards you comes with a much lower cost compared to misinterpreting someone who is actually walking towards you, as walking away.

To explore whether biological or social relevance drives the facing bias, Brooks et al. (2008; Chapter 3) recently investigated the role of the gender of the point-light walker on its perceived in-depth orientation. Many studies have shown that observers easily pick up on the cues that signal the gender of a point-light figure (see e.g., Barclay, Cutting, & Kozlowski, 1978; Cutting, 1978a; Hiris, & Ewing, 2010; Jordan, Fallah, & Stoner, 2006; Kozlowski & Cutting, 1977; Mather & Murdoch, 1994; Pollick, Kay, Heim, & Stringer, 2005; Troje, 2002; Troje, Sadr, Geyer, & Nakayama, 2006; van der Zwan et al., 2009). Assuming that male and female point-light walkers differ in biological or social relevance, male figures might elicit different degrees of the facing bias than female figures. For instance, an approaching male might be more threatening than an approaching female. Brooks et al. (2008) found a strong effect of figure gender on perceived in-depth orientation. Consistent with Vanrie et al. (2004), male point-light walkers elicited a strong facing bias. However, the more strongly the figure was perceived as female, the more strongly it was perceived as facing away from the viewer. Schouten et al. (2010; Chapter 4) further investigated the effect of figure gender and the role of observer sex in a larger sample of male and female observers. While much weaker compared to what was observed in Brooks et al. (2008), Schouten et al. (2010) again observed a significant figure gender effect: Male figures elicited a stronger facing bias than female figures. Remarkably, Schouten et al. (2010) also observed a small but significant interaction between figure gender and observer sex: The facing bias for male point-light walkers appeared to be stronger for male observers than for female observers.

The results of Brooks et al. (2008) and Schouten et al. (2010) might be interpreted as evidence that perceived gender is an important causal variable in the facing bias.

However, although both studies demonstrated a correlation between perceived gender and perceived in-depth orientation, they did not show that the relation is causal, namely that the male figure was perceived as facing the viewer because it was interpreted as a male walker and the female figure was perceived as facing away because it was seen as a female walker. Indeed, particular configural or kinematic stimulus properties that are related to the change in figure gender might, regardless of the perceived gender they elicit, be responsible for the change in perceived in-depth orientation. Vanrie and Verfaillie (2006), for example, have shown that purely kinematic aspects of an action performed by the same male actor strongly affect depth assignment: The presence of a preferred interpretation depends on the performed action. One could thus argue that the hypothesis of a causal link between perceived gender and perceived in-depth orientation becomes less probable when taking into account the fact that not all male point-light actions elicit a facing bias. Nevertheless we reasoned that the correlation found in Brooks et al. (2008) and Schouten et al. (2010) for walking needed to be investigated more directly.

The goal of the current study was therefore to find out whether the change in perceived gender causes the figure gender effect or whether stimulus-based factors bias the perceived in-depth orientation. In Experiment 1 we tested the role of structural and kinematic information. More specifically, we tested whether structural and kinematic variations that lead to the same changes in perceived gender, also lead to comparable changes in perceived in-depth orientation. In Experiment 2 we explored whether the facing bias arises from information in a particular part of the stimulus. In particular, we verified whether the upper and lower parts of the point-light figures elicit comparable or different perceived in-depth orientations. In Experiment 3 we investigated the interaction between the variables that were manipulated in Experiments 1 and 2.

EXPERIMENT 1

INTRODUCTION

If biological or social relevance in terms of perceived gender is causing the facing bias then changes in perceived in-depth orientation should be independent of the way the point-light configuration signals gender. Stimulus changes that lead to similar changes in perceived gender should thus lead to similar changes in perceived in-depth orientation. With this in mind, we utilized Troje's (2002) method to decompose the original set of

point-light figures into structure-only and kinematics-only point-light figures, signalling gender from extremely female to extremely male. Structure-only stimuli only vary in structural cues (neutral kinematics) whereas kinematics-only stimuli only vary in kinematic cues (neutral structure). Troje (2002) showed that orthographic frontal/back projections of both sets of stimuli lead to a highly comparable perceived gender.

The goal of Experiment 1 was to elicit a comparable change in perceived gender with two types of gender cues: structure and kinematics. If changes in perceived in-depth orientation are caused by changes in perceived gender, and both gender cues elicit the same perceived gender, then structure-only stimuli should elicit the same change in perceived in-depth orientation as kinematics-only stimuli. If in such case structure-only stimuli and kinematics-only stimuli were to elicit different changes in perceived in-depth orientations, this would suggest that changes in perceived in-depth orientation are not directly caused by changes in perceived gender.

METHOD

Participants. Fourthy-four psychology students (22 males and 22 females) of the University of Leuven participated for course credit. All participants had normal or corrected to normal vision and were naïve to the purpose of the experiment. None of them had participated in previous experiments investigating the facing bias.

Stimuli and apparatus. From the same set of coordinates as used in Brooks et al. (2008) and Schouten et al. (2010), two sets of 7 stimuli signalling gender from extremely female to extremely male (z-scores: -6, -4, -2, 0, 2, 4, 6 SD) were created: structure-only stimuli and kinematics-only stimuli. The structure of a figure is defined in terms of the average positions of each of the 15 dots over one step cycle. The kinematics of a figure is defined in terms of the trajectory of each dot relative to its average position. Structure-only stimuli were intended to contain only structural differences and no differences in kinematics. Hence, structure-only stimuli were created by keeping the individual structural information of each figure and replacing its kinematics by the kinematics of the neutral figure (see Movie 6). Kinematics-only stimuli were meant to contain only kinematic differences and no differences in structure. So, kinematics-only stimuli were created by keeping the individual kinematics of each figure and replacing its structure by the structure of the neutral stimulus (see Movie 7). For more details on the creation of the stimuli see Troje (2002) and Troje, Westhoff, and Lavrov (2005). Each point-light walker was

orthographically and frontally presented in the centre of the screen (subtending a size of 8º of VA height), and each consisted of 15 black dots (radius = 15 arc mins of VA) on a grey background. Each trial started with a 500 ms fixation cross followed by the stimulus presentation consisting of 3 walking cycles (1.5 s per cycle), each consisting of 2 steps (128 frames per cycle at 85 Hz). Participants could only respond after each stimulus presentation with a key press (unlimited time). Pressing a response key initiated the next trial. The start position of the animation cycle was randomized across trials.

Procedure. All observers were tested individually. They were seated in a dimly lit and sound attenuated room in front of a 21 inch CRT monitor. Viewing distance was 57 cm. During instructions, participants were shown an extremely male (6 SD), extremely female (-6 SD), and neutral (0 SD) stimulus. They were instructed to indicate on each trial by a key-press whether the presented stimulus was perceived as oriented towards or away from them (arrow down for facing the viewer [FTV], arrow up for facing away [FA]). They were instructed to respond according to their own subjective experience and it was stressed that an equal distribution of both response alternatives was not necessary. When instructions were clear, participants completed 40 practice trials (random selection of all possible conditions). Then it was checked again whether the task was clear and the experiment commenced. Both the group of 22 males and the group of 22 females were randomly divided into two groups of 11 participants each. The first group judged the in-depth orientation of the 7 structure-only point-light figures in the first half of the experiment and the gender of the same stimuli in the second half. The second group did the same for the kinematics-only stimuli. Observers were only informed about what would be the task in the second phase (gender judgements) after they had completed the first phase (facing judgements). Within each phase the order of the trials was randomized. In total each participant completed 420 trials (7 genders * 2 phases * 30 repetitions), divided in 6 sessions of 35 trials in each phase. Between sessions participants could take a break. In total the experiment lasted about one hour.

RESULTS

In Figure 8A we plot the mean proportions of MALE responses across observers as a function of figure gender for the structure-only stimuli and the kinematics-only stimuli. From the plot it is clear that, as was demonstrated by Troje (2002), perceived gender of

frontally presented point-light figures is highly comparable for structure-only and kinematics-only stimuli.

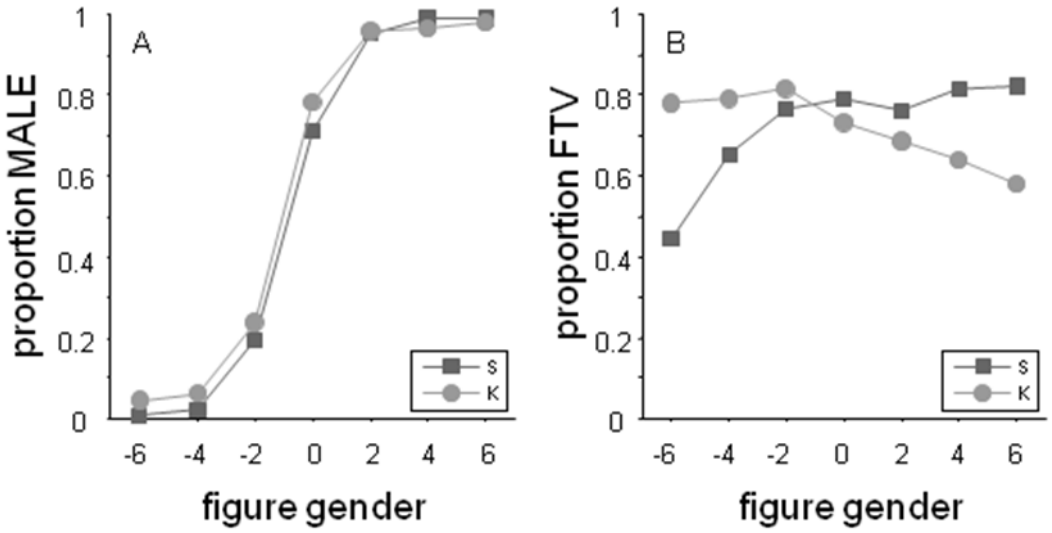


Figure 8. Proportions of MALE responses (A) and proportions of FTV responses (B) as a function of figure gender (from extremely female -6 SD, to extremely male 6 SD) for structure-only (S) and kinematics-only (K) point-light walkers.

The proportions of MALE responses as a function of figure gender were fitted with a cumulative Gaussian for structure-only and kinematics-only stimuli separately. Confidence intervals of the means and slopes were computed based on 10000 Monte Carlo simulations (see Wichmann & Hill, 2001a, 2001b). The slopes of the curves fitted to the proportions of MALE responses for the structure-only stimuli (*slope* = 0.27; $CI_{.95}$ = [0.25, 0.30]) and the kinematics-only stimuli (*slope* = 0.32; $CI_{.95}$ = [0.29, 0.36]) do not differ. There is, however, a slight but significant difference in the means of the structure-only (*mean* = -0.76; $CI_{.95}$ = [-0.86, -0.65]) and kinematics-only curve (*mean* = -1.00; $CI_{.95}$ = [-1.12, -0.89]). Both curves are shifted somewhat to the left, indicating a bias towards responding MALE more often than FEMALE. This so-called "male bias" has been observed in several studies before (Schouten et al., 2010; Troje et al., 2006; van der Zwan et al., 2009). Here, the male bias appears to be slightly stronger for kinematics-only stimuli. Taken together, we can conclude that changes in proportions of MALE responses as a function of figure gender are highly comparable for structure-only and kinematics-only point-light walkers.

Hence, if a change in perceived in-depth orientation were to be caused by a change in perceived gender, then the change in perceived in-depth orientation should be comparable for structure-only and kinematics-only stimuli. Inspection of Figure 8B, however, reveals an *opposite* relation between perceived gender and perceived in-depth orientation for structure-only stimuli compared to kinematics-only stimuli. Thus, stimulus manipulations that give rise to the same change in perceived gender give rise to an opposite change in perceived in-depth orientation.

To analyse the relation between figure gender and perceived in-depth orientation for structure-only and kinematics-only stimuli in more detail, we transformed for each observer the proportions of FTV responses for each figure gender to z-values (probit analysis). We then fitted for each observer a second-order polynomial to the probit transformed FTV responses as a function of figure gender, returning three meaningful parameters (constant, slope, and curvature) for each observer. This transformation allowed analyses of variance (ANOVA) and t-tests on the parameters across observers.

The constant parameter reflects the overall facing bias (across figure gender). A positive value indicates a facing bias. A negative value indicates a tendency to see the figure as facing away. Values near zero (FTV values around 50%) indicate that the figure is perceived as depth ambiguous. The slope parameter represents the slope with which FTV responses change as a function of figure gender. Positive values indicate that the number of FTV responses increases when the figure gender changes from extremely female to extremely male, and vice versa. The curvature parameter represents the curvature in FTV responses as a function of figure gender. Negative values indicate that the number of FTV responses for extremely female or extremely male gender figures tend to be smaller than for gender figures in between.

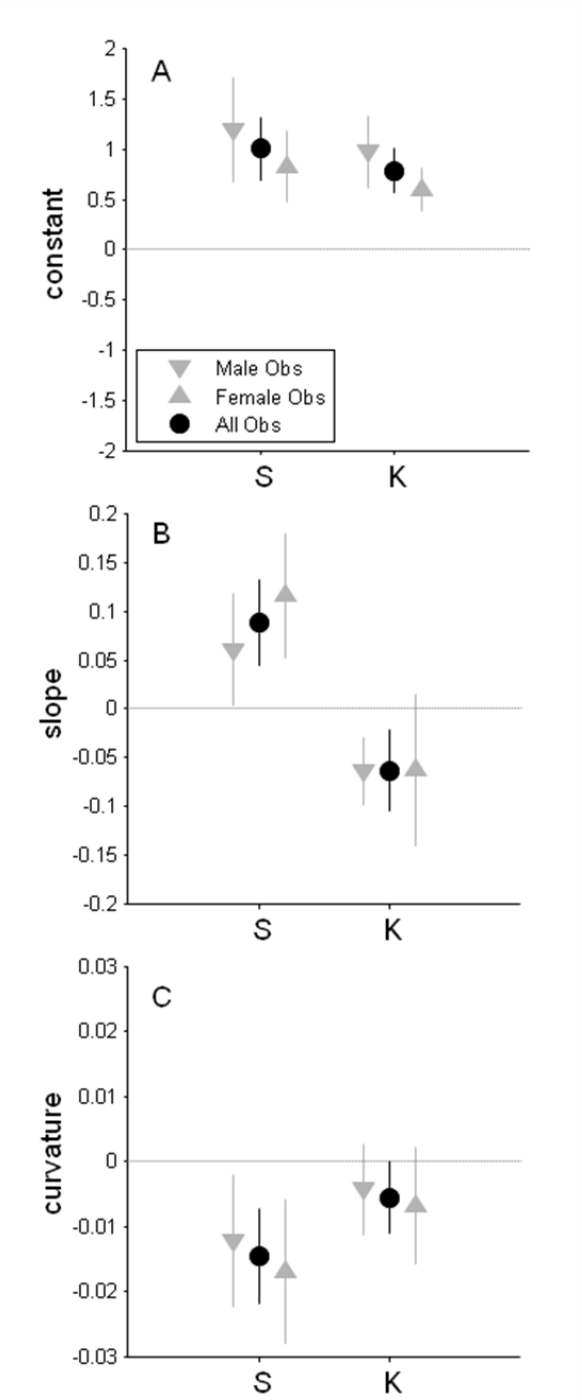


Figure 9. Mean parameter values and 95% confidence intervals of the second-order polynomial fit on the probit transformed FTV responses as a function of figure gender for structure-only (S) and kinematics-only (K) point-light walkers in Experiment 1. Black circles represent means and confidence intervals across all observers. Grey arrows pointing downwards/upwards represent the data of male and female observers, respectively.

In Figure 9 we plot the means and 95% confidence intervals across observers for each of the three parameters, including parameter values for male and female observers

separately. First, data for the constant variable indicate that, on average, the structure-only (*FTV* = 72.21%) and kinematics-only (*FTV* = 71.78%) stimuli elicit a comparable facing bias. Second and most importantly, Figure 9B confirms what was observed in Figure 8B. Structure-only stimuli elicit an opposite slope compared to kinematics-only stimuli. The slope of structure-only stimuli is significantly positive ($t(21) = 4.00$, $p < 0.001$) and the slope of kinematics-only stimuli is significantly negative ($t(21) = -2.97$, $p = 0.007$). Third, both types of stimuli elicit negative curvature, significant for the structure-only stimuli ($t(21) = -3.84$, $p < 0.001$), and a negative tendency for the kinematics-only stimuli ($t(21) = -1.90$, $p = 0.07$). Although the tendency is not significant, negative curvature is somewhat stronger for structure-only stimuli compared to kinematics-only stimuli ($t(21) = -2.0102$, $p = 0.06$).

From Figure 9 no strong differences between the parameter values of male and female observers are apparent. This was confirmed by a 2 (gender cue: structure-only vs. kinematics only) by 2 (observer sex: male vs. female observer) repeated measurement ANOVA on each of the three parameters separately. Only for the constant parameter a tendency of an effect of observer sex was found ($F(1,40) = 3.72$, $p = 0.06$): Male observers (FTV = 76.97%) tended to show a somewhat stronger facing bias than female observers (FTV = 67.02%). For none of the parameters, an interaction between observer sex and gender cue was observed.

DISCUSSION

As was expected based on the results of Troje (2002), the perceived gender of the fronto-parallel projected structure-only and kinematics-only point-light walkers was, apart from the small difference in the mean, found to be highly comparable. Hence, under the hypothesis of a causal relation between perceived gender and perceived in-depth orientation, changes in proportions of FTV responses as a function of figure gender for structure-only stimuli and kinematics-only stimuli should also be comparable. The results, however, show that the structure-only stimuli elicited an opposite relation between figure gender and perceived in-depth orientation compared to the kinematics-only stimuli. Structure-only stimuli on average induced a positive relation between perceived in-depth orientation and figure gender: The facing bias is much stronger for male figures and absent for female figures. Kinematics-only female stimuli, in contrast, elicited a facing bias and the more male the figure was, the weaker the facing bias was.

Contrary to what was observed in Schouten et al. (2010), here we did not find a significant interaction between figure gender and observer sex. Note, however, that the interaction found in Schouten et al. (2010) resulted from male observers exhibiting a stronger facing bias than female observers only for male figures. Although not significant, here male observers tend to show a higher constant than female observers indicating a higher facing bias across all gender figures for male observers than for female observers.

The crucial finding of Experiment 1 is that structural and kinematic variations between point-light figures that lead to comparable changes in perceived gender, lead to opposite changes in in-depth perceptions. The data thus indicate that the change in facing bias as a function of figure gender that was observed by Brooks et al. (2008; Chapter 3) and Schouten et al. (2010; Chapter 4) did not result directly from a change in the perceived gender. In contrast, stimulus properties that also lead to changes in perceived gender or that co-vary with other stimulus properties that lead to changes in perceived gender seem to have played an important role.

EXPERIMENT 2

INTRODUCTION

The results of Experiment 1 show a strong effect of structural and kinematic stimulus properties on the facing bias. Other stimulus manipulations that globally change the structure or kinematics of the point-light figure, such as perspective cues (Schouten & Verfaillie, 2010; Vanrie et al., 2004) or the type of action (Vanrie & Verfaillie, 2006) have also been demonstrated to affect the facing bias. However, the fact that stimulus manipulations that affect the entire figure also have an effect on the facing bias does not imply that the facing bias arises from information distributed over the entire figure. Some participants of the Schouten et al (2010; Chapter 4) study but also of the present Experiment 1 spontaneously reported that, to some degree, the perceived in-depth orientation seemed to depend on the part of the figure that they were attending to. When asked to further describe their impressions, these observers reported that when looking at the lower part they tended to perceive the figure as facing towards them in most of the cases, while when looking at the upper part the facing away interpretation occurred more often. These subjective reports suggest that the facing bias does not arise from information

that is distributed over the entire figure but rather from information that is present in the lower part of the point-light walker.

Vanrie and Verfaillie (2010) reported preliminary evidence that is in line with these subjective reports, albeit for a jumping action and not for a walking action. Indeed, whereas eye movement recordings of observers reporting switches in the perceived orientation of point-light *walkers* delivered somewhat inconsistent data, eye movement data of observers reporting switches in the in-depth interpretation of a *jumping* point-light figure suggested that fixations in the lower part of the figure resulted in relatively more FTV interpretations and fixations in the upper part of the figure in relatively more FA interpretations. This was not only the case when participants were instructed to freely saccade to a point-light figure presented in the periphery (internally determined initial fixations), but also when the initial fixation location was controlled experimentally by manipulating stimulus appearance relative to the fixation location (externally determined fixations).

In Experiment 2, we explicitly manipulated the available stimulus information, in order, first, to more directly test the hypothesis that the facing bias in the perception of point-light walkers depends on the figure part that observers attend to and, second, to examine whether this modulates the correlation between perceived gender and depth orientation. To this end, we included, besides the original point-light figures as used in Schouten et al. (2010), two extra conditions in which either only the upper parts or only the lower parts of the point-light walkers were shown. Proportions of FTV responses were recorded.

METHOD

Participants. Fourteen psychology students (7 males and 7 females) from the University of Leuven participated for course credit. All had normal or corrected to normal vision and were naïve to the aims of the experiment. None of them had participated in previous experiments investigating the facing bias.

Stimuli and apparatus. Apart from occluding either the upper or the lower part of the walker, stimuli were identical to the stimuli used in Schouten et al. (2010; Chapter 4). The gender stimuli thus differed in structural as well as in kinematic information. Two variables were manipulated: figure part and figure gender. There were three types of figure part information: Either only the upper part or only the lower part, or the whole figure was

visible. In the upper-part condition, only 8 dots representing the head, shoulders, sternum, elbows, and wrists were presented. In the lower-part condition, only 7 dots representing the pelvis, hips, knees, and ankles were visible. In the whole-figure condition, all 15 dots were shown. The spatial locations and motions of individual dots in the upper part and lower part condition were exactly the same as those of the dots representing the same body location in the whole figure condition. Similar to Schouten et al. (2010; Chapter 4) and Experiment 1, from the set of 13 gender stimuli (Troje 2002, Brooks et al. 2008) 7 were chosen so as to be equally spaced between extremely female and extremely male (z-scores: -6, -4, -2, 0, 2, 4, 6 SD). The factorial combination of figure gender and figure part resulted in 21 different point-light stimuli. Stimulus duration, start position, and randomisation method were similar to Experiment 1.

Procedure. The procedure differed from Experiment 1 on the following points. Before the start of the experiment, participants were shown a whole extreme male (6 SD), extreme female (-6 SD), and neutral (0 SD) stimulus and were informed that these kind of stimuli or only their upper part or lower part were going to be presented and that their task consisted of reporting whether the figure was seen as facing the viewer or as facing away (arrow down for FTV, arrow up for FA). When instructions were clear, participants completed 40 practice trials (random selection of all possible conditions). For each of the 21 stimulus combinations the 14 observers completed 20 repeats. Within each observer all 420 trials were randomized and divided in 7 sessions of 60 trials. Between sessions participants could take a break. In total the experiment lasted almost one hour.

RESULTS

As in Experiment 1, for each observer in each of the figure part conditions (whole, upper-only, lower-only) the probit transformed proportions of FTV responses (z-values) as a function of figure gender (z-values) were fitted with a second-order polynomial, returning a constant, slope, and curvature parameter.

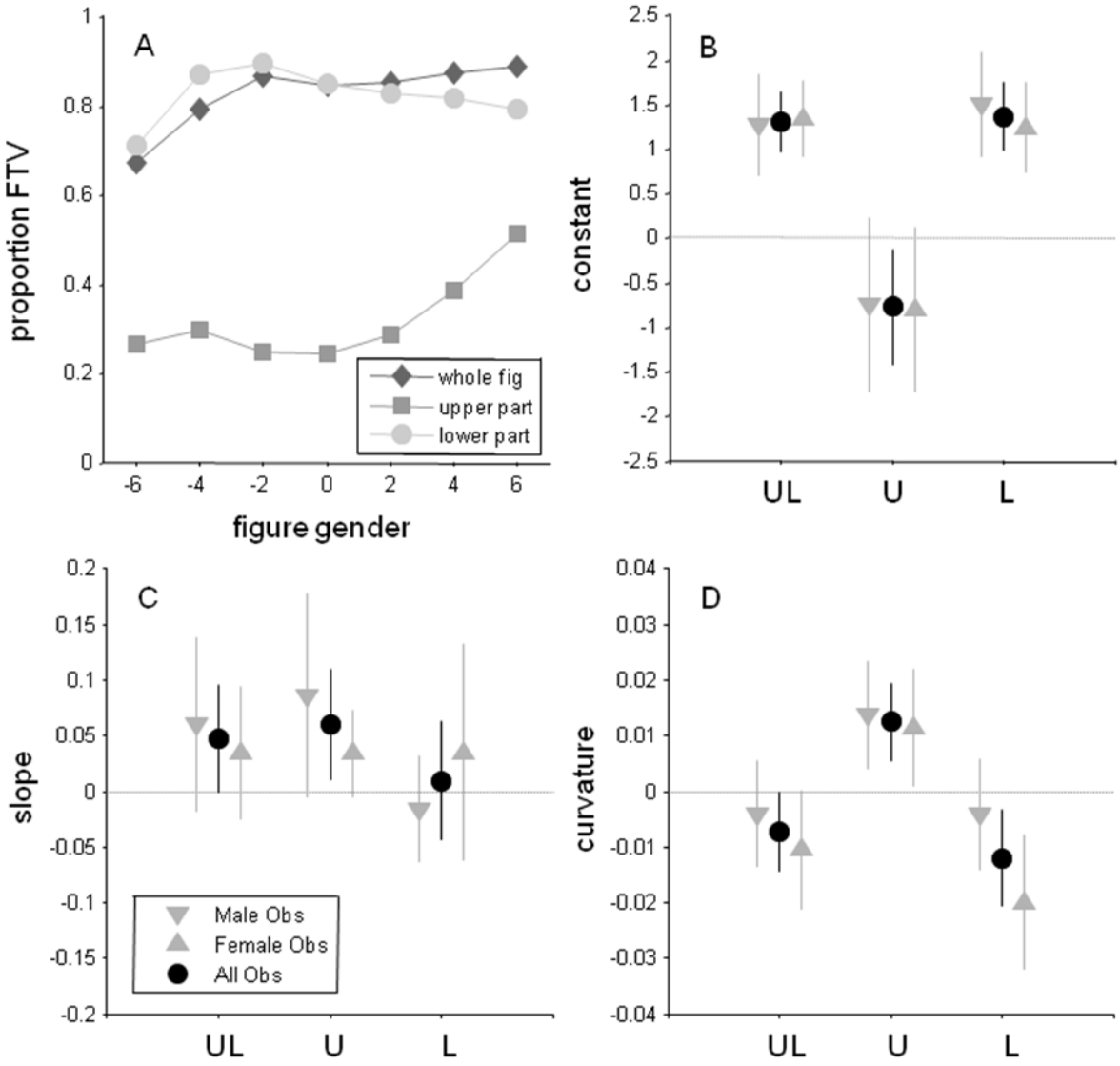


Figure 10. Data of Experiment 2 for whole figures (UL), upper-part-only figures (U) and lower-part-only figures (L). A: Mean proportions (across all observers) of FTV responses as a function of figure gender. B-D: Mean parameter values of the second-order polynomial fit on the probit transformed FTV responses as a function of figure gender. Black circles represent means and 95% confidence intervals across all observers. Grey arrows pointing downwards/upwards represent the data of male and female observers, respectively.

In panel A in Figure 10 we plot the mean proportions of FTV responses across observers for the whole figures, upper parts and lower parts as a function of figure gender. In panels B, C, and D in Figure 10 we plot the mean and 95% confidence intervals across observers and across male and female observers separately, for each of the three parameters in each condition. The crucial question concerned the dependency of the facing

bias on figure part. From panel A in Figure 3 it is clear that there is a strong effect of figure part on the overall proportions of FTV responses and thus also on the constant parameter (see panel B). Consistent with previous findings (Brooks et al., 2008; Schouten et al., 2010; Vanrie et al., 2004) a strong facing bias in the whole-figure condition was observed (82.81% FTV responses; mean z-value significantly positive; $t(13) = 7.48$, $p < 0.001$). A similar facing bias was found for the lower-part-only condition (82.40% FTV responses; mean z-value significantly positive; $t(13) = 7.07$, $p < 0.001$). In the upper-part-only condition, however, no facing bias is apparent. Instead, observers appeared to perceive the figure as facing away (31.89% FTV responses; mean z-value significantly negative; $t(13) = -2.37$, $p = 0.03$). Paired t-tests showed that the mean constant parameters for upper-part-only figures and lower-part-only figures strongly differed ($t(13) = -5.79$, $p < 0.001$) as did the mean constant parameters of the upper-part-only and whole figures ($t(13) = -7.10$, $p < 0.001$). The mean constant parameter for the lower-part-only and whole figures did not differ ($t(13) = 0.28$, $p = 0.79$). These results suggest that information in the lower part of the figure is responsible for the facing bias.

From panels A and C in Figure 10 no strong differences between the slopes are apparent. A one way repeated measurement ANOVA on the slope values revealed no effect of figure part ($F(2,26) = 1.31$, p = 0.29). However, based on previous research (Brooks et al., 2008; Schouten et al., 2010) a positive figure gender effect and thus a positive slope was expected in the whole figure condition. Here, in the whole figure condition, the mean slope was found to be significantly different from zero at the 0.05 level in a one-tailed t-test ($t(13) = 1.95$, $p = 0.04$; not significant in a two-tailed t-test). The mean slope for the upper-part-only condition for which we had no a priori expectations was also found to be significantly positive (two tailed t-test; $t(13) = 2.37$, $p = 0.04$). The mean slope in the lower-part-only condition did not differ from zero ($t(13) = 0.36$, $p = 0.73$).

Panels A and D in Figure 10 suggest a strong effect of figure part on the curvature parameter. This is confirmed by a significant effect of figure part in a one way repeated measurement ANOVA on the curvature values of the three part conditions ($F(2,26) = 19.78$, $p < 0.001$). Two-tailed t-tests reveal that there is a tendency of a negative curvature in the whole figure condition ($t(13) = -2.01$; $p = 0.07$) and a significant negative curvature in the lower-part-only condition ($t(13) = -2.73$; $p = 0.02$). The curvature in the upper-part-only condition is significantly positive ($t(13) = 3.57$; $p = 0.003$).

To test for observer sex effects we performed a 3 by 2 repeated measurement mixed ANOVA (figure part condition as a within-subjects variable and observer sex as a between-subjects variable) on the values of each of the three parameters. These analyses revealed no main effects of observer sex and no interactions of figure part and observer sex.

DISCUSSION

The results of Experiment 2 confirm what was reported by some participants in our previous experiments. When presented in isolation, the lower part of the point-light walker elicits a strong facing bias. The upper part, in contrast, on average seems to be perceived as facing away. The facing bias that is observed for the whole point-light walkers is quite similar to the facing bias for the lower parts. This suggest that the facing bias arises from stimulus properties in the lower part of the point-light walker. To establish and further explore the effects reported in Experiments 1 and 2, we designed a third experiment.

EXPERIMENT 3

INTRODUCTION

The first goal of Experiment 3 was to test the robustness of the effects found in Experiments 1 and 2. On the one hand, we aimed to replicate the positive relation between figure gender and facing bias for structure-only stimuli and the negative relation between figure gender and facing bias for kinematics-only stimuli. On the other hand, we intended to replicate that whole figures and their lower parts elicit a strong facing bias and that the upper parts are on average perceived as facing away.

The second goal of Experiment 3 was to explore a potential interaction between the variables that were manipulated in Experiments 1 and 2. It could be possible, for instance, that the structural or kinematic differences that induce changes in the perceived in-depth orientation only affect the perceived in-depth orientation of the upper or lower part of the figure. As such, the presence of an interaction between gender cue and figure part could provide more detailed information on the particular stimulus properties playing a role in the perceived in-depth orientation of point-light figures.

Therefore, in the current experiment we manipulated within subjects the gender of the point-light figure (7 levels: from extremely female to extremely male) as well as the gender cue (3 levels: original, structure-only, or kinematics-only point-light figures) and the figure part (3 levels: the whole point-light figure or only the upper or lower part). Observer sex was the between-subjects variable. FTV responses for all stimulus combinations were recorded.

METHOD

Participants. Twenty-two paid observers (11 males and 11 females) participated. All observers had normal or corrected to normal vision and were naïve to the purpose of the experiment. None of them had participated in previous experiments investigating the facing bias.

Stimuli and apparatus. All stimuli were derived from a combination of the stimulus manipulations that were described in the Method sections of Experiment 1 and 2. The combination of manipulations led to the following 9 sets of 7 point-light stimuli: SKUL, SUL, KUL, SKU, SU, KU, SKL, SL, KL in which S and/or K stand for the presence of structural and/or kinematic differences between the figures and U and/or L stand for the presence of the upper and/or lower part of the point-light figures.

Procedure. Procedural details that differ from Experiment 1 or 2 were the following. All variables (except for observer sex) were manipulated within subjects. Trials were randomized and divided in 21 sessions of 24 trials. Between sessions participants could take a break. Each participant completed 504 trials (63 stimuli * 8 repetitions) which lasted about an hour.

RESULTS

As in Experiment 1 and 2, for each observer for each of the conditions, the probit transformed proportions of FTV responses as a function of figure gender were fitted with a second-order polynomial, returning a constant, slope, and curvature parameter. In Figure 11 we plot the mean proportions of FTV responses across all observers as a function of figure gender (panels A-C) and the means and 95% confidence intervals across observers and across male and female observers separately for each of the three parameters in each condition. Data for point-light figures varying only in structural, kinematic, and both

structural and kinematic information are vertically aligned in the left, middle and right colum, respectively. For each parameter separately, a 3 (gender cue: SK, S, K) by 3 (figure part: UL, U, L) repeated measurement ANOVA was performed. Other ANOVAs and t-tests will be mentioned separately.

First, we checked whether we replicated the findings of Experiment 1 and 2. In panels A, B, G, and H in Figure 11 we can see that, as was observed in Experiment 1, the slopes of the number of FTV responses as a function of figure gender for structure-only (SUL) and kinematics-only figures (KUL) strongly differ. This is confirmed by a significant main effect of gender cue for the slope parameter ($F(2,168) = 28.33$, $p < 0.001$). Whole structure-only figures (SUL) elicit a positive slope ($t(21) = 3.39$, $p = 0.003$), whereas whole kinematics-only figures (KUL) elicit a negative slope ($t(21) = -3.61$, $p = 0.002$).

As was the case in Experiment 2, panels D, E, and F in Figure 11 suggest a strong effect of figure part for the constant parameter. This is confirmed by a strong main effect of figure part in the repeated measurement ANOVA on the constant parameters ($F(2,168) = 199.31$, $p < 0.001$). The original point-light figures (SKUL; 85.71% FTV responses; mean z-value significantly positive; $t(21) = 10.04$, $p < 0.001$) and their lower-part-only versions (SKL; 84.33% FTV responses; mean z-value significantly positive; $t(21) = 11.66$, $p < 0.001$) elicit a strong facing bias. The upper-part-only versions (SKU; 34.74% FTV responses; mean z-values significantly negative; $t(21) = -2.73$, $p = 0.01$) are on average again perceived as facing away. As was observed in Experiment 2, curvature for the lower-part-only figures generally seems to differ from curvature for the upper-part-only figures. This is reflected in a significant main effect of the curvature parameter for figure part ($F(2,168) = 22.61$, $p < 0.001$). In general, the current results thus confirm the basic findings of Experiment 1 and 2.

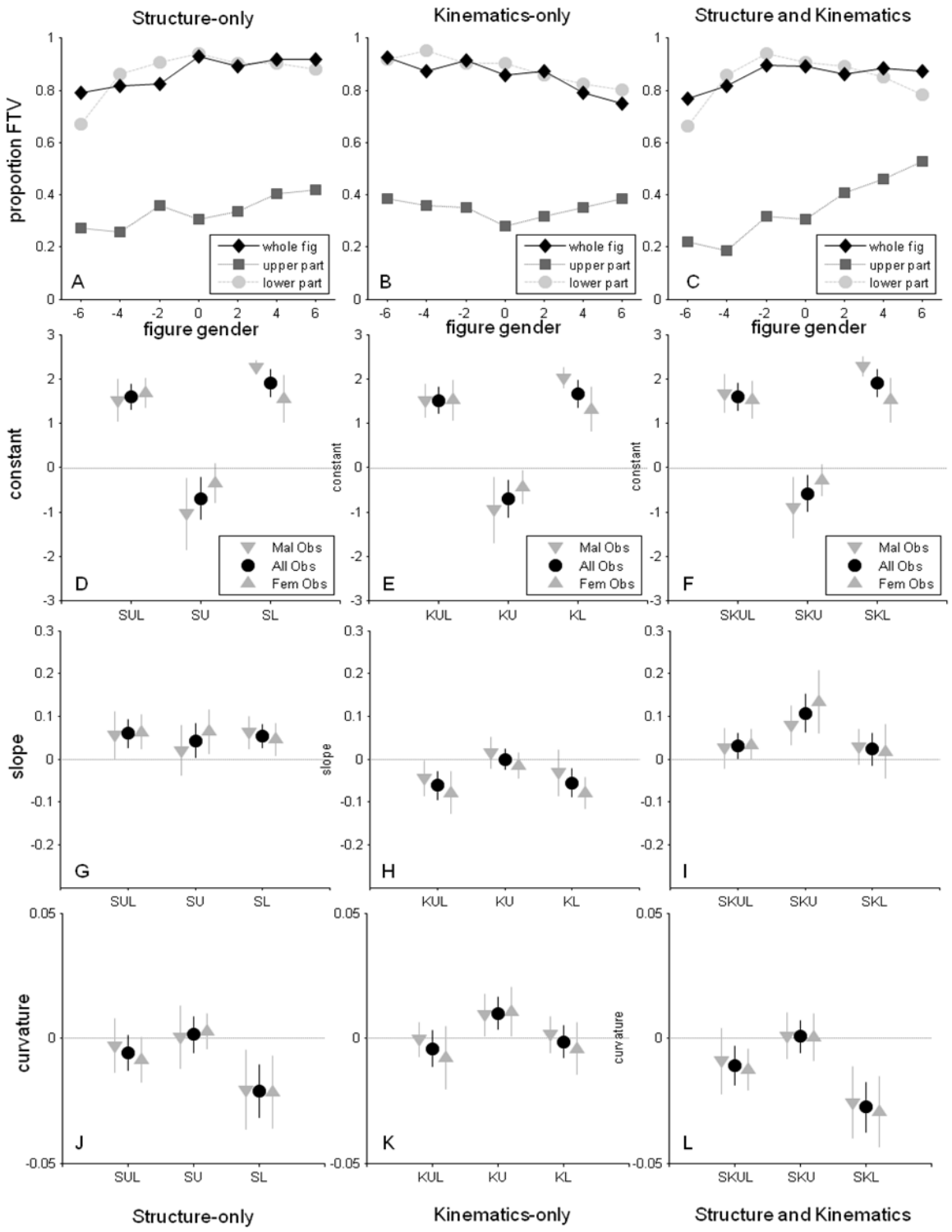


Figure 11. Data of Experiment 3. A-C: Mean proportions (across all observers) of FTV responses as a function of figure gender. D-L: Mean parameter values of the second-order polynomial fit on the probit transformed FTV responses as a function of figure gender for each of the conditions. Black circles represent means and 95% confidence intervals across all observers. Grey arrows pointing downwards/upwards represent the data of male and female observers, respectively.

Second, we explored the interaction between gender cue and figure part for the constant, slope, and curvature parameters. For the constant parameter we found no significant interaction between gender cue and figure part ($F(4,168) = 0.14$, $p = 0.97$). For the slope parameter, however, we observed a significant interaction ($F(4,168) = 2.81$, $p = 0.027$). Inspection of panel G in Figure 11 suggests that the positive figure gender effect that is found for whole structure-only figures (SUL) is present in the upper part (SU) as well as in the lower part (SL) of the figure. As confirmed by a one way repeated measurements ANOVA on figure part for the structure-only stimuli, the slopes of the whole structure-only (SUL), structure-upper-part-only (SU) and structure-lower-part-only (SL) figures do not differ ($F(2,42) = 0.31$, $p = 0.73$). A closer look at the slopes for the kinematics-only figures in panel H of Figure 11 suggests that kinematics-upper-part-only figures (KU) do not show a figure gender effect, whereas kinematics-lower-only figures (KL) show the same negative figure gender effect that is also observed for the whole kinematics-only figures (KUL). Paired t-tests on the slopes of the three figure part conditions for the kinematics-only stimuli showed that the slope for kinematics-upper-part-only figures differed significantly from the slope for whole kinematics-only figures ($t(21) = 2.97$, $p = 0.007$) as well as from the slope for kinematics-lower-part-only figures ($t(21) = 2.83$, $p = 0.01$). The slope of whole-kinematics-only figures and kinematics-lower-part-only figures, however, did not differ ($t(21) = -0.30$, $p = 0.77$). This suggests that the negative relation between perceived in-depth orientation and figure gender that is found for the kinematics gender cue acts mainly in the lower part of the figure. For completeness we mention the marginally significant interaction between gender cue and figure part for the curvature parameter ($F(4,168) = 2.39$, $p = 0.05$). Taken together, the current data thus reveal an interaction between gender cue and figure part.

At first sight, from panels D to L in Figure 11 no systematic parameter differences between male and female observers are apparent. However, closer inspection of panel D to F in Figure 11 suggests a difference for the constant parameter. For upper-part-only stimuli (SKU, SU, and KU) on average male observers seem to show smaller constant values when compared to female observers. For lower-part-only stimuli (SKL, SL, and KL), in contrast, the reverse appears to be true: Male observers systematically show higher constant values compared to female observers. These observations are confirmed by a 2 (upper-part vs. lower part) by 2 (male vs. female observers) ANOVA on the constant parameter. While there was no main effect of observer gender ($F(1,20) = 0.06$, $p = 0.81$)

we found, on top of the strong effect of figure part ($F(1,20) = 93.40$, $p < 0.001$) a significant interaction between figure part and observer sex ($F(1,20) = 7.04$, $p = 0.02$).

DISCUSSION

First, consistent with what was observed in Experiment 1, we found structure-only point-light figures to elicit a positive and kinematics-only figures to elicit a negative relation between figure gender and facing bias. Moreover, as was found in Experiment 2, in Experiment 3 whole figures and lower-part-only figures evoked a strong facing bias and upper-part-only stimuli on average evoked the tendency to perceive the point-light figure as facing away. We can thus conclude that the gender cue effect, observed in Experiment 1 and the figure part effect, observed in Experiment 2, were reliably replicated in Experiment 3.

Second, the results of Experiment 3 indicated that the gender cue effect (slope) depends on the part of the figure that is available. More specifically, the data indicated that the negative relation between figure gender and the facing bias that is observed for whole kinematics-only point-light figures, mainly results from information in the lower part. We also found a marginally significant interaction for the curvature parameter. Overall, the interaction between the gender cue effect and the figure part effect indicates that stimulus properties are affecting the perceived in-depth orientation differently in different parts of the figure.

Third, responses to stimuli that contain both structural and kinematic gender cues seem to resemble more the responses to the structure-only stimuli than the responses to the kinematic-only stimuli. This suggests that observers seem to rely more on the structural cues than on the kinematic cues when both are available.

Also worthy to note is that in Experiment 3 again significant differences in the FTV responses of male and female observers were found. Male observers show a stronger facing bias for whole and lower-part-only figures but they also show a stronger tendency to perceive the upper-part-only figure as facing away. It remains unclear, however, what exactly is causing this difference in response pattern between male and female observers.

GENERAL DISCUSSION

To test whether behavioral or social relevance is causing the facing bias for depth-ambiguous point-light walkers, Brooks et al. (2008) explored whether the gender of the point-light figure had an effect on the facing bias. A strong effect of figure gender on the facing bias was observed: The facing bias was only found for male and gender neutral point-light figures. Extreme female figures were perceived as facing away. In addition to a (weaker) figure gender effect, Schouten et al. (2010) found an observer sex effect: The facing bias for male point-light walkers was stronger for male observers than for female observers. While these observations are consistent with the account assuming that the perceived gender biases the perceived in-depth orientation of a point-light walker, they do not provide conclusive evidence. Indeed, crucial for the validity of such an account is the demonstration of a causal link between perceived gender and perceived in-depth orientation. The goal of the current study was to test whether or not this relation is causal.

The present study provided the following findings. First, the results of Experiment 1 demonstrated that the relation between perceived gender and perceived in-depth orientation is not causal. Structural and kinematic stimulus changes that induced comparable changes in perceived gender were found to elicit opposite changes in perceived in-depth orientation. While we observed a positive relation between figure gender and facing bias for structure-only stimuli, the sign of the relation inverted for kinematics-only stimuli. From this we can conclude that stimulus properties, irrespective of the perceived gender they elicit, play a crucial role in biasing the in-depth perception of depth-ambiguous point-light walkers. Second, the results of Experiment 2 indicated that the facing bias arises from information in the lower part of the point-light figure. Moreover, the data showed that the perceived in-depth orientation of whole figures is highly comparable to the perceived in-depth orientation of lower-part-only stimuli, but quite different from the perceived in-depth orientation of upper-part-only stimuli. This suggests that, to judge the in-depth orientation, observers preferably rely on the lower part of the point-light figure. Third, in Experiment 3 an interaction between the effects reported in Experiments 1 and 2 was demonstrated. This interaction was observed for the parameters that express the relation between perceived in-depth orientation and figure gender (significant for the slope and marginally significant for the curvature) suggesting that structural and/or kinematic properties differentallly affect the upper and lower parts of the stimulus. For example, the negative relation between figure gender and facing bias that was observed for kinematic-

only whole figures was also found for the kinematic-only lower parts but not for the kinematic-only upper parts.

Obviously, several questions remain. First, the results of the current study do not yet provide a clear insight into the causes of the facing bias. Our findings point to an important role for structural and kinematic information in the lower part of the point-light walker. When presented in isolation kinematic properties in the lower part of the female figures seem to elicit a strong facing bias. Kinematic properties in the lower part of male figures also elicit a facing bias but much weaker. Structural properties in the lower part of gender neutral figures give rise to a strong facing bias but the more male or female the figure is the weaker the facing bias is. The drop in facing bias is strongest at the female side of the continuum. More experiments are necessary to find out in detail which stimulus properties elicit the facing bias, but it nevertheless might be interesting to speculate. Recent studies (Chang & Troje, 2008, 2009a, 2009b; Saunders, Suchan, & Troje, 2009; Troje & Westhoff, 2006) showed that the local inversion effect in biological motion perception is carried by the dots representing the feet. Observers are able to correctly judge the left-right direction of a sagittal scrambled point-light figure. However, when the scrambled point-light display or only the trajectories of the foot dots are inverted, this ability disappears. These studies suggested the existence of a 'life detector': a visual filter that is tuned to quickly and automatically detect the presence of a moving living organism and direct attention to it. The mechanism is found to be specifically sensitive to the gravitational forces on the legs and to acceleration in the motion. The filter is believed to be evolutionarily old and innate and its main function would be to alert the observer to a potentially dangerous or otherwise demanding situation (Troje & Chang, in press). Because of the apparent similarities between the type of information that is necessary to elicit the facing bias and to trigger the life detector, namely the presence of particular kinematic information in the feet, here, we speculate that the facing bias might be somehow related to the life detector mechanism. For example, the vertical accelerations in trajectories of the feet that signal the potential presence of a moving organism might also trigger the perceptual system's tendency to err at the side of caution. Such a mechanism could provide a considerable evolutionary benefit, especially in situations in which the direction of motion of the feet is ambiguous. One interesting direction for future research on the in-depth perception of point-light walkers could thus be to investigate in detail the role of accelerations in the foot dots.

Second, it remains unclear why upper-part-only stimuli elicit a bias to perceive the figure as facing away. The positive relation between figure gender and facing bias for structure-only upper parts (Figure 11A,G) suggests that the perceived in-depth orientation of the upper part depends on the spatial relations between the dots. The presence of kinematic gender variations on top of structural variations (Figure 11C,I) in the upper parts seems to amplify this positive relation. Note that while across one step cycle the structure of kinematics-only gender figures remains constant, at comparable points within one step cycle, the structure of the kinematic-only figures can differ. As such, kinematic properties could amplify particular structural configurations when the gender figures vary in both structure and kinematics. For upper part figures that vary both in structure and kinematics, the bias to perceive the stimulus as facing away is strongest for upper parts of female figures but almost absent for upper parts of extreme male figures.

At first sight, the present results seem to be at odds with the results of Vanrie and Verfaillie (2006), who suggested that the upper limbs play the primary role in eliciting the facing bias. However, the point-light stimuli for which Vanrie and Verfaillie (2006) observed the facing bias, likely to be caused by information in the upper part, were quite different from the stimuli used in the present experiment. In their Experiment 3, Vanrie and Verfaillie (2006) explored the role of the direction of movement of the upper and lower limbs. Movement directions were either lateral or perpendicular to the plane of the torso. Observers had to judge the in-depth orientation of four semantically meaningless point-light stimuli. The point-light figures were either making a lateral or a perpendicular sway with the upper limbs and either making a lateral or perpendicular step with the lower limbs. Stimuli were presented in a frontal view and a ¾ view (rotation of 45 degrees). The results of this experiment showed that the direction of movement of the lower limbs had no effect on the proportions of FTV responses. The direction of movement of the upper limbs, however, had a strong effect. Point-light figures featuring a perpendicular arm sway elicited a strong facing bias whereas figures featuring a lateral arm sway did not. A crucial aspect of the results of Vanrie and Verfaillie (2006) that should be mentioned here is that the upper limb effect was only observed for the point-light figures in a ¾ view. One explanation put forward by the authors assumed that the facing bias arises for the upper part because the arms may function as a figure on the background of the torso. This line of reasoning by Vanrie and Verfaillie (2006) might not only provide an explanation of their results, it might also explain the difference between their results that the upper part seemed to elicit the facing bias and the present finding that the upper part of a point-light walker

elicits mainly 'facing away' interpretations. When the *wrist dots* move within the area delineated by dots making up the torso (shoulder and hip dots) some form of local, motion-induced figure-ground segregation may take place. This happens for the ¾ view figures of Vanrie and Verfaillie (2006) in which the upper limbs move perpendicularly. However, this does not occur in the frontal view stimuli of Vanrie and Verfaillie (2006) nor in any of their other stimuli in which the upper limbs swing laterally. Similarly, in the point-light walkers used in the present study, the wrist dots never enter the region of the torso. In contrast, in our stimuli *the elbow dots*, especially for female figures do enter the region of the torso. Contrary to the wrists, the elbows in a figure featuring a female structure tend to move *behind* the torso instead of in front of the torso and therefore lead to 'occlusion violations' (see Vanrie & Verfaillie, 2006). That is, if the figure was to be oriented towards the viewer, the elbow dots should in principle be occluded. In our stimuli the elbow dots are actually never occluded, potentially triggering the facing away interpretation for figures in which the elbow dots enter the region of the torso. This might also explain the positive relation between figure gender and perceived in-depth orientation that we observed for the upper part stimuli. The more male the figures are, the more the elbows point outwards. For extreme male figures the elbow dots never enter the region that is normally occupied by the torso.

Together, the present results and those of Vanrie and Verfaillie (2006) thus suggest that the perceived in-depth orientation of a point-light figure, or even a part of a point-light figure, strongly depends on its spatiotemporal characteristics. This means that findings on the perceived in-depth orientation of a particular point-light action cannot be generalized to other point-light actions.

Third, as revealed by the results of Experiment 3, there is also a positive figure gender effect for structure-only lower parts. This suggests, on top of the importance of the spatial relations between the dots of the upper part as discussed above, an important role for the spatial relations between the dots of the lower part. Inspection of the structural properties in the lower part that vary as the figures change along the gender dimension does, however, not suggest a comparable explanation as for the upper part. Instead, we speculate that the effect of structural information in the lower part might also be related to the life detector mechanism. The results of Chang and Troje (2009a) suggested that for retrieving the direction from scrambled biological motion displays the processing of the local spatiotemporal cues inherent to the isolated motion of one foot are not sufficient. Critically, for the life detector to be activated, the foot's elemental cues must be evaluated

with reference to the motions of other elements, or at least to other parts of the same element's trajectory. Hence, if the facing bias were to be driven by kinematic properties in the feet of the point-light walker, then the reference that one foot dot constitutes for the other might be crucial. The more female/male the point-light figure is the smaller/larger is the distance between the dots of the feet. Notably, the more into the female extreme of the gender continuum (-2, -4, -6 SD) the more progressively the dots of the feet overlap, thereby potentially disrupting the reference that is crucial to trigger the life detector. For neutral figures (0 SD) the strongest facing bias is observed, while when entering more into the male side of the gender continuum (2, 4, 6 SD) a saturation or even a small decrease in the strong facing bias for structure-only lower parts can be observed (Figure 11A), leading to a negative curvature (Figure 11J). This might reflect that neutral lower parts (0 SD) carry, amongst all gender figures, the optimal distance between the feet dots for the reference and that, also when the distance between the feet dots grows too big, the reference gradually gets disrupted. These hypotheses should of course also be tested in future research. This could be done, for example, by explicitly manipulating for one foot dot the presence of, and relative distances to, particular reference points.

Fourth, in Experiments 1 and 3 we showed that structural and kinematic differences between point-light figures have a strong effect on the perceived in-depth orientation. The data of Experiment 3 seem to suggest that, to judge the perceived in-depth orientation, observers rely more on the structural cues than on the kinematic cues. Neurophysiologically plausible models on the perception of biological motion (e.g. Giese & Poggio, 2003) generally incorporate two pathways: a form pathway and a motion pathway, that process the structural and kinematic properties of biological motion in a relatively independent manner before converging into common areas such as STS, premotor cortex and fusiform and occipital face area (also see Jackson & Blake, 2010). Recently, Vangeneugden, Pollick, and Vogels (2009) reported evidence for the existence of "snapshot" neurons - that code for form/posture - in the lower bank, and "motion" neurons - that code for visual kinematics - in the upper bank/fundus of the macaque STS. Future neurophysiological and behavioral research should investigate whether the dependency of the perceived in-depth orientation on the manipulation of structural and kinematic components in biological motion figures reflects the selective invasion of these specifically tuned neural structures.

Fifth, observer sex effects are repeatedly observed. However, the way in which they show up seems rather inconsistent. In Schouten et al. (2010; Chapter 4), male observers

showed a stronger facing bias than female observers for male but not for female point-light figures. In Experiment 1 of the present study, male observers generally showed a somewhat stronger facing bias than female observers. In Experiment 3 male observers showed a stronger facing bias than female observers for whole and lower-part-only figures but for upper-part-only figures male observers more strongly perceived the figure as facing away. Possibly due to the smaller sample size (7 females, 7 males) in Experiment 2 no observer sex effects were observed. Given the strong effects of figure part and gender cue, a possible explanation of observer sex effects could be that male observers attend to different stimulus parts or different stimulus properties than female observers (Johnson & Tassinary, 2005; Hewig, Trippe, Hecht, Straube, & Miltner, 2008). Why the effects are not always observed in a consistent form, remains unclear.

In summary, the current study provided two key findings concerning possible causes of the facing bias. First, we showed that the link between perceived gender and perceived in-depth orientation is not causal. Second, we demonstrated that the facing bias is entirely carried by information in the lower part of the point-light walker. Our data do thus not support accounts assuming that the facing bias is driven by the potential behavioral or social relevance inferred from the global percept of a human figure. They are, however, consistent with more locally acting stimulus properties affecting the perceived in-depth orientation of parts of the point-light walker. Yet, which particular stimulus properties consistently bias the perceptual in-depth organisation and whether these biases reflect innate mechanisms or learned priors remains unclear. Experimentally testing the hypotheses brought forward in this chapter could be the next step forward.

Chapter 6

The facing bias in biological motion perception: Some additional observations on observer sex differences, perspective, and response times

INTRODUCTION

In this closing chapter of Part I we took the opportunity to briefly explore in a single experiment three remaining issues on the in-depth perception of point-light figures. The three issues were the following.

First, in Chapter 4 we reported besides the figure gender effect also an interaction between figure gender and observer sex. While not consistent, in Chapter 5 again observer sex effects were observed. As the data of Chapter 5 indicate that the change in perceived in-depth orientation is not caused by a change in perceived gender, it would be interesting to test whether observer sex effects in the in-depth perception of point-light figures can also be observed for stimuli that do not vary in depicted gender, but vary, for example, in the amount of perspective information they carry.

Second, in Chapter 2 we presented a technique to manipulate the perceived in-depth orientation and to derive the point of subjective ambiguity. As mentioned in the discussion of that chapter, a limitation of our technique was that the distortions at the projection plane involve distortions of the shape, size, and position of the dots. The changes in the shape and size of the dots may act as local cues. It is thus possible, in principle, that in Chapter 2 observers merely based their responses on the shape and size changes of a single dot. Therefore, it would be interesting to check whether our perspective technique still works when all local shape and size cues are removed.

Third, in all our previous experiments on the in-depth perception of point-light figures we recorded proportions of FTV responses. We always assumed that the proportions of FTV responses reflected the perceived in-depth orientation of our stimuli. Particularly, proportions of FTV responses that considerably deviated from 0.50 were assumed to reflect perceptual biases. Yet, in the tasks used in the previous chapters it is unclear to what degree response biases, in addition to perceptual biases, were involved. Designing experiments that avoid this problem is one of the important challenges for future research on the in-depth perception of point-light figures (see Concluding remarks and future research). However, in the present experiment, using the familiar paradigm used in the previous chapters, we try to make a first attempt to develop additional measures of perceived in-depth orientation by instructing observers to respond as soon as they reach a stable percept and recording response times. We are aware of the fact that response times, as proportions of FTV responses, can be subject to criterion shifts. However, it could be interesting to explore how response times relate to proportions of FTV responses.

In summary, in the current experiment we tested whether observer sex effects can be observed for point-light figures of which the perceived in-depth orientation was manipulated with perspective cues. In contrast to Chapter 2, where the dots of the point-light figure changed in shape or size depending on the field of view angle, here we employed our perspective technique to compute the 2D positions of the dots of the point-light figure. Dots with a constant shape and size were then placed on these positions, leading to point-light figures in which only the relative locations and motions of the dots provided information on the figure's in-depth orientation. Proportions of FTV responses and reaction times were recorded.

EXPERIMENT

METHODS

Participants. 22 observers (11 females and 11 males) participated in this experiment. All observers had normal or corrected to normal vision and were naïve to the purpose of the experiment.

Stimuli and apparatus. The point-light stimulus used in this experiment was the neutral walker (0 SD) derived from the continuum described in detail by Troje (2002). The amount of perspective information that was provided by this stimulus was manipulated according to the perspective technique described in detail in Chapter 2 (Schouten & Verfaillie, 2010). However, one important aspect of this technique was adapted. Whereas in Chapter 2, the dots changed in shape and size depending on the field of view angle of the perspective projection, here first the 2D positions of the dots were computed. Then, dots of a constant shape and size were placed on these positions. This resulted in a set of point-light walkers of which only the relative positions and motions of the dots were informative on the point-light figure's in-depth orientation. As in Chapter 2, the perspective cues were gradually manipulated from absent (orthographic projection) to very strong. Note that here, as in Experiment 2 of Chapter 3 (Brooks et al., 2008) for half of the figures containing perspective cues, perspective was added to make the figures appear more facing the viewer (negative field of view angle values). In the other half of the figures perspective cues signalled a walker facing away (positive field of view angle values). In total, we used 13 levels of perspective information that were composed of one orthographic projection and 12 distance (D) manipulations of the convergence point, six on

the front side and six on the back side of the walker, corresponding to 16, 8, 6, 4, 3, and 2 times the height of the walker. This resulted in field of view angles of 14, 28, 37, 53, 67, and 90 degrees, respectively. As in Chapter 2, the height of the point-light figures (adjusted according to the viewing distance) subtended about seven degrees of visual angle. Hence, the height of the projected image (screen height, four times the height of the walker) subtended about 28 degrees. Each dot subtended about 15 arc mins. The start position of the animation cycle (117 frames/cycle) was randomized across trials.

Procedure. Observers were seated in a dimly lit and sound attenuated room. Viewing distance was 57cm from a CRT monitor (refresh rate = 120 Hz). On each trial, after a 500 ms fixation cross, observers had to indicate by a key-press (arrow down for FTV, arrow up for FA) whether the point-light figure was perceived as oriented towards or away from them. Reaction times were recorded. Observers were instructed to respond according to their own subjective experience and it was stressed that an equal distribution of both response alternatives was not necessary. The point-light walker was presented until response. Observers were encouraged to respond as fast as possible as soon as they experienced a stable in-depth percept. After instructions were given, observers completed a practice block (random selection of all possible conditions) in which they were familiarized with the stimuli and the task. Then it was checked again whether they had understood the task and the experiment commenced. In total, each observer completed 520 trials (13 levels * 40 repetitions). Trial order was randomized and trials were divided in 13 blocks of 40 trials each. In total, the experiment lasted about 35 minutes.

RESULTS

First, as in Chapter 2, we verified whether the proportions of FTV responses as a function of the field of view angle were well fit with a cumulative Gaussian. Deviance of the observations to the cumulative Gaussian fit (Wichmann & Hill, 2001a, 2001b) was not significantly higher than simulated deviance (10,000 Monte Carlo simulations) for 19 out of 22 observers (all p values >.05, except for three observers). Slopes varied between -0.021 and 0.0059 with a mean of -0.0073 (SD = 0.0064). Mean PSA values varied between -13.28° and 357.20° with a mean of 62.57°. Compared to what was observed in Chapter 2 (mean PSA = 40º), here the PSA seems to converge to a higher field of view angle. However, this value is strongly affected by outliers (some observers had a very strong facing bias leading to extremely high PSA values). When a single cumulative Gaussian

was fit to all data[7], the mean PSA converges to 47.30° which is more in the range of the mean PSA that we found in Chapter 2. Based on these data we can conclude that also without the distortions of the shape and size of the dots, the perspective technique works well to manipulate the perceived in-depth orientation and to determine the PSA.

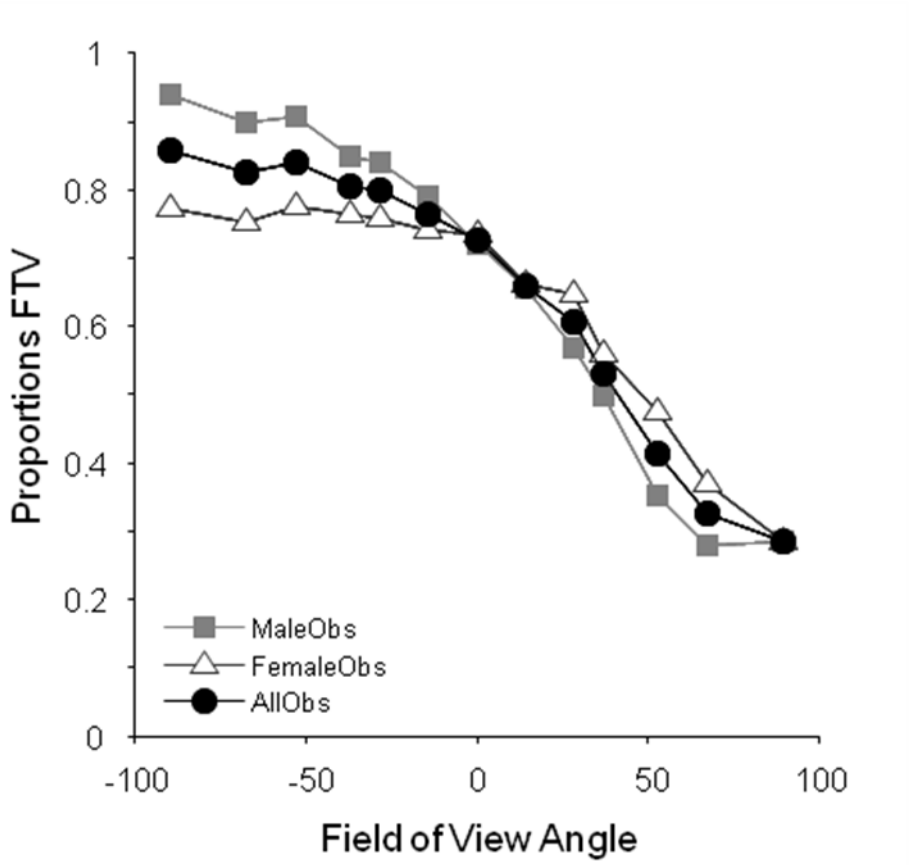


Figure 12. Proportions of FTV responses as a function of field of view angle for all observers and for male and female observers, seperately.

Second, to compare the present results with the results of Chapter 4 (Schouten et al., 2010) we performed a repeated measurement ANOVA on the probit transformed proportions with the field of view angle as a within variable and observer sex as a between variable. As expected based on what was described in the previous paragraph we found a significant effect of field of view angle on the perceived in-depth orientation ($F(20,240) = 30.18$, $p < 0.001$). Although observer sex itself did not show a main effect on the perceived in-depth orientation ($F(1,20) = 0.71$, $p = 0.41$), we found a significant interaction between field of view angle and observer sex ($F(12,240) = 2.44$, $p < 0.01$). In this respect, the

[7] In the fitting procedure of Wichmann and Hill (2001a, 2001b) values are weighted according to their reliability

present results are reminiscent of what was observed in Chapter 4. In Chapter 4, figure gender had a stronger effect on male observers than on female observers. Here, the perspective projection had a stronger effect on male observers than on female observers. In Figure 12 we plot the proportions of FTV responses as a function of field of view angle for male and female observers separately. From the figure it is clear that the changes in proportions of FTV responses as a function of field of view angle were stronger for male than for female observers. This was also confirmed by a significant difference in the slope parameters of two cumulative Gaussians, one fitted to the male (slope = -0.0068, 95% C.I. = [-0.0075, -0.0063]) and one fitted to the female observer data (slope = -0.0034, 95% C.I. = [-0.0038, -0.0030]). In addition, note that one of the female observers had a positive instead of a negative slope and that each of the three earlier mentioned observers for which the deviance exceeded the simulated deviance (10,000 Monte Carlo simulations) were female. Female observers thus respond quite differently to the same stimuli than male observers.

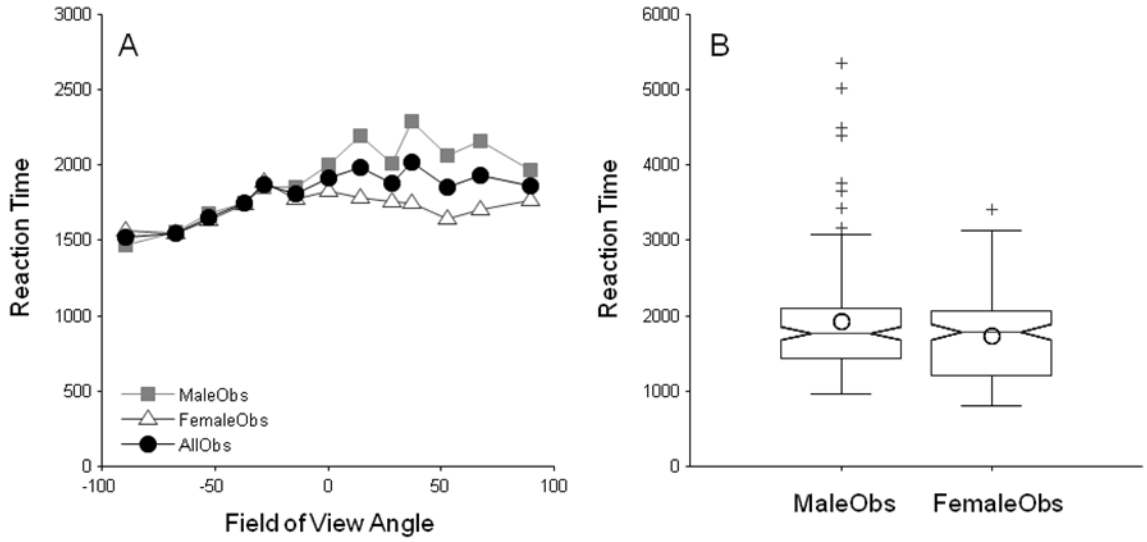


Figure 13. A. Reaction times as a function of field of view angle for all observers and for male and female observers separately. B. Box plots of reaction times for male and female observers.

Third, we explored response times and their relation to the proportions of FTV responses. To be able to visually capture potential observer sex differences in response times, in Figure 13A we plotted the reaction times as a function of field of view angle for male and female observers, separately. Mean response time of all responses was 1818 ms.

To analyse response times, first we fit for each observer a second-order polynomial to the response times as a function of field of view angle. We then analysed the values of the curvature, the slope and the constant with (paired) T-tests. From Figure 13A it appears as if female observers have shorter response times than male observers, mainly for figures containing 'facing away' perspective cues. However, the difference between the constants of female observers (1780 ms) and male observers (2012 ms) was not significant ($t(10) = 1.01$, $p = 0.34$). The box plots in Figure 13B clarify that the somewhat larger mean for male compared to female observers (circles) resulted from outliers (crosses) in the data of the male observers. The notches of the box plots (triangles pointing inwards) that indicate the 95% confidence interval around the median are overlapping, showing that median reaction times of male and female observers did not differ. Curvature ($t(10) = -0.58$, $p = 0.57$) and slope ($t(10) = 1.5235$, $p = 0.16$) parameters did also not differ significantly between female and male observers. However, response times as a function of field of view angle across all observers showed a significantly positive slope ($t(21) = 2.16$, $p = 0.043$) and a significantly negative curvature ($t(21) = -2.16$, $p = 0.043$). A repeated measurement ANOVA (with field of view angle as a within and observer sex as a between factor) on the response times showed that only the effect of field of view angle was significant ($F(12,240) = 3.71$, $p < 0.001$). The effect of observer sex ($F(1,20) = 0.64$, $p = 0.43$) and the (outlier driven) trend of the interaction of observer sex with the field of view angle ($F(12,240) = 1.69$, $p = 0.069$) were not significant.

Finally, we checked whether the response times correlated with the uncertainty reflected by FTV responses. It could be expected, for example, that the closer the proportion of FTV responses are to 0.50, the longer the response time is, because uncertainty about the perceived in-depth orientation of such point-light figures is highest. In contrast, the more the proportions of FTV responses differ from 0.50, the less uncertain an observer is about the perceived in-depth orientation. This was tested by probit transforming the proportions of FTV responses to z-scores. The probit link transforms proportions of FTV responses close to 0.50 to values near zero and proportions of FTV responses above and below 0.50 to positive and negative values, respectively. The absolute value of the z-scores could then be conceived as a measure of certainty about the response. If our expectations were right, these values should negatively correlate with response times. Across all conditions and observers the correlation between response times and the absolute value of z-scores of FTV responses was indeed significantly negative, $r(284) = -0.26$, $p < 0.001$.

DISCUSSION

The present experiment lead to the following observations. First, we show that our perspective technique also can be used when the shape and size distortions of the dots of the point-light figure are removed. This indicates that observers do not simply base their response on the local shape and size changes of a single dot. Instead, most observers seem to easily pick up on the perspective information that is provided by the relative positions and motions of the dots. While care should be taken when directly comparing the data of the current experiment and the data of Chapter 2 (e.g., different stimulus levels, different software, etc.), the slopes found in the present experiment seem to be somewhat flatter than the slopes we found in Chapter 2 suggesting that at least some of the perspective information that is provided by the shape and size changes of the dots was used by observers in Chapter 2. Some observers, apparently mainly females (see below), show a rather flat slope or a more atypical data pattern and seem to be worse in picking up the cue provided by the relative locations and motions of the dots.

Second, reaction times seem to capture to a certain degree the same uncertainty/certainty that is assumed to be reflected by proportions of FTV responses. The closer proportions of FTV responses are to 0.50, the slower response times and the further away proportions are from 0.50, the faster observers tend to respond. In general, as suggested by the significant positive slope of response times as a function of field of view angle, observers are faster in responding to figures that carry cues that the figure is facing towards the viewer and are slower in responding to figures that carry 'facing away' cues. In other words, observers more quickly reach a stable percept of a walker that is facing towards them than a walker that is facing away from them, an observation that is consistent with the facing bias. The significant curvature indicates that the increase in reaction time saturates or even reverses to a slight decrease for figures that carry strong 'facing away' cues. Hence, as could be expected, uncertainty about the in-depth orientation of those figures seems to drop again.

Third, we found an effect of observer sex on the perceived in-depth orientation. As in Chapter 4, we observed a significant difference between male and female observers in how strongly proportions of FTV responses depend on stimulus variations that cause variations in perceived in-depth orientation. This finding raises two questions that might inspire future research. First, is the resemblance between the present data and those of Chapter 4 due to the fact that changes in perspective information resemble somehow the

stimulus changes that occur when the gender of the point-light figure changes from female to male or from male to female? Or put differently, is the figure gender effect that we observed in the previous chapters caused by changes in the relative locations and motions of the dots that the perceptual system tends to interpret as perspective cues? Second, does the difference in slope between male and female observers that we found in the current study reflect a response bias or does it point to a higher sensitivity for the perspective cue for male observers compared to female observers? When looking at the literature, women seem to perform at a higher level than men on most verbal tests but men outperform women on visual-spatial tasks (Weiss, Kemmler, Deisenhammer, Fleischhacker, & Delazer, 2003). Gender differences in favor of men in spatial abilities are commonly found (Cooke-Simpson & Voyer, 2007; Voyer, Voyer, & Bryden, 1995). On the Mental Rotations Test (Vandenberg & Kuse, 1978) these sex differences in favor of male observers are most pronounced (Linn & Petersen, 1985). Hirnstein, Bayer, and Hausmann (2009) recently showed that sex-specific response strategies may contribute to the magnitude of sex differences in the original MRT. However, the authors also found that sex differences in response strategies do not fully account for the robust sex differences in mental rotation. In dynamic spatial tasks that test the ability to perceive and extrapolate real motion, to predict trajectories of moving dots, and to estimate arrival times of two or more objects, the employed strategy is also reported to play an important role in sex differences (Pena, Contreras, Shih, & Santacreu, 2008). However, what exactly the differences in strategy in these tasks are, remains unclear and vaguely described[8]. Apart from Akay and Celebi (2009), who found no observer sex differences using random dot stereograms and evoked potentials, as far as we know of, there is no research explicitly investigating or reporting on observer sex differences specifically related to depth perception.

While this cannot be derived from the present data, observer sex differences in in-depth perception performance could be easily tested in a 2AFC same-different task in which observers have to indicate whether two point-light figures, that are either the same or carry slightly different perspective cues, are the same or different. If male observers are perceptually more sensitive they should outperform female observers in such a task. If the

[8] Strategy selection in spatial problems is usually described as a continuum whose extremes are determined by 'holistic' and 'analytic' strategies. In holistic strategies (more typical for males) participants use information about the spatial relations between stimuli. When using analytic strategies (more typical for females), participants reduce spatial information to nonspatial formats, i.e., labeling parts of the spatial stimulus (Pena, Contreras, Shih, & Santacreu, 2008)

male and female performance turns out to be equal in such a task, this would suggest that male and female observers do not differ in sensitivity to perspective cues in point-light figures but rather show a different response behaviour in a naming task.

The fact that the observer gender effect occurs also for the perspective manipulation corroborates the findings of Chapter 5, demonstrating that the figure gender effect and its interaction with observer sex is not driven by a change in perceived gender. However, strictly one could still argue that the perspective manipulation affects the perceived gender and therefore changes the perceived in-depth orientation. In our opinion this explanation is quite unlikely because changes in perceived in-depth orientation for the present stimuli are much stronger compared to the changes in perceived in-depth orientation elicited by changes in figure gender. This would imply that changes in *perceived* gender are stronger for the present stimuli (which do not change in *depicted* gender) than for the well documented set of gender stimuli (Troje, 2002).

Part II

Audiovisual perception of biological motion

Chapter 7

Introduction to Part II

Traditionally, perception researchers have studied the senses in isolation. Actually, this is somewhat surprising because in daily life the senses are bombarded with countless simultaneous stimulations. Perception is thus, in essence, multisensory (Vroomen & Keetels, 2010), a point made already by Gibson (1966, 1979). Indeed, the multisensory nature of the world is highly advantageous because it increases perceptual reliability and saliency (Ernst & Bulthoff, 2004), and as a result, it can enhance learning and discrimination, and speed up our reaction (Hershenson, 1962; Morrell, 1968) to environmental stimuli (Keetels & Vroomen, in press). However, only since the past decade, multisensory perception gained scientific interest. Indeed, recently research on multisensory perception has made up arrears and at present the domain is truly mushrooming (for extensive overviews see Calvert, Spence, & Stein, 2004; Keetels & Vroomen, in press; Koelewijn, Bronkhorst & Theeuwes, 2010; Spence & Driver, 2002; Spence & Santangelo, 2009; Vroomen & Keetels, 2010; also see http://www.imrf.info/).

To introduce this relatively young research domain, in the first part of this introduction some of the often used concepts are discussed, immediately followed by a short overview of the most well known effects and phenomena. As the focus in the present dissertation is on the interaction between the auditory and the visual modality, we mainly restrict ourselves to literature on audiovisual interactions. In the second part of the introduction, we review the most relevant neurophysiological and behavioral research on the audiovisual perception of actions.

AUDIOVISUAL PERCEPTION

INTRODUCING SOME CONCEPTS

Multisensory integration, unity assumption, and the principle of inverse effectiveness

In the Encyclopedia of Neuroscience, Holmes, Calvert, and Spence (2009) define *multisensory integration* as "the neural integration or combination of information from different sensory modalities ... which give rise to changes in behavior associated with the perception of and reaction to those stimuli". Typically, information is integrated across sensory modalities when the sensory inputs share certain common features, or in other words, when information coming from different senses is congruent. This is often also referred to as the *unity assumption*. The assumption of unity states that as information from

different modalities share more (amodal) properties, the more likely the brain will treat them as originating from a common object or source (see e.g., Keetels & Vroomen, in press; Spence, 2007; Spence & Parise, 2010; Stein & Meredith, 1993; Vroomen & Keetels, 2010; Welch & Warren, 1980). Two main types of congruency are usually distinguished: *Temporal congruency* and *spatial congruency*. Temporal congruency points to the fact that multisensory integration is most pronounced when stimuli from different sensory modalities are presented within a specific temporal window. In the present dissertation, we will spend most attention to this type of congruency (also see below). Spatial congruency refers to the fact that stimuli that arise from the same external spatial location generally result in enhanced multisensory integration (for reviews see Spence & Driver, 2002; Spence & Santangelo, 2009; Koelewijn et al., 2010).

Perceptual enhancement of information in one modality that is congruent with information in another modality would be strongest at the behavioral level when the information in both modalities is quite poor, for example as a result of noise or when the interpretation is ambiguous. This is known as the principle of *inverse effectiveness*: The multisensory enhancement is inversely proportional to the effectiveness of the best unisensory response (Meredith & Stein, 1983). Recently, however, the rule of inverse effectiveness has become the subject of debate. Behavioural evidence sometimes contradicts the rule and direct links between multisensory neurons and multisensory behaviour remain unclear (see Holmes, 2007, 2009).

The point of subjective simultaneity (PSS) and the simultaneity window

To express the temporal congruency between auditory and visual stimuli two concepts or measures are of importance: The point of subjective simultaneity and the simultaneity window. Both measures strongly depend on the task in which they are assessed. Typically, perception of audiovisual simultaneity is measured in either a simultaneity judgement task (SJ) or in a temporal order judgement task (TOJ).

In an SJ task, audiovisual stimulus pairs are presented at various temporal offsets and observers explicitly judge whether the visual and auditory stimuli were simultaneous or not (see Figure 14). Proportions of simultaneous responses as a function of audiovisual offset are then fitted with a Gaussian curve. The peak of this curve is taken as the *point of subjective simultaneity (PSS)*, in that it is assumed that, at this particular temporal offset, the auditory and visual information is perceived as being maximally simultaneous

(Vroomen & Keetels, 2010). Another measure that is usually derived from the Gaussian curve is its standard deviation, or its width (in ms), which is taken as a measure of the *simultaneity window*, that is, the range of audiovisual offsets at which the brain treats the auditory and visual information as belonging to the same event. This range of audiovisual offsets is also referred to as the window of simultaneity, the temporal integration window, or the temporal window of integration.

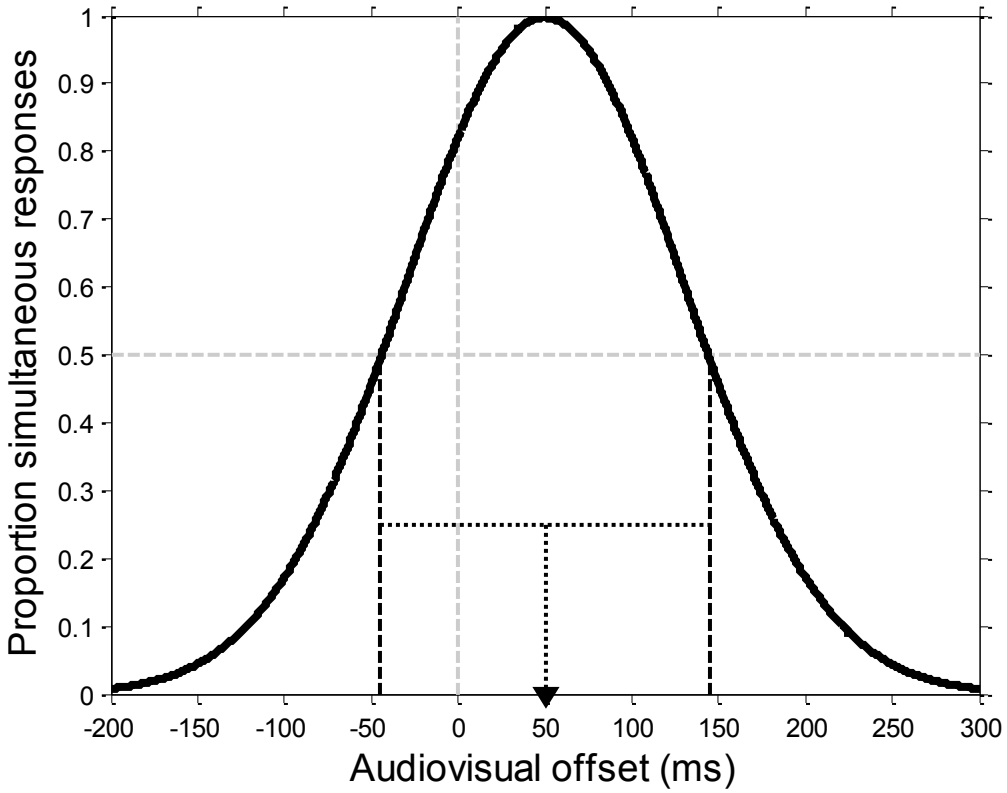


Figure 14. Schematic illustration of how the PSS and the simultaneity window can be derived from the data of an SJ task.

In a TOJ task, audiovisual stimuli are also presented at various temporal offsets but, rather than judging whether stimuli were simultaneous or not, observers judge which stimulus came first (or second). Usually, then proportions of 'audio first' or 'video first' responses as a function of audiovisual offset are fitted with a cumulative Gaussian curve (see Figure 15). The mean of this curve, that is, the 50% threshold, is taken as a measure of the PSS because it is assumed that at this point observers were maximally unsure about the temporal order of the auditory and visual stimuli. The slope of the curve or the distance between the 25% and the 75% threshold are then taken as a measure of the temporal

integration window (Spence & Squire, 2003; Van Eijk, 2008). Note that the latter measure is related to the *just noticeable difference* (JND), a well known concept in psychophysics.

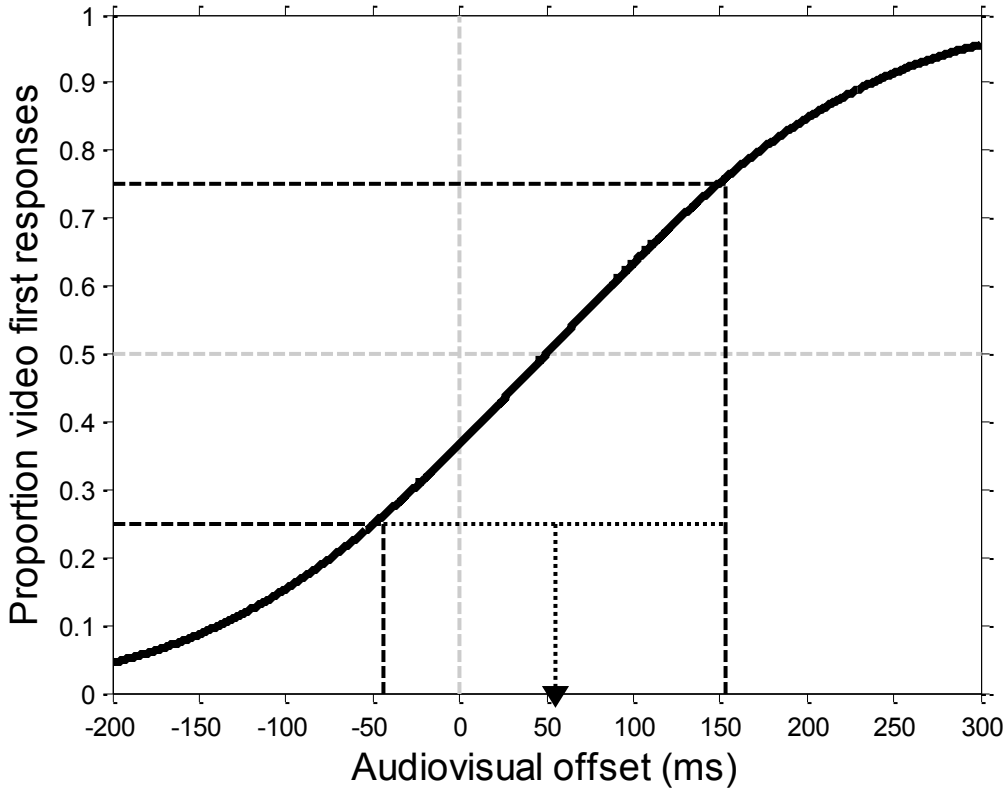


Figure 15. Schematic illustration of how the PSS and the simultaneity window can be derived from the data of a TOJ task.

As the SJ task and the TOJ task are meant to measure the same concepts, one would expect the PSS and simultaneity window measured in both tasks to highly correlate. Remarkably, this is not what is usually observed. Data of SJ and TOJ tasks usually correlate very poorly (van Eijk, Kohlrausch, Juola, & van de Par, 2008; Petrini, Holt, & Pollick 2010)[9]. For good overviews on audiovisual synchrony perception, see Keetels and Vroomen (in press), Vroomen and Keetels (2010), and van Eijk (2008).

[9] The general goal of a planned line of research was to investigate the factors that affect the PSS. In our opinion, the PSS as a measure of maximum perceived simultaneity is often misleading. Indeed, as already shown by Dixon and Spitz (1980), the distribution of simultaneous responses is typically positively skewed. Nevertheless, the PSS is usually defined as the mean of a symmetrical Gaussan fit to positively skewed data (Arrighi, Alais, & Burr, 2006; Petrini, Holt, & Pollick, 2010; Petrini, Russel, & Pollick, 2009), in our view, leading to an overestimation of the PSS. One potential way to overcome this problem is to explicitly deal with the asymmetry in the data by fitting two cumulative Gaussians, one at the rising and one at the descending side (see e.g., van Eijk et al., 2008) or to fit the data with a Gaussian including an extra free parameter that allows to capture the skewness (see e.g., Azzalini & DallaValle, 1996).

Audiovisual integration and the level of audiovisual interactions

While at present some consistency on the naming and definition of concepts starts to arise (Spence, Senkowski, & Röder, 2009), in the young literature on multisensory perception 'multisensory interaction' and 'multisensory integration' were often treated as synonyms. At present, it is clear that 'integration' points to a different process than 'interaction' (see below). Complicating the situation, inconsistencies did not only exist about 'integration' or 'interaction'. To refer to the involvement of multiple sensory modalities, 'multisensory', 'multimodal', 'intersensory', or 'crossmodal' were used as synonyms as well. In an attempt to simplify things, we will refer to 'audiovisual' instead of multisensory, crossmodal, or other kinds of naming when we aim to point to the interaction and integration between the auditory and visual modality.

Two types of findings are often taken as evidence for audiovisual *integration*: subjective reports about changes in perceptual experiences and changes in the precision of perceptual decisions. However, neither of them necessarily provides unambiguous evidence for integration (Arnold, Tear, Schindel, & Roseboom, 2010).

Subjective reports about changes in perceptual experiences are the weakest evidence for integration. For example, while changes in visual experiences through sound could be the result of audiovisual integration, these changes may just as well be the result of a decision-level interaction between modalities which does not necessitate integration. Subjective reports are thus not necessarily informative of the level of audiovisual interactions (also see Chapter 11).

Improvements in the precision of perceptual decisions (see e.g. Chapter 9 and Chapter 10) could, depending on the amount of improvement that is observed, be roughly related to one of the following three processes: probability summation, statistical optimal integration, or linear summation[10] (Arnold et al., 2010).

Probability summation can occur when redundant information is encoded by independent sensory systems. Each system has an independent probability of exceeding the requisite threshold for an accurate sensory decision. Summing both probabilities can lead to a slight improvement. Note however, that probability summation does not necessitate

[10] For example, take $AV_s = (A_s^k + V_s^k)^{1/k}$, where AV_s represents the sensitivity to combined auditory and visual signals, A_s represents sensitivity to auditory signals, and V_s represents sensitivity to visual signals. The exponent k then signifies the degree of summation. A probability summation corresponds with k = 3 to 4. Statistical optimal integration corresponds with a quadratic summation, or k = 2, and linear summation corresponds with k= 1 (Arnold et al., 2010).

sensory integration. Instead, it could be more appropriate to describe this process as 'interaction' rather than 'integration'.

Statistical optimal integration refers to the process in which a weighted summation of two initially independent sensory estimates is computed. When the weighting is inversely proportonial to the variance, associated with each of the initially independent sensory estimates, the process is often described as a Bayesian maximum likelihood estimation (MLE). Note that the weighting is inversely proportional to the variance of the unimodal estimate, because, the higher the reliability of the estimate, the smaller the variance, and vice versa. In other words, the weighting in statistical optimal integration is directly proportional to the reliability of the unimodal estimates (also see Ernst & Banks, 2002; Ernst & Bulthoff, 2004).

Linear summation is an even stronger type of integration than statistical optimal integration. It can occur when two sensory cues are encoded by a common physiological process. Sensitivity for the combined signals can then be predicted by simply adding sensitivities for the isolated signals.

Related to the discussion whether interactions between the sensory modalities can be called integration or not is the discussion about whether the interaction effects are situated at an early, perceptual, pre-attentive level or at a later, more cognitive, attentional level. Only a decade ago the majority of researchers studying visual perception were still highly sceptical of the existence of early sensory interactions. Sensory modalities were studied in isolation because it was assumed that only at a later level of processing the information of the different senses was combined (Schroeder & Foxe, 2005). This view gradually changed in the past years when a growing number of papers were published that employed, often new, behavioral (e.g., de Gelder & Bertelson, 2003; Vroomen & de Gelder, 2000; Vroomen & de Gelder, 2004b) and neurophysiological (e.g., Calvert & Thesen, 2004; Giard & Peronnet, 1999) experimental setups to provide evidence for early multisensory interactions.

Below, we will shortly discuss some of the most well known *audiovisual* interaction effects. Note that in the literature for several other modality combinations (vision-touch, audition-touch, etc.), interactions are reported that are comparable to the audiovisual phenomena presented here. The present overview is thus far from extensive and confined to the literature related to the focus of Part II.

AUDIOVISUAL INTERACTION EFFECTS

The McGurk effect

The McGurk effect (McGurk & MacDonald, 1976) is a perceptual phenomenon in which vision alters speech perception. For example, a sound of /ba/ tends to be perceived as /da/ when it is coupled with a visual lip movement associated with /ga/. The effect is robust and has been studied extensively. At the time of this writing, the paper of McGurk and McDonald (1976) has been cited 1184 times (source: ISI Web of Knowledge). For an illustration see Movie 8.

The ventriloquist effect

The ventriloquist effect (Howard & Templeton, 1966), or ventriloquism, refers to the phenomenon that the perceived location of a sound source is drastically influenced by visual stimulation. This effect is a common experience in daily life, such as when one is watching tv and voices are perceived to originate from the actors on the screen despite the considerable discrepancy between the image and the sound source (e.g., speakers on the side). The effect is called ventriloquism because it was considered a stripped-down version of what the ventriloquist was doing when performing on stage (Bertelson, & Aschersleben, 1998). Many experiments have demonstrated that the ventriloquist effect acts at a pre-attentive, perceptual level. For example, ventriloquism has been demonstrated to occur in patients with visual neglect (Bertelson, Pavani, Ladavas, Vroomen, & de Gelder, 2000), not to depend on the direction of deliberate visual attention (Vroomen & de Gelder, 2000), and to elicit aftereffects (Bertelson & Aschersleben, 1998). The ventriloquist effect thus seems to be a pre-attentive, perceptual audiovisual phenomenon (for reviews see Bertelson, 1999; Vroomen & de Gelder, 2004a).

Temporal ventriloquism

The perceived timing of a stimulus in one modality can be actively shifted toward the time of occurrence of an event in another modality. This phenomenon is named *temporal ventriloquism*, on the analogy of the spatial ventriloquist effect (Vroomen & de Gelder, 2004b; Vroomen & Keetels, 2010). The temporal ventriloquist effect is thus similar to the spatial variant, except that, in temporal ventriloquism a sound attracts a

visual event in the time dimension rather than vision attracting sound in the spatial dimension.

The remarkable aspect of this phenomenon is that sounds can increase the visual temporal resolution. That is, when the temporal order of two successive visual flashes, e.g., one at the left and one at the right, has to be detected ("which was first?"), then a non-spatial sound before the first flash and a non-spatial sound after the second flash will increase visual performance because the sound can separate the visual events in time by attracting them. There are by now many demonstrations of this phenomenon (for some examples, see Aschersleben & Bertelson, 2003; Bertelson & Aschersleben, 2003; Fendrich & Corballis, 2001; Freeman & Driver, 2008; Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Vroomen & de Gelder, 2004b). These studies show that small lags between vision and sound are reduced and thus may go unnoticed because the perceived timing of the visual events is flexible and adjusts immediately to the timing of the sound.

The stream - bounce effect

Two identical visual targets moving across each other can be perceived either to bounce off or to stream through each other (Metzger, 1934). Nonetheless, the majority of observers report a perception of streaming instead of bouncing. Sekuler, Sekuler, and Lau (1997) demonstrated that when a brief sound is presented at the moment at which the targets coincide, the visual perception is strongly biased to the bouncing interpretation (for an illustration see http://www.michaelbach.de/ot/mot_bounce/index.html).

The auditory induced bounce perception persists as the probability of a motion reversal is reduced (Grove & Sakurai, 2009) and when the sound is subliminal (Dufour, Touzalin, Moessinger, Brochard, & Despres, 2007) suggesting that the activation of the human auditory system, even without reaching consciousness, can affect the perception of an ambiguous visual motion display. Watanabe and Shimojo (2001) demonstrated that flanker sounds presented 300 ms before and 300 ms after the simultaneous sound that normally induces the bouncing percept, attenuate the effect. However, the attenuation only occurs when the simultaneous sound and auditory flankers have similar acoustic characteristics and the simultaneous sound is not salient. Thus, when the auditory stimuli are similar enough, so that they are perceptually grouped together, auditory grouping seems to win from audiovisual grouping.

Note that earlier O'Leary and Rhodes (1984) already demonstrated that the perceptual organization of tones could influence the perceptual organization of moving dots. They found that when a sequence of high and low tones was heard as two streams, a dot that moved up and down was more likely to be seen as two streams of dots moving horizontally.

The freezing effect

When observers are shown a rapidly changing visual display, an abrupt sound can "freeze" the display with which the sound is synchronized. Perceptually, the display appears brighter (also see Stein, London, Wilkinson, & Price, 1996) or appears to be presented for a longer period of time (Vroomen & de Gelder, 2000, 2004a).

In a series of six experiments, Vroomen and de Gelder (2000) showed that the freezing phenomenon occurs at an early, perceptual level. Whether or not the freezing phenomenon was observed was closely related to the perceptual organization of the tone in the auditory modality. More specifically, the effect was observed when the tone could easily segregate from a sequence of tones, but it was greatly attenuated (or it even disappeared) when exactly the same tone was less abrupt or part of a melody (note the analogy to the effect of flanker sounds on the stream – bounce effect). The phenomenon is unlikely to be accounted for by crossmodal attentional cuing (Spence & Driver, 1997), because the effect disappeared when the abrupt sound preceded the target by an SOA at which one may expect a crossmodal attentional cuing effect (i.e., 254 ms), and it was unaffected by whether the onset of the target was predictable or not. The freezing phenomenon is thus also an early, perceptual effect (Vroomen & de Gelder, 2004a).

The illusory flash effect

In the illusory flash effect, sound also radically changes the phenomenological quality of the percept of a non-ambiguous visual stimulus. When a single brief visual flash is accompanied by multiple auditory beeps, the single flash is perceived as multiple flashes (Shams, Kamitani, & Shimojo, 2000). Control conditions, catch trials, and many other observations indicate that the illusory flashing phenomenon is indeed a perceptual illusion and is not due to the difficulty of the task or to some cognitive bias caused by sound (Shams, Kamitani, & Shimojo, 2000, 2002, 2004). The effect was also found to be resistant to feedback training (Rosenthal, Shimojo, & Shams, 2009).

Using Visual Evoked Potentials (VEP), Shams, Kamitani, Thompson, and Shimojo (2001) provided neural evidence for the illusory flash effect. VEP's were compared in the presence and absence of sound. Activity was modulated extensively and with short latency in trials in which an illusory flash was perceived. Moreover, brain potentials for the illusory flash were qualitatively very similar to those for a physical flash suggesting that the same mechanism underlies the percept of both illusory and physical flashes (also see Watkins, Shams, Josephs, & Rees, 2007; Watkins, Shams, Tanaka, Haynes, & Rees, 2006). The illusory flash effect is sometimes also referred to as the 'Shams illusion' (for demos see http://shamslab.psych.ucla.edu/demos/).

The pip pop effect

Van der Burg, Olivers, Bronkhorst, and Theeuwes (2008) recently showed that a simple auditory pip drastically decreased search times for a synchronized visual object that is normally very difficult to find. In a visual search task observers had to find a horizontal or vertical line segment of which the color repeatedly and abruptly changed between red and green. The target had to be found between a number of similar red-green alternating distractor line segments with another orientation. Van der Burg et al. (2008) found that an abrupt and non-spatial auditory pulse synchronised with the abrupt color change of the visual target strongly facilitated the speed of target identification, and did so independently of the number of distractor elements. Phenomenologically, the synchronous pip made the visual object pop out from its complex environment, providing a direct demonstration of spatially nonspecific sounds affecting competition in spatial visual processing.

The pip pop effect differs from the freezing phenomenon in that objects, as they do in the freezing phenomenon (Vroomen & de Gelder, 2000), do not appear serially at the same spatial location. The pip pop effect thus shows how a sound, not carrying any relevant spatial information, can still strongly affect the competition between multiple objects concurrently present in a spatial layout, simply because of its temporal congruency with the visual target.

Actually, the demonstration of the pip pop effect came somewhat as a surprise. Indeed, some years before, Fujisaki, Koene, Arnold, Johnston, and Nishida (2006) had reported no enhancement in a visual search task for a target changing in synchrony with an auditory signal. In their study, participants were asked to detect a flashing or rotating visual target among a number of flashing or rotating distractor objects. The target dynamics were

synchronized with either amplitude-modulated pips or frequency-modulated sweeps. In contrast to what was recently found by Van der Burg et al. (2008), the presence of the sounds did not result in more efficient search. According to Van der Burg et al. (2008), one of the reasons for this discrepancy in results might be that in Fujisaki et al.'s displays, the sound, the distractors, and the target, were changing continuously. Combined with the high range of modulation frequencies they used (up to 40 Hz), such circumstances may not have been optimal for spatiotemporal audiovisual integration. Indeed, Fujisaki and Nishida (2005) themselves showed that multisensory integration becomes difficult at temporal frequencies higher than 4 Hz (the repetition frequency of the sounds used in Van der Burg et al., 2008 was about 1.11Hz). Moreover, to find the unique audiovisual coupling in their displays, the auditory and visual streams needed to be integrated across rather lengthy intervals (up to 2 s), whereas in Van der Burg et al.'s paradigm, only single synchronized audiovisual events occurred, which were temporally isolated.

Very recently, Van der Burg, Cass, Olivers, Theeuwes, and Alais (2010) reported a set of experiments to further investigate the discrepancy between their own results and those of Fujisaki et al. (2006). By manipulating the kind of temporal modulation (sine wave vs. square wave vs. difference wave) they showed that abrupt visual events are required for the audiovisual search efficiency (pip pop effect) to occur. Sinusoidal audiovisual modulations, as were used in Fujisaki et al. (2006), did not support efficient search. What seems to be crucial to observe the effect is that the synchronized audiovisual stimuli are transient. Van der Burg et al. (2010) propose that transient signals are necessary in synchrony driven binding to avoid spurious interactions with unrelated signals when these occur close together in time. Also see the General Discussion of Chapter 9. For a demo of the pip pop effect see http://www.psy.vu.nl/pippop/.

Neural correlates of audiovisual integration

During the past years, many efforts have been done to find neurophysiological evidence for multisensory interactions (Calvert & Thesen, 2004; Giard & Peronnet, 1999). Here we will summarize the most relevant findings (also see Koelewijn et al., 2010).

In the heteromodal regions of the brain three main neural structures were identified that are involved in multisensory interactions: The ventral and lateral intraparietal areas (VIP and LIP), the Superior Colliculus (SC), and the Superior Temporal Sulcus (STS) of which the latter two (SC and STS) are most often mentioned in neurophysiological studies

on audiovisual integration. Figure 16 (with thanks to Thomas Koelewijn) provides a schematic anatomic overview of these areas.

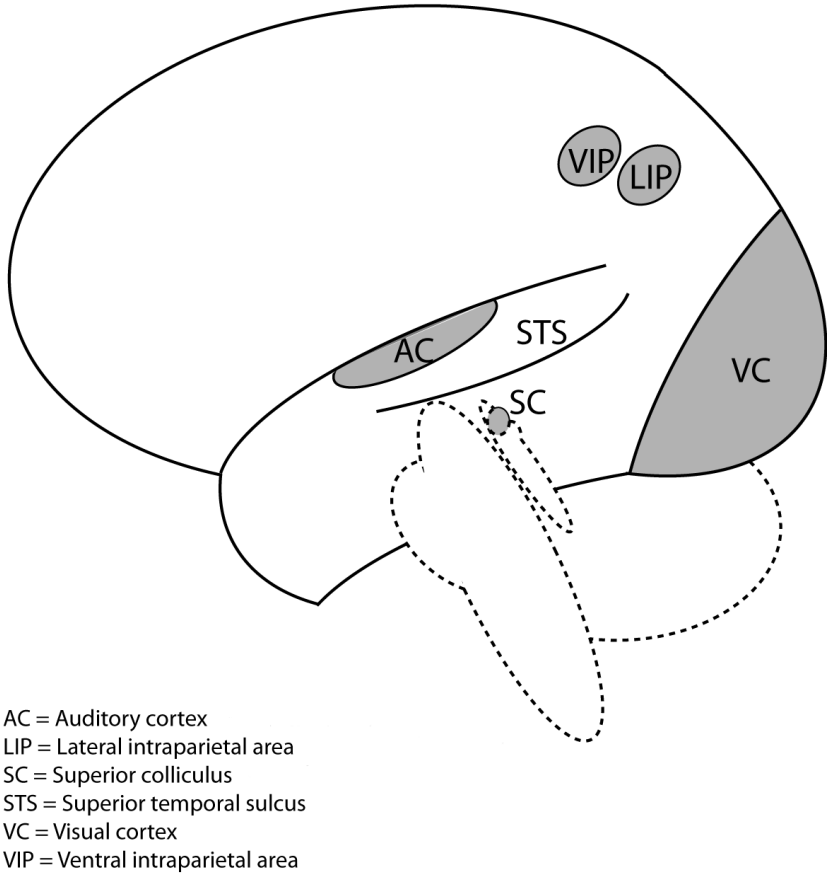


Figure 16. Brain areas involved in audiovisual integration. (Image reproduced with kind permission from Koelewijn et al., 2010).

The SC is a subcortical structure that is known to be involved in saccadic eye movements. It receives input from the visual cortex and many other cortical areas and has a strong topographic organization. Neurons within the SC not only respond to visual but also to somatosensory and auditory input (Meredith, Nemitz, & Stein, 1987; Meredith & Stein, 1986). The receptive fields of these different modalities overlap. For example, a sound or visual event presented at the same location will activate the same SC neuron (Meredith & Stein, 1996). Bimodal stimulation within the same receptive field will result in a super-additive neuronal response (Wallace, Meredith, & Stein, 1998). However, not only spatial but also temporal proximity of multisensory input results in stronger neural activity (Calvert, Brammer, Iversen, 1998; Calvert, Hansen, Iversen, Brammer, 2000; 2001;

Colonius & Diederich, 2004; Diederich & Colonius, 2004; Meredith & Stein, 1983; Meredith et al., 1987).

Quite a number of studies found the STS to be involved in audiovisual integration (see Hein & Knight, 2008). Using functional magnetic resonance imaging (fMRI), Calvert, Campbell, and Brammer (2000) investigated whether neural correlates could be found for the observation that congruent multisensory stimuli lead to better detection and shorter reaction times than unisensory stimuli (Hershenson, 1962; Morrell, 1968). Ten human subjects were exposed to varying epochs of semantically congruent and incongruent audiovisual speech and to each modality in isolation. Brain activations to matched and mismatched audio-visual inputs were contrasted with the combined response to both unimodal conditions. This strategy identified an area of heteromodal cortex in the left STS that exhibited significant supra-additive response enhancement to matched audiovisual speech inputs and a corresponding subadditive response to mismatched audiovisual speech inputs (also see Calvert et al., 1997). Calvert, Hansen, Iversen, and Brammer (2000, 2001) showed that this effect also extends to non-speech stimuli such as white noise pulses paired with black and white checkerboard alternations. Barraclough, Xiao, Baker, Oram, and Perret (2005) later showed that matching sights and sounds of actions integrate in STS (also see furtheron).

Note that not only heteromodal areas but also unimodal areas, as for example the primary visual or auditory cortices, that were long believed to receive strictly unisensory input, respond to input from other modalities. We already mentioned the study of Shams et al. (2001), showing that visual evoked potentials can be elicited by a sound leading to an illusory visual flash. Using fMRI, Calvert et al. (1997) found that even in the absence of auditory speech sounds, the visual perception of the movements of a speakers lips activates the auditory cortex. Calvert et al. (1999) found significant response enhancements (specific to semantically congruent audiovisual inputs) in both auditory (BA 41/42) and visual (V5) cortex under audiovisual speech stimulation. See Koelewijn et al. (2010) for more examples.

In summary, during the past years research on multisensory perception in general and on audiovisual perception in particular has blossomed. While a decade ago the debate about the possibility of early sensory interactions was highly vivid and sometimes heated, the results of numerous behavioral and neurophysiological studies have gradually shifted opinions. It is now commonly believed that multisensory interactions occur at various processing levels, ranging from very early, perceptual, or pre-attentive integration to late,

cognitive, attentional interactions. Whether or not multisensory interaction occurs, seems to be highly dependent on whether the *assumption of unity* is fulfilled, that is, whether the stimuli invading the different senses are spatially, temporally or even semantically congruent. The type of stimuli or their complexity may be relevant to the level at which the effects occur (Spence & Parise, 2010; van Eijk et al., 2008; Stekelenburg & Vroomen, 2007).

While audiovisual speech stimuli often lead to strong interactions, it is yet unclear whether speech is a special case of audiovisual interaction (Calvert et al., 2004; Spence & Parise, 2010; Stekelenburg & Vroomen, 2007; Tuomainen, Andersen, Tiippana, & Sams, 2005). Recently, a subgroup of researchers has started to focus on ecologically valid audiovisual stimulation of another type, namely the audiovisual perception of non-speech actions. This type of research will be the topic of Part II.

AUDIOVISUAL PERCEPTION OF BIOLOGICAL MOTION[11]

As may appear from Chapter 1, the focus in the research on biological motion perception traditionally was restricted to the visual modality. However, humans live in a multisensory world, and sounds are often available during action observation. In fact, it is hard to imagine everyday actions that do not produce any sound. The most obvious audiovisual action that is involved in many daily activities is probably speech. However, apart from speech, numerous other actions provide observers with synchronized audiovisual stimulation, as in the case of walking, clapping, hammering, putting down a cup, etc. The perceptual system seemingly without any effort integrates this multisensory stream of information. Therefore, it is sometimes hypothesized that actions are not represented merely visually but rather audiovisually.

Recent neurophysiological research supports this idea. First, audiovisual mirror neurons have been localized in the ventral premotor cortex of monkeys, neurons that not only fire when the monkey executes a particular action or sees the action, but also when the monkey hears the action related sound. Audiovisual mirror neurons are believed to represent actions independently of whether these actions are performed, heard, or seen (Keysers, Kohler, Umilta, Nanetti, Fogassi, & Gallese, 2003; Kohler, Keysers, Umilta, Fogassi, Gallese, & Rizzolatti, 2002). Second, using fMRI in humans, Kaplan and Iacoboni

[11] Note that some of the papers that we refer to in this introduction were not yet published at the time we were running the experiments that are reported in the empirical chapters of Part II (papers that appeared after 2005).

(2007) investigated the neural systems responding to the sight and sound of bimanual paper tearing. Audiovisual facilitation was found in the ventral premotor cortex for the action stimuli but not for non-action control stimuli, suggesting that this region is involved in integrating multimodal information about actions. Third, with TMS Alaerts, Swinnen, and Wenderoth (2009) explored responses of the human primary motor cortex (M1) to visual and auditory stimuli. They observed a selective response increase to congruent audiovisual manual actions, as compared to unimodal and incongruent multimodal manual actions. According to the authors, their findings speak in favour of 'shared' action representations in the human motor system. Fourth, Bidet-Caulet, Voisin, Bertrand, and Fonlupt (2005) report that listening to the footsteps of a walker activates the posterior STS, a region that is, as mentioned in Chapter 1, active during the visual perception of biological motion. Finally, using single cell recordings Barraclough et al. (2005) investigated whether STS neurons coding the sight of actions also integrated the sound of those actions. For 23% of neurons responsive to the sight of an action, the sound of that action significantly modulated the visual response. The sound of the action increased or decreased the visually evoked response for an equal number of neurons. In the neurons whose visual response was increased by the addition of sound (but not those neurons whose responses were decreased), the audiovisual integration was dependent upon the sound of the action matching the sight of the action. These results suggest that neurons in the STS form multisensory representations of observed actions.

Several research groups have very recently started to investigate, behaviorally, the audiovisual perception of biological motion. For instance, Brooks, van der Zwan, Billard, Petreska, Clarke, and Blanke (2007) reported, unlike data from previous studies investigating audiovisual integration in linear motion processing (Alais & Burr, 2004; Meyer & Wuerger, 2001), a left-right direction selective enhancement of the detection of a visually masked point-light walker. Relative to a control condition (stationary sound), auditory motion in the same direction as the visually defined biological motion target increased its detectability (as suggested by decreased reaction times). Auditory motion in the opposite direction had the inverse effect (increased reaction times). Second, using visually masked point-light displays of tap dancing feet and congruent synchronized sounds, Arrighi, Marini, and Burr (2009) demonstrated audiovisual enhancement better than the optimal maximum likelihood prediction (for more details see Chapter 9). Another group of researchers (Saygin, Driver, & De Sa, 2008), designed a task in which observers had to decide whether a periodically moving point-light walker had the same temporal

frequency as a series of auditory beeps that in some cases coincided with the footsteps of the walker. Performance in this task was consistently better for upright point-light walkers compared to inverted or scrambled walkers. The advantage with upright walkers disappeared when the visual ‘‘footsteps’’ were not phase-locked with the auditory events (and instead offset by 50% of the gait cycle) suggesting that visual disruption of multisensory judgments occurred only when the sound was perceived as a consequence of the portrayed action. Petrini et al. (2010) further investigated how experience with a certain multisensory event affects audiovisual synchrony perception of upright and rotated biological motions. Their results showed that the gestalt of upright point-light drumming enhances the detection of audiovisual asynchrony for musical novices, but not for expert drummers. Hence, as shown by Saygin et al. (2008) the nature of the visual stimulation can affect the perceived synchrony between the two sensory signals, but the extent of this effect is constrained by the level of experience with a particular multisensory event. In a previous study, Petrini, Russel, and Pollick (2009) had already shown that prior knowledge acquired through years of experience with a certain audiovisual event appears to be an effective factor in enhancing the ability to detect asynchrony. In their study, six drummers and six novices were asked to judge audiovisual synchrony for point-light drumming actions where the visual information was manipulated to eliminate or include the drumstick–drumhead impact point. In the condition with only the arm information, novices were unable to detect asynchrony whereas drummers were able to. In the conditions that included the impact point, drummers perceived the best alignment when the sight preceded the sound, while in the arm-only condition they perceived the best alignment when the sound occurred together with or preceded the sight, as it would be expected if they were predicting the beat occurrence (also see Petrini, Dahl, et al., 2009). Petrini, Russel, and Pollick (2009) conclude that humans can acquire, through practice, internal action models which can be used to replace missing information when integrating multisensory signals from the environment. Finally, van der Zwan, MacHatch, Kozlowski, Troje, Blanke, and Brooks (2009) paired point-light walkers signalling gender on a continuum from extremely female to extremely male (Troje, 2002) with auditory footsteps that were perceived to be unambiguously female. Figures paired with female footsteps were judged more female. Moreover, adaptation to a gender-neutral walker paired with the female footsteps elicited male after-effects, indicating that the effects operate at the perceptual level.

Part II consists of four experimental chapters reporting on studies on audiovisual action perception, each of which takes a somewhat different approach. In Chapter 8, we

present an adjustment task to estimate the audiovisual point of subjective simultaneity for a repetitive stimulus like the point-light walker. In Chapter 9, we investigate whether a synchronized sound can enhance the visual detection of a point-light walker. In Chapter 10, we explored whether auditory and visual information on the speed of a point-light walker are integrated in a statistically optimal fashion. Finally, in Chapter 11, we present an experiment that examines whether looming and receding sounds affect the perceived in-depth orientation of depth-ambiguous point-light walkers.

Chapter 8

Finding the point of subjective simultaneity in a phase adjustment task

INTRODUCTION AND GOAL

One of the necessary conditions for perceptual integration of an auditory and visual event is their temporal congruency or proximity. In other words, the auditory and visual event need to be presented close enough in time to perceive them as one integrated event. The point of subjective simultaneity (PSS) is the point at which distinct physical events (e.g. the sound and the sight of a footstep) are perceived as maximally simultaneous or synchronized. Recently Arrighi, Alais, and Burr (2006) showed that for rhythmic audiovisual streams the auditory event needs to be delayed relative to the visual event to perceive both events as simultaneous (non-zero, positive PSS). More importantly, they reported evidence that this delay decreases with increasing tempo of the rhythm (frequency dependent PSS). The higher the frequency of the rhythm, the smaller the auditory delay necessary to perceive the audiovisual information as synchronized. Since planned experiments involved presenting point-light walkers accompanied by synchronized sounds representing the footsteps (see Experiment 1 of Chapter 9) at different tempos (see Chapter 10) we first designed an adjustment task to measure the PSS of a sound (auditory event) and the footsteps (visual events) at different walking speeds. Note that as the actual sounds generated by the footsteps of the walker were never recorded, the real audiovisual offset or the objective point of simultaneity remains unknown.

METHOD

Visual stimuli consisted of a 2D projection of the 3D coordinates of a walking figure provided by Vanrie and Verfaillie (2004). At each trial the point-light walker walked as if on a treadmill in the center of the screen in either the left or the right sagittal orientation (random) at one of three speeds: low, middle and high speed. The duration of one step cycle (= two steps) for every speed was respectively 2124ms, 1416ms, or 1062ms. This equals a frequency of the footsteps of respectively 0.94Hz, 1.41Hz, or 1.88Hz. Individual auditory events were presented by professional headphones and consisted of a 30 ms sine wave (800 Hz) of which the first 10 ms were multiplied by the left half of a Gaussian function to avoid ‘clicks’. The frequency of the auditory rhythm (temporally equally spaced events) always matched the step frequency of the point-light walker, but the audiovisual phase relation was randomized.

After a practice block of 18 trials five participants completed 10 blocks of 18 trials (three speeds, six observations per block). On each trial a point-light walker appeared together with the rhythmic auditory stimuli at one of the three speeds (low, middle, high). Participants had to move the entire rhythmic auditory pattern forward (left key) or backward (right key) in time until the sound was perceived as synchronized to the footsteps of the point-light walker. Participants had to press enter to store the audiovisual offset and to initiate the following trial. They were encouraged to make their response within one minute.

RESULTS AND DISCUSSION

Of all stored audiovisual offsets 91% fall between the local vertical maximum and the global vertical minimum of the front foot dot trajectory (Figure 17 C). The Lilliefors test for goodness of fit to a normal distribution rejects the hypothesis that data are normally distributed. The distribution is positively skewed. Skewness = 0.1350 (ses = 0.0816). This is congruent with the general finding that our perceptual system is more tolerant for late arriving sounds compared to early arriving sounds (e.g. Dixon & Spitz, 1980). When normalized to the step cycle there is no difference in PSS between speeds (Figure 17 D: symbols are overlapping). Note that this not necessarily contradicts the findings of Arrighi et al. (2006). The PSS expressed in absolute values (ms) depends on the objective visual reference point, which is unknown.

These data show that for each natural speed of the walker the PSS converges to about the same physical posture (see Figure 18). However, it is not clear to which visual information participants adjusted the auditory ticks. We expected participants to adjust the sound to the global vertical minimum (most right blue dashed line), that is, when the dot representing the ankle reaches its lowest point in the step cycle. The data suggest that this is not the action frame to which people are aligning the auditory information. Participants seem to synchronize the sound to an earlier salient visual event. This may be the local vertical direction change (left blue dashed line), the horizontal direction change (red dashed line) or another salient trajectory change of the dot in between.

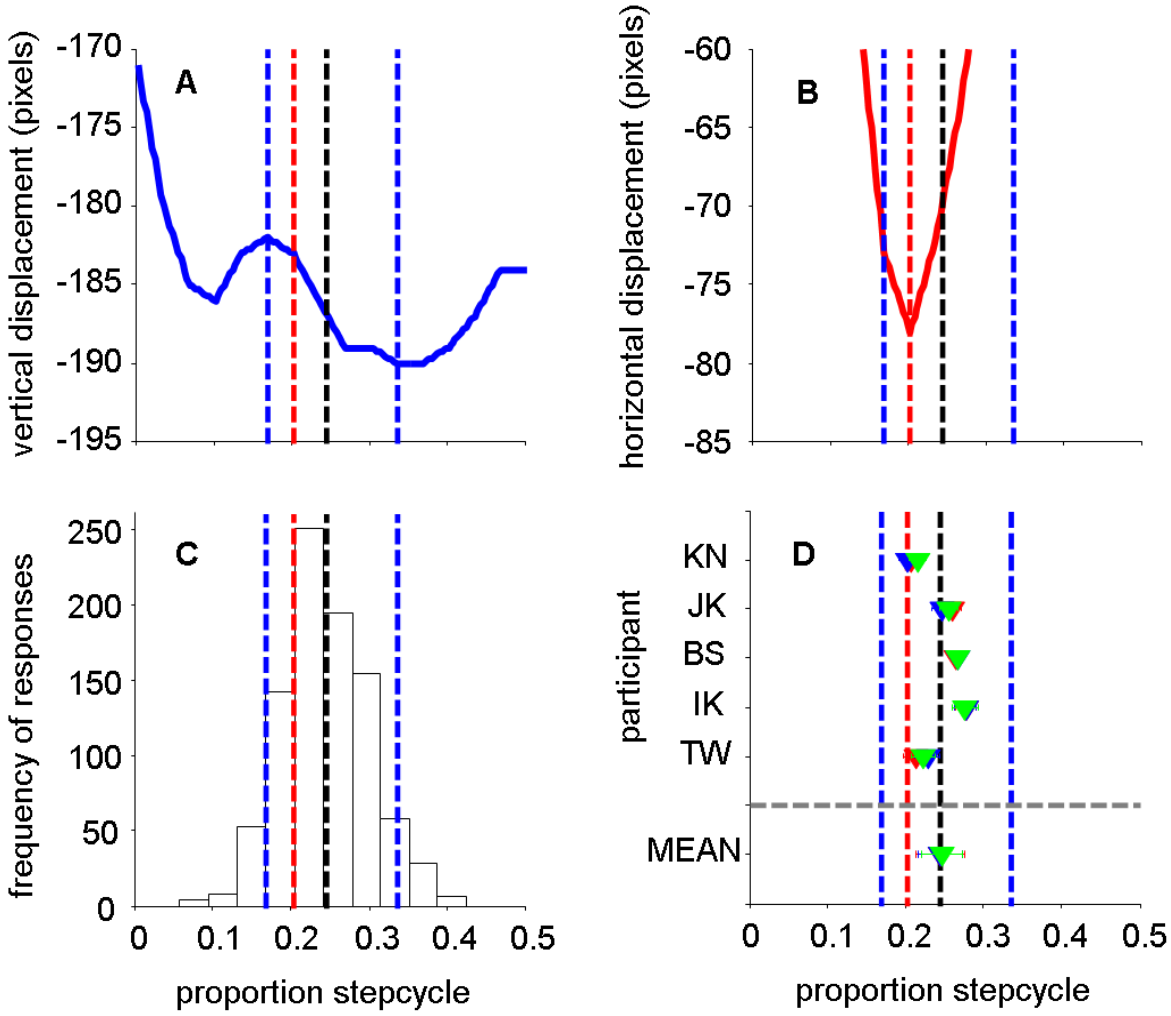


Figure 17. Summary of the data together with some objective relative temporal markers. Depicted as a function of the normalized step cycle: (A) the vertical (blue line) and (B) horizontal (red line) front foot dot trajectory (displacement in pixels), (C) the frequency distribution of all responses and (D) the PSS (red, blue, green triangles) for each participant for each speed (red = low, blue = middle, green = high speed), and the mean PSS for each speed. In all panels the blue dashed lines represent the local vertical extrema (local vertical maximum and global vertical minimum) of the front foot dot trajectory. The red dashed lines represent the front horizontal extremum. The black dashed line represents the overall mean PSS.

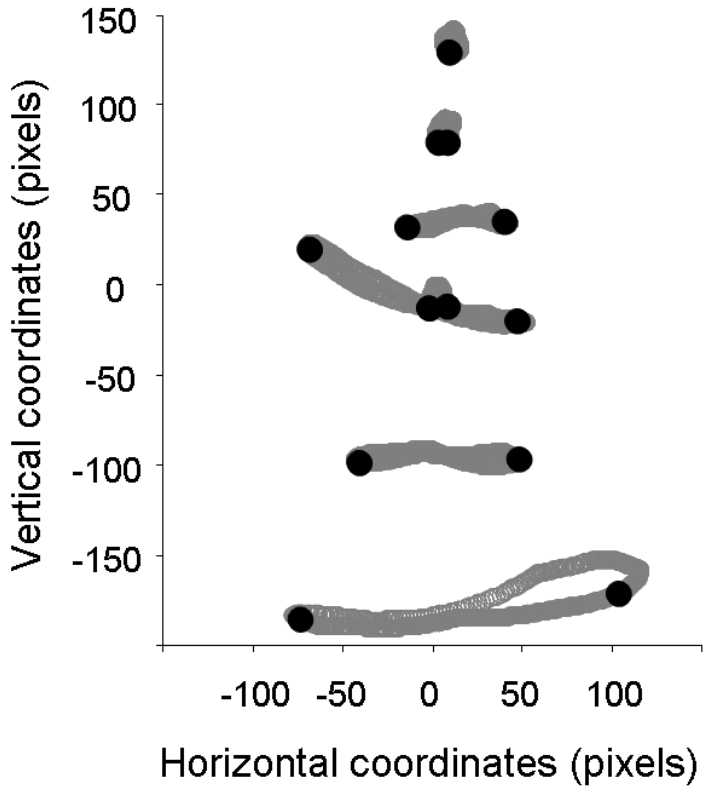


Figure 18. Spatial positions of the dots (in pixels) of a leftward oriented point-light walker. Grey circles represent the spatial positions of each dot of the point-light figure during one entire cycle. The black dots indicate the posture that represents the overall mean PSS.

CONCLUSION

An adjustment task in which participants adjust the sound to a visual event is well suited to estimate the PSS. Points of subjective simultaneity were relatively stable across participants and across realistic walking speeds. With this method it is possible to derive to which specific posture of a sequence of postures a sound has to be coupled to perceive the auditory event as synchronized to the visual events.

Chapter 9

Does synchronized sound enhance the visual detection of a point-light figure?

EXPERIMENT 1

It has been shown that visual processing of point-light actions can be hampered (cf. principle of inverse effectiveness) by a simultaneous mask of moving point lights (e.g. Bertenthal & Pinto, 1994; Cutting, Moore, & Morrison, 1988; Hiris, Humphrey, & Stout, 2005; Thornton, Pinto, & Shiffrar, 1998; Jordan, Reiss, Hoffman, & Landau , 2002). In this experiment we wanted to examine whether the availability of temporally congruent auditory information expedites global visual processing of the action and thereby makes action recognition more resistant to visual masking. More specifically, the goal of the present experiment was to verify whether rhythmic sounds synchronized to the footsteps can enhance the detection of a visually masked point-light walker.

METHOD

The visual stimulus consisted of a limited lifetime point-light walker (Beintema & Lappe, 2002; see Movie 2). Contrary to the classical point-light walker in which the dots remain on the main joints of the figure, in the limited lifetime point-light walker used in this experiment the dots on each new frame of the action are repositioned to a random location between the main joints of the figure (ankles, knees, hips, shoulders, elbows, wrists). The advantage of this kind of point-light figure is that its perceptual detectability can be systematically varied by manipulating the number of dots of which it consists. Moreover, it can be easily masked with a random noise mask (dots added at a random position on the display on each frame).

In a temporal two alternative forced choice task (2AFC) on each trial a visually masked point-light walker was presented for two step cycles (four steps, approximately two seconds) in the left (random interval) and the right (other interval) in plane orientation in the center of the screen (as if on a treadmill). Participants indicated in which of the two intervals the point-light figure was walking to the left. The point-light figures were embedded in different levels of background noise. There were three sound conditions: rhythmic synchronized sound, random sound, and no sound. For each combination of noise level and sound condition the number of dots of which the point-light figure consisted (signal dots) was manipulated and thresholds were estimated by means of three staircases (two down, one up; three down, one up; and four down, one up) converging to three thresholds (70,71 %; 79,40 %; 84,09 % correct, respectively). A cumulative Gaussian was

fitted to these thresholds to derive the 75 % correct threshold for each condition (Wichmann & Hill, 2001a).

RESULTS AND DISCUSSION

Figure 19 displays the estimated thresholds. Generally results revealed no systematic difference in thresholds between the sound conditions. Participants JK and ED show no difference between a rhythmic synchronized sound, a random sound an no sound for three noise levels (30, 60 and 90 dots, see Figure 19 A, B). For participant BS (experimenter) there is no difference in thresholds between a rhythmic synchronized sound an no sound for seven noise levels (30, 40, 50, 60, 70, 80, 90 dots, see Figure 19 C).

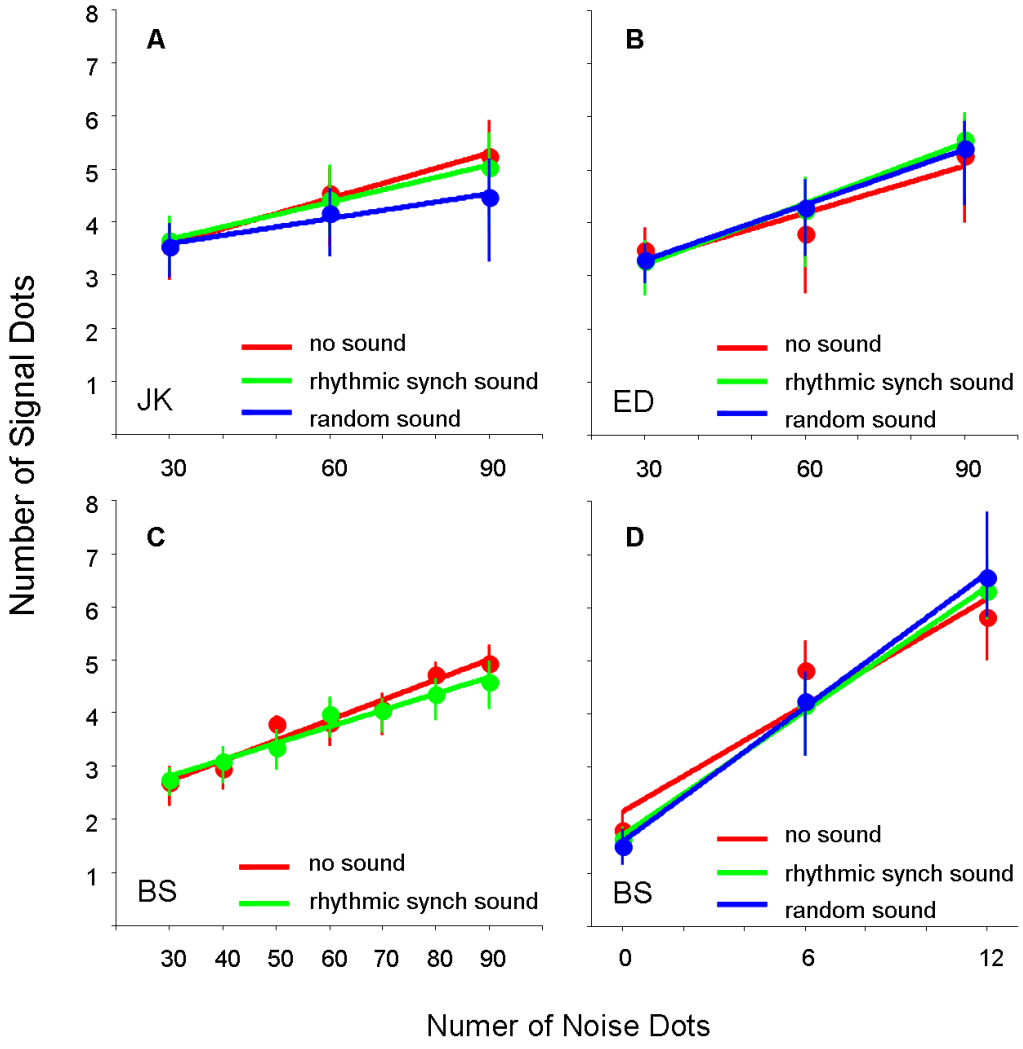


Figure 19. Number of signal dots as a function of number of noise dots for orientation discrimination (A-C) and discrimination between an impossible and veridical walker (D). Error bars indicate 95% confidence intervals.

Possibly, visually the task (orientation discrimination) may have been too easy to allow for any audiovisual enhancement. Therefore we slightly modified the task to demand a much more elaborative processing of the stimulus. Figure 19 D shows the results of BS in a task in which a veridical point-light walker (same stimulus as in previous task) had to be discriminated from an impossible walker - a walker in which the upper part of the body is rotated 180° in depth compared to the lower part, as in Verfaillie (1993, 2000). Huge differences in signal to noise ratios between the tasks (compare abscissas Figure 19 A-C with D) illustrate that indeed this task is much more difficult than orientation discrimination. However, as in the previous task, again no difference in thresholds between the sound conditions was found.

The current data indicate that rhythmic sounds synchronized to the footsteps do not enhance the detection of a visually masked point-light walker. Sound does not help in an orientation discrimination task nor in a task in which a veridical walker has to be discriminated from an impossible walker.

EXPERIMENT 2[12]

In sharp contrast to the findings of Experiment 1, Arrighi, Marini, and Burr (2009) recently reported strong audiovisual enhancement effects in the perception of biological motion. These authors recorded auditory and visual information of a tap dancing action. Visual stimuli consisted of six point-lights representing only the feet of the tap dancer (three dots per foot). Noise dots, construed based on the motions of the feet, were generated to dynamically mask the point-light feet. A signal display contained the point-light feet embedded in noise dots, a noise display contained only noise dots. For three observers (two of which were authors) in a 2AFC task (signal display in one interval, noise display in the other) visual detection thresholds were measured under three different conditions: Synchronised sound (the original sound with the original audiovisual offset; 'in phase'), desynchronised sound (the original sound but with a modified audiovisual offset[13]; 'out of phase'), and no sound. Results showed that for all observers sensitivity was highest when the auditory information was present and in phase with the visual sequence. The increase factor in sensitivity compared to the no-sound condition was on average 1.4

[12] We thank Valeria Manera and Iwona Pomianowska for their help in conducting this experiment.

[13] The size of the audiovisual offset or its direction (sound leading or audio leading) was not reported in Arrighi et al. (2009). In a reply on a question by email concerning this matter, Roberto Arrighi clarified that the offset was randomly chosen (R. Arrighi, personal communication, July, 19th, 2010).

(consistent with the square root of 2). The authors conclude that auditory information, not informative on its own, can facilitate visual discrimination of tap dancing.

In their second experiment Arrighi et al. (2009) investigated summation between auditory and visual signals when both are informative and at threshold. First, discrimination thresholds for visual and auditory signals presented on their own were measured. The visual signal and noise displays were the same as described above, without sound. Auditory stimuli were standardized by substituting the taps with a single template tap sound. In the auditory condition the signal interval comprised the soundtrack of the tap dance (standardized taps) with added noise taps inserted at random times, the other interval comprised only noise taps. In the summation 2AFC task visual and auditory stimuli were normalized to be equally detectable. One interval comprised both the visual and the auditory tap dance sequences, and the other only visual and auditory noise. In one condition the visual and auditory signals were synchronized, in another they were presented out of phase (see Movie 9 for an ilustration of a signal trial in the synchronized condition). Compared to unimodal performance, on average the in-phase condition yielded a relative sensitivity improvement of about 1.6 (better than square root of 2) and the out-of-phase condition an improvement of 1.4 (square root of 2). Although small, this difference was significant. Thus, while performance under audiovisual out-of-phase conditions matched the optimal maximum likelihood prediction[14], performance was significantly better than the optimal maximum likelihood prediction when the audiovisual streams were synchronized. The authors conclude that unlike simple linear translational motion, biological motion shows 'complete or supra-additive audiovisual integration'.

The differences in results between the study of Arrighi et al. (2009) and the results of the present Experiment 1 are puzzling. While Arrighi et al. (2009) claim that biological motion perception shows supra-additive audiovisual integration, we even didn't find any evidence for audiovisual facilitation in Experiment 1. It is unclear what causes this strong discrepancy in results. However, comparing our stimuli with the stimuli used by Arrighi et al. (2009) might reveal an important difference. The movements of the actor, and thus the motions of the dots, in the stimuli of Arrighi et al. (2009) are highly variable and unpredictable. That is, the overall temporal uncertainty of the visual information is relatively high. In contrast, the movements of the actor in our stimuli are quite predictable. In fact, the point-light walker that was used in Experiment 1 consisted of one entire step

[14] When the unimodal reliabilities are matched, the maximum performance increase that can be obtained by statistical optimal integration is the square root of 2 (= 1.41). Also see Chapter 10.

that is looped over and over[15]. Consequently, the time between two footsteps is a constant. In principle, once two successive footsteps are detected, sufficient information is available to predict the third step. Hence, the temporal uncertainty of the visual information remains rather restricted. The enhancing effect of sound on the visual detection of biological motion as reported by Arrighi et al. (2009) might be carried in large by reduction of temporal uncertainty. The lack of temporal uncertainty in our stimulus might thus have been the reason why we did not observe audiovisual enhancement.

The goal of Experiment 2 was twofold. First, we aimed to replicate the results of Arrighi et al. (2009) with point-light stimuli that featured high temporal uncertainty. Second, the out-of-phase audiovisual offset was not reported in Arrighi et al. (2009). We wanted to explore in more detail the role of the audiovisual offset on detection performance under audiovisual conditions. To this end, we captured a new action and designed another experiment. To explore performance under audiovisual conditions as a function of objective as well as subjective simultaneity, we manipulated the audiovisual offset in the performance task and measured the perceived simultaneity as a function of these offsets in a simultaneity judgement task (SJ).

METHOD

Participants. Twelve paid observers, naive as to the aims of the experiment, took part in this experiment. All had normal hearing and normal or corrected to normal vision.

Stimuli. Following the procedure described by Dekeyser, Verfaillie, & Vanrie (2002) a 13 dot point-light action was constructed based on the 3D positions and motions of 30 infrared reflective markers attached to the body of the actor, performing a random jumping action. The action consisted of hand claps (bare hands) and footfalls (shoes on a wooden plate) that were generated by the actor as randomly as possible. The audiovisual action was also simultaneously captured with a standard DV camcorder (Sony DCR TRV 15e). This allowed us to synchronize the video, and hence also the audio (stereo, 48 KHz, 16 bit) which was recorded by an external microphone (Sony F-99EX) attached to the camcorder, to the point-light action (see Movie 10).

The point-light action used in the experiment consisted of a single 6 s random dancing segment of which the first 2 s were linearly faded in and the last 2 s were linearly

[15] The walking action was temporally and spatially perfectly symmetrical: The left side of the body was mirrored to be completely identical to the right side of the body (see Vanrie & Verfaillie, 2004).

faded out. Dots were black on a grey background. The number of dots presented per display could be manipulated (see below) and each individual dot had a limited lifetime of 120 ms (see Neri, Luu, & Levi, 2006; Neri, Morrone, & Burr, 1998). Thereafter, the dot appeared randomly at another of the available positions. Beginning and ending of the lifetime was randomized across dots. Thus, for a particular display on each frame a fixed number of dots was presented and onset and offset of dots were distributed over display time. Besides the signal display, constructed as described above, also distractor displays were created. Distractor displays consisted of temporally scrambled signal displays. Thus, while the signal display carried the original relative temporal relations, in the distractor display the relative temporal relations were randomized. On average, however, positions and motions of the dots of each distractor display equalled those of the signal display (Neri et al., 2006). On each trial of the experiment, the visual display appeared after a fixation cross of 500 ms and a random waiting time between 0 and 350 ms.

Procedure. The experiment consisted of three sequential tasks. The first task was a pretest in which the visual threshold, to be used in the second task, was measured. To avoid the need for a dynamic noise mask[16] to control performance, we designed a spatial 2AFC visual search task. On each trial eight point-light displays appeared on the screen, four on the top row, four on the bottom row. One of the eight displays (randomly chosen on each trial) was the signal display, the other seven displays were distractor displays. Observers had to indicate on each trial which of the two spatial intervals contained the signal display, the upper (arrow up) or the lower row (arrow down). The pretest consisted of two phases. No sound was presented. In each phase the number of signal dots was manipulated (3 levels) and in each phase observers completed 60 trials per signal level (180 randomly chosen trials per phase, divided in 6 sessions of 30 trials). In the first phase, each of the displays contained two, four, or eight dots. This phase of the pretest familiarized the observers with the task and the stimuli. In the second phase of the pretest each display consisted of two, three, or four dots. The second phase of the pretest served to determine for each observer individually the visual threshold, that is, the number of dots to be used in the second task for optimal performance (around 70% percent, to avoid ceiling and floor effects). A cumulative Gaussian was fitted to proportions correct for each of the

[16] The major challenge with masking point-light figures concerns controlling for the density of the signal dots which often serves as a cue to observers whether or not a signal is present in a cloud of noise dots. Especially when observers are well trained (for stable performance in psychophysical experiments), they may finally pick up on this cue. In such cases, task performance tends to be stable but it does not provide relevant information on the detection of biological motion.

three conditions and the 70% threshold was determined. Threshold values were rounded to the closest integer value.

The second task was an audiovisual search task. It was similar to the pretest, with two exceptions. First, instead of manipulating the number of dots of which the displays consisted, the number of dots was now fixed to the rounded threshold value determined in the pretest. Second, except for the no-sound-condition, the presentation of the visual stimulus was accompanied by the sound-track of the random jumping action with a particular audiovisual offset (-300, -200, -150, -100, -50, 0, 50, 100, 150, 200, 300 ms; negative values indicate that audio is leading; positive values indicate that audio is delayed relative to video). Thus, in the audiovisual search task there were 12 conditions: A no-sound condition and 11 audiovisual offset conditions (see Movie 11 for an example of a trial in the synchronized condition). In this task observers completed 50 repeats of each condition (600 randomly ordered trials in total, divided in 25 sessions of 24 trials).

The third task was an SJ task. On each trial a single point-light dancer, consisting of the same number of signal dots as was determined by the pretest was presented in the centre of the screen. The visual display was always accompanied by the sound-track with one of the same eleven audiovisual offsets as used in the audiovisual search task. Observers had to indicate by a key press on each trial whether they perceived the sound as synchronized (key 1) to the visual display or not (key 2). In total each participant completed 275 randomly ordered trials (11 offsets * 25 repeats), divided in 11 sessions of 25 trials.

RESULTS AND DISCUSSION

In Figure 20 we plot the mean proportions (across observers) of correct responses (black circles) in the audiovisual search task and the proportions of simultaneous responses (grey squares) in the SJ task as a function of audiovisual offset. The dashed black line represents the proportion of correct responses in the no-sound condition (audiovisual search task). The mean proportion of correct responses was 71.53% across all sound conditions, 71.52 % across the 11 audiovisual offsets, and 71.67% in the no-sound condition. From these observations it is evident that there was no effect of sound condition. This was confirmed by a repeated measurement ANOVA on the d-primes with sound condition as a within-subjects variable ($F(11,121) = 0.44$, $p = 0.93$). Overall mean reaction

time was 4545 ms. A repeated measurement ANOVA on the reaction times revealed no effect of sound condition ($F(11,121) = 0.75$, $p = 0.67$).

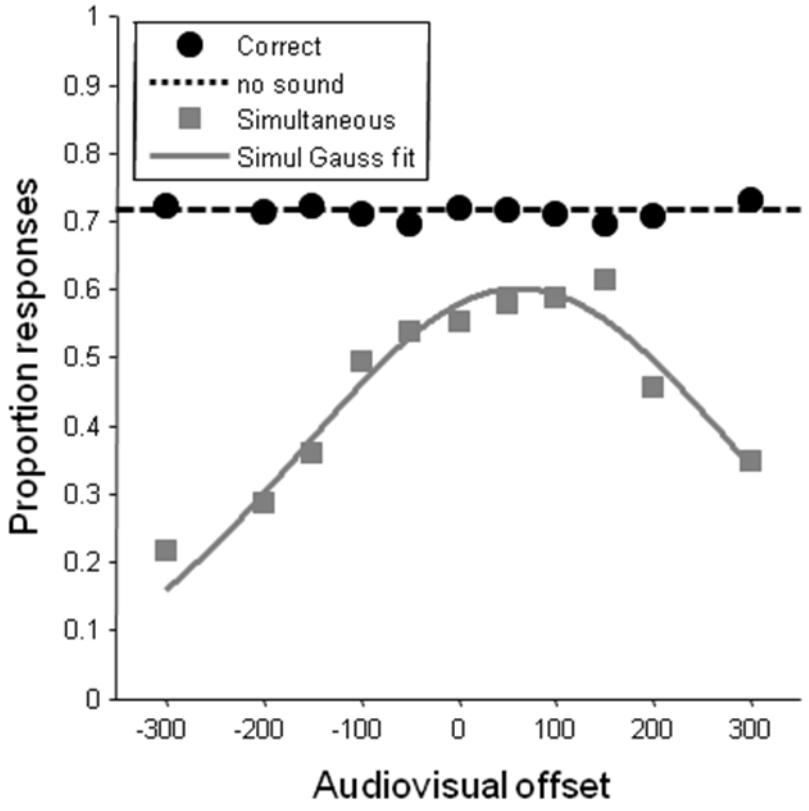


Figure 20. Proportions of correct (audiovisual search task) and proportions of simultaneous responses (SJ task) averaged across observers as a function of audiovisual offset. The black dashed line represents the average proportion correct in the no-sound condition. The grey solid line represents the Gaussian fit to the simultaneous responses.

For the proportions of simultaneous responses a rather atypical data pattern is observed. Based on previous research (e.g., Arrighi, Alais, & Burr, 2006; Dixon & Spitz, 1980; Van Eijk, Kohlrausch, Juola, & Van de Par, 2009) we expected proportions of FTV responses as a function of audiovisual offset to rise relatively fast, to peak between -50 and + 100ms and then to drop slowly. Instead, the average proportions of simultaneous responses seem to rise slowly, peak at about +150 ms and then drop, relative to the rise, quite steeply. Following traditional procedures (in which the PSS is estimated as the mean of a symmetrical Gaussian fit to the proportions of simultaneous responses) the estimated PSS was +63 ms (CI_{95} = [38, 89]). Thus, if only the Gaussian fit is considered, based on the expectations described above, the pattern of simultaneous responses may look rather

normal (see our earlier remark in Chapter 7, p. 102 on the often used procedure to fit a symmetrical Gaussian to proportions of simultaneous responses).

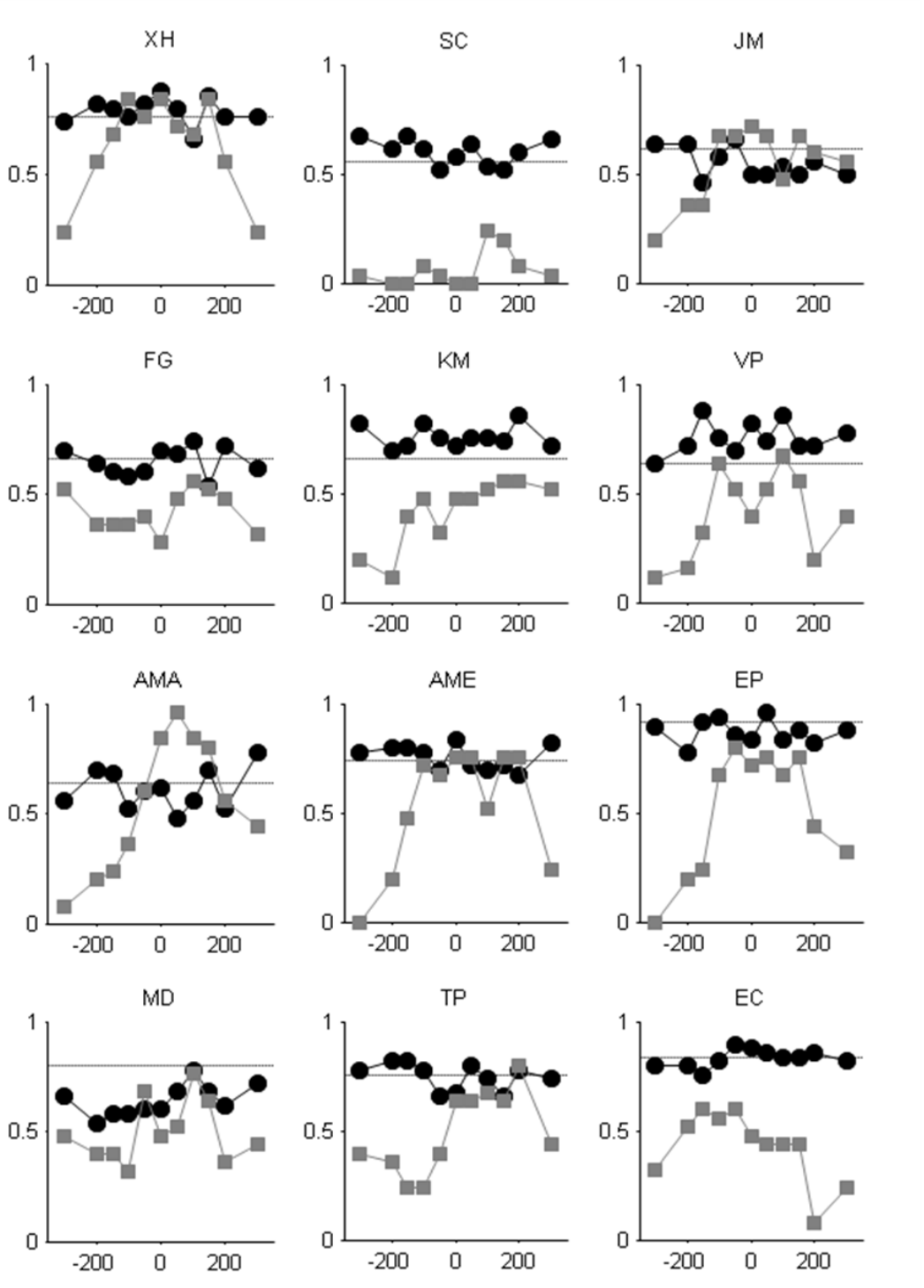


Figure 21. Proportions of correct (black circles) and proportions of simultaneous responses (grey squares) for each of the twelve observers as a function of audiovisual offset. The horizontal line represents the average proportion correct in the no-sound condition.

To clarify the origin of the atypical data pattern of proportions of simultaneous responses across observers, in Figure 21 we plot the same data but for each observer separately. Only one out of the 12 observers, AMA, shows a more typical simultaneity judgement pattern. For most of the observers, however, simultaneity judgement data appear quite distorted. Notably, AMA was one of the few observers who converged to four signal dots in the pretest (only EC also converged to four signal dots in the pretest). That is, for AMA, in the audiovisual search task as well as in the SJ task, all point-light displays consisted of four signal dots while for most other observers (except for EC) in both tasks the point-light displays consisted of three or two dots. This may indicate that, to accurately judge the audiovisual simultaneity of the point-light dancing action, a minimum amount of dots of which the figure consists, is required. However, the fact that a rather distorted pattern of simultaneous responses was found also for EC indicates that having perceptually available the point-light dancer consisting of four dots is not sufficient to obtain typical simultaneous judgement data. Also worthy to note is that, after the experiment, observer MD spontaneously reported to not have recognized the meaning of the sound. This observer did not perceive the sound as related to the dancing action but as generated by "bouncing billiard balls". This suggests that for this observer (and thus maybe also for other observers) no audiovisual binding occurred between the degraded visual information of the dancing action and the sound that it produced.

GENERAL DISCUSSION

The general goal of the two studies reported in the present chapter was to investigate whether synchronized sound can enhance the visual detection of a point-light action. In Experiment 1 we found no evidence of audiovisual enhancement for the detection of a visually masked point-light walker. Contrary to our results, Arrighi et al. (2009) reported strong audiovisual enhancement effects for the detection of a masked point-light tap dancing action. Assuming that to observe audiovisual enhancement temporal uncertainty was a crucial property of the point-light action in Experiment 2 we tried to replicate the results of Arrighi et al. (2009) with a point-light dancing action in an audiovisual search task. In this experiment we also aimed to explore in more detail the role of objective and subjective audiovisual simultaneity. Contrary to what was expected based on the results of Arrighi et al. (2009), but consistent with the results of Experiment 1, in Experiment 2 we did not find any evidence for audiovisual enhancement.

The discrepancy between our results and the results of Arrighi et al. (2009) are vast and puzzling. As finding the causes of this discrepancy will help to identify the necessary conditions for audiovisual enhancement effects to occur, in the remainder of this chapter we will discuss some potentially important factors.

First, it is possible that in Experiment 2 we did not completely tackle the lack of temporal uncertainty in our point-light stimulus. From the methods described by Arrighi et al. (2009), it appears that in their study on each trial one of three possible action sequences was displayed. In our Experiment 2, on each trial the same action segment was presented. Note, however, that the starting time of the action segment in our experiment was randomized (random waiting time before the 6s display). It could be argued in principle, that the different number of action segments (three vs. one) constituted a relevant difference between both studies, because relative to Arrighi et al. (2009), the temporal uncertainty in our experiment was still lower. However, in our opinion it is unlikely that this is what caused the differences between the results. In Arrighi et al. (2009) the same observers participated in both of their experiments making it likely that, at least to some degree, participants learned the temporal structure of the three sequences. Also important to note here is that when interpreting the results of their summation task, Arrighi et al (2009) suggest that it is not the reduction of temporal uncertainty but a meaningful integration of the audiovisual stimulation that is causing the increase in performance.

Second, whereas in our studies the limited lifetime technique (Beintema & Lappe, 2002; Neri et al., 1998, 2006) was used to impair visual detection, in Arrighi et al. (2009) visual detection performance was impaired only by adding different levels of noise. It is thus possible that in our studies the limited lifetime technique prevented audiovisual integration. Relevant to refer to here is a recent study of Van der Burg, Cass, Olivers, Theeuwes, and Alais (2010) that elaborates their findings on the pip pop effect (see Chapter 7 or Van der Burg, Olivers, Bronkhorst, Theeuwes, 2008). In a set of four experiments Van der Burg et al. (2010) showed that audiovisual synchrony will only lead to benefits in visual search when the changes are transient. That is, the change of the visual signal only binds to the auditory signal if it is short enough in time. This suggests that in the stimuli of Arrighi et al. (2009) the sound is binding to the sudden impact of the feet on the floor (vertical direction changes or a sudden stop of the motions of the tap dancing feet). In contrast, in our studies sound may not have had the opportunity to bind to the relevant impacts or direction changes. Indeed, in Experiment 1 the vertical direction change that represents the footfall (impact of the foot on the ground) is gradual and rather

spread over time (see Figure 18 in Chapter 8). Moreover, in Experiment 1 none of the dots themselves actually ever changed direction because the dots contained no motion but were redrawn on another location on each frame of the animation. In Experiment 2, direction changes of the feet of the underlying action were more abrupt. However, in this experiment we also used the limited lifetime technique. Dots were redrawn after a lifetime of 120 ms (see Neri et al., 2006). It could thus be possible that by using the limited lifetime technique we disrupted the visual information that is crucial to bind with a synchronized sound. In fact, the general poor performance in the SJ task seems to corroborate this idea. Indeed, the limited lifetime technique and the reduction of the number of signal dots seem to disrupt the visual ankers that are necessary for adequate simultaneity judgement and audiovisual binding.

Third, in our studies audiovisual binding may have occurred for other than the intended audiovisual stimulus or stimulus properties. Several other salient visual properties of our stimuli might have competed with the abrupt direction changes to bind with the abrupt sounds. For example, if indeed sound specifically binds to an abrupt visual onset (or offset) as shown by Van der Burg et al. (2010), sounds may have bound to the repetitive appearances (or disappearances) of the dots. The limited lifetime technique may thus not only disrupt the necessary visual information of the relevant direction changes, it may also provide the perceptual system with an overload of visual onsets (and offsets) that all meet the conditions to bind with the abrupt sounds. Other examples of potential false audiovisual matches in Experiment 2 may be the distractor stimuli or even the movements of other dots within the signal stimulus. Indeed, as the distractor stimuli were temporally randomized signal stimuli, it is possible that direction changes of at least some of those dots were simultaneous with the sounds of at least some of the audiovisual offsets. Note, in contrast, that in Arrighi et al. (2009) the signal stimuli were represented by six dots, three on each foot. The three dots representing each foot constituted a rigid structure. That is, the relative spatial positions between those dots remained constant. Within their signal stimulus there were thus no competing dot movements that could have bound to the sound.

An important question that arises when taking the above mentioned research into consideration is whether the audiovisual enhancement reported by Arrighi et al. (2009) is specific to biological motion perception. Do the stimuli really need to be ‘meaningful’? What is clear from Arrighi et al. (2009), though, is the fact that the presence of a meaningful entire human structure is not necessary to observe strong audiovisual enhancement effects. Rather, the motions and in particular the direction changes or impacts

of six dots representing two feet seem to carry the crucial visual property to which the sound is binding. But, if this property is nothing more than the clear presence of visual transients that do not have to compete with many other irrelevant visual transients, then it could be questioned what the difference is between the results of Arrighi et al. (2009) and the studies of Van der Burg et al. (2008, 2010) that feature meaningless visual and auditory stimuli. Critically, Arrighi et al. (2009) did not include a control condition with meaningless stimuli. We will come back to this issue in the last section of this dissertation (Concluding remarks and future research).

Chapter 10

Are auditory and visual information on the speed of walking integrated in a statistically optimal way ?

INTRODUCTION

If information on the same perceptual dimension is available to different modalities the brain has the opportunity to integrate the different sources of information in a statistically optimal fashion (see Chapter 7). For example, when both the auditory and visual stimuli are informative on the walking speed of a person, then a more reliable perceptual estimate of the walking speed can be obtained if both sources of information are combined or integrated, compared to when only the auditory or the visual information is used. The signature of integration is a reduction of variance (Ernst & Banks, 2002; Ernst & Bulthoff, 2004).

The goal of this study was to verify whether or not auditory rhythmic ticks, synchronized to the footsteps of a point-light walker, are integrated with the visual information of the point-light walker in a statistically optimal fashion.

EXPERIMENT

METHOD

In a temporal 2AFC task two naïve participants (EE and JK) judged which of the two successive presentations of the stimuli was the fastest. There were three conditions: (1) an auditory rhythmic pattern (rhythmic ticks at a certain frequency), (2) a visual rhythmic pattern (point-light walker with footsteps at a certain frequency), and (3) an audiovisual rhythmic pattern (rhythmic auditory ticks synchronized with the footsteps of a point-light walker at a certain frequency). On each trial, one of four higher and four lower speeds (test stimuli, random interval) from one of the three conditions was compared with a middle speed (reference stimulus, other interval) of the same condition which allowed to estimate the variance of the psychometric speed discrimination curves (cumulative Gaussian fits) for each condition. The smaller the variance, the higher the reliability, the steeper the curves and the better the discrimination. Given that the noise of individual estimates are independent and Gaussian, the estimate with the lowest variance is the Maximum Likelihood (ML) Estimate (Ernst & Banks, 2002; Ernst & Bulthoff, 2004). Based on the unimodal variances in the auditory modality (σ^2_A) and the visual modality (σ^2_V) the ML variance of the combined modalities can be predicted ($\sigma^2_{AV\ pred} = (\sigma^2_A * \sigma^2_V)/(\sigma^2_A + \sigma^2_V)$) and compared to the observed (σ^2_{AV}) audiovisual variance. As for the variance, also the

ML mean of the audiovisual psychometric curve can be predicted. The ML mean is the weighted sum of the unimodal means ($\mu_{AV\ pred} = (\mu_A * w_A) + (\mu_V * w_V)$), in which the weights are the normalised reliabilities ($w_A = r_A/(r_A + r_V)$ and $w_V = r_V/(r_A + r_V)$) with the reliabilities being the inverse of the unimodal variance ($r_A = 1/\sigma^2_A$ and $r_V = 1/\sigma^2_V$).

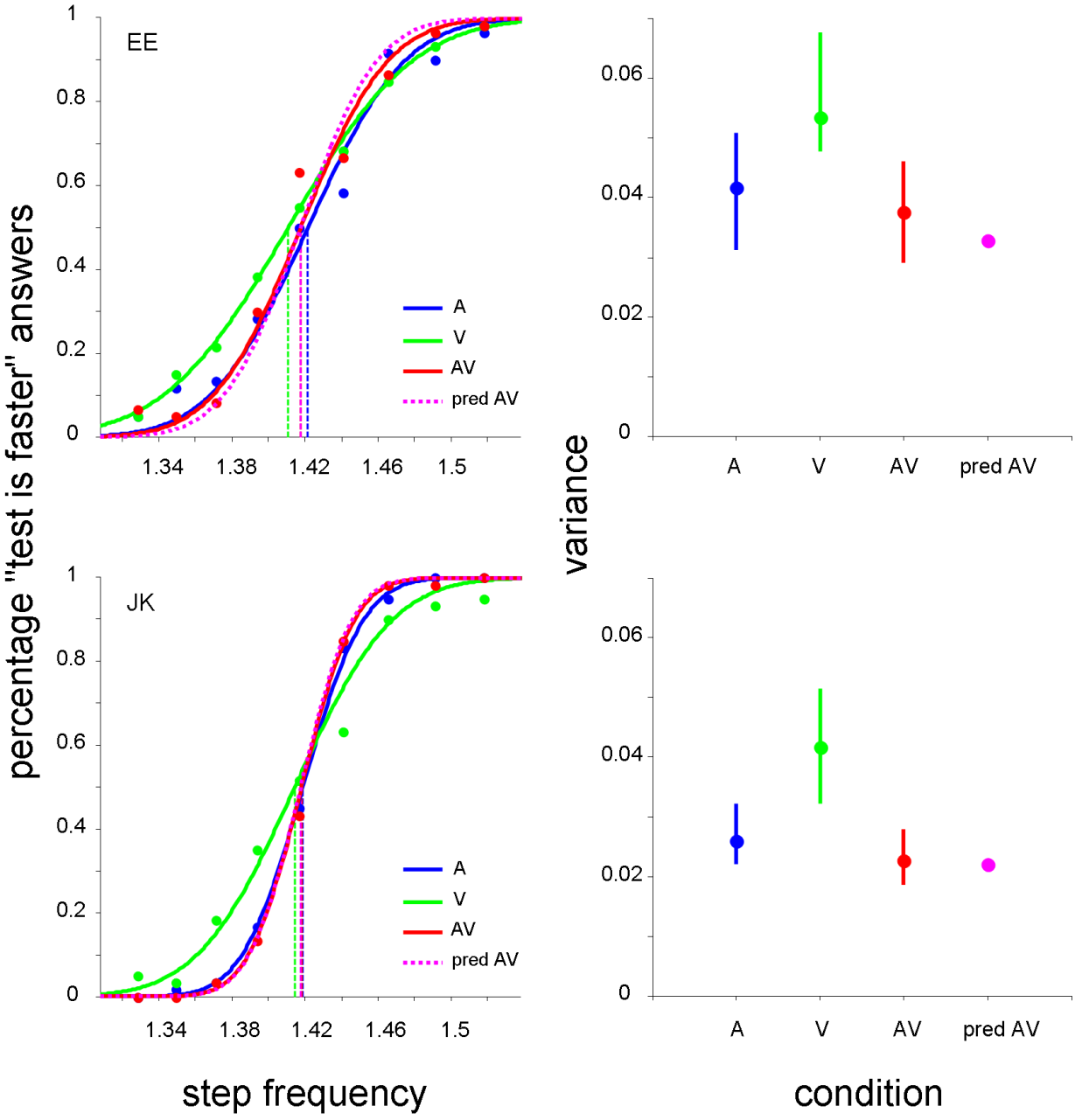


Figure 22. Data, cumulative Gaussian fits, means (left) and variances (right) of the auditory (blue), visual (green) and audiovisual (red) performance of two naïve participants EE and JK compared to the predicted Maximum Likelihood audiovisual (magenta) variance and mean. Error bars (right) indicate 95% confidence intervals on the estimated variances.

RESULTS AND DISCUSSION

Despite the fact that both EE and JK completed 60 trials for each data point on each psychometric curve (1620 trials in total) performance was relatively variable (large confidence intervals on estimated variances). The results (see Figure 22) showed that there was no significant difference between the observed audiovisual variance and the predicted audiovisual Maximum Likelihood variance. This is consistent with statistical optimal integration. However, there was no significant difference between the variance in the best unimodal condition and the predicted audiovisual variance either. To be able to discriminate between statistical optimal integration and using the most reliable unimodal cue the predicted audiovisual variance should be significantly lower than the lowest unimodal variance.

CONCLUSION

This experiment could not answer the question whether the auditory and visual information on the speed of walking are integrated in a statistically optimal fashion. The results do not exclude statistical optimal integration but they do not exclude the use of the best unimodal cue either. The current data allow concluding that audiovisual performance was as least as good as the best unimodal performance. To optimize this experiment, more control on the reliability of the unimodal stimuli is necessary. An experiment in which both the visual and auditory modalities are moderately and equally reliable will provide the largest power in discriminating between statistical optimal integration and the use of the best unimodal cue.

Note:

Without going into the details of the results, we shortly report another experiment that is a variant on the study reported above. Instead of comparing auditory, visual and audiovisual conditions, in this experiment three audiovisual conditions were tested. The audiovisual conditions differed only by their audiovisual phase-offset. In the first condition the auditory ticks were perfectly synchronized to the footsteps of the point-light walker (phase-offset = 0ms, mean PSS found in Chapter 8). In the second condition the sound occurred before the footstep (phase-offset = -200 ms). In the third condition the sound occurred after the footstep (phase-offset = 200 ms). Similar to the experiment described

above, the variances of the psychometric curves in each condition were measured. Based on the assumption that temporal congruency is crucial for audiovisual integration, the variances in the synchronized condition were expected to be significantly lower than in both non-synchronized conditions. Remarkably, this expected significant difference only showed up for participant B.S. (experimenter). Three naïve subjects did not show any significant differences between the phase conditions. This may be due to the fact that independent of its phase relation to the visual stimulus the auditory modality remains the most reliable in this task. Combining or integrating it with the visual information may not significantly enhance audiovisual speed discrimination performance. Alternatively, people may have combined the audiovisual information in the non-zero offset conditions as well. Indeed, one participant reported coupling the sound to the arm movement in the non-zero offset conditions. This may underlie the absence of any difference between the offset conditions for the naïve subjects. The single expected and significant difference for BS could be due to the 'willingness' to combine both sources of information in the zero-offset condition and the 'unwillingness' to couple the auditory information to something different than the footsteps (cfr. unity assumption, see Chapter 7). However, this is close to an interpretation of a pure experimenter effect. Taken together this study does not allow drawing firm conclusions on audiovisual integration in the estimation of walking speed (also see Concluding remarks and future research).

Chapter 11

Do looming and receding sounds affect the perceived in-depth orientation of depth-ambiguous biological motion figures?

.

This chapter is based on:

Schouten, B., Troje, N. F., Vroomen & Verfaillie, K. (2010). Looming and receding sounds affect the perceived in-depth orientation of depth-ambiguous biological motion figures*, Plos One,* Manuscript under revision.

ABSTRACT[17]

The focus in the research on biological motion perception traditionally has been restricted to the visual modality. Recent neurophysiological and behavioural evidence, however, supports the idea that actions are not represented merely visually but rather audiovisually. The goal of the present study was to test whether the perceived in-depth orientation of depth-ambiguous point-light walkers is affected by the presentation of looming or receding sounds synchronized with the footsteps.

Orthographic frontal/back projections of point-light walkers were presented either without sound or with sounds of which the intensity level was rising (looming), falling (receding) or stationary. Despite instructions to ignore the sounds and to only report the visually perceived in-depth orientation, point-light walkers accompanied with looming sounds were more often perceived to be facing the viewer whereas point-light walkers paired with receding sounds were more often perceived to be facing away from the viewer.

The present results suggest that looming and receding sounds can alter the perceived in-depth orientation of depth-ambiguous point-light walkers. The challenge for future research will be to find out at what processing level these effects occur.

INTRODUCTION

The goal of the current study was to explore whether sound affects the visual in-depth perception of point-light figures. Recently, the visual perception of depth-ambiguous biological motion attracted scientific interest (Brooks, Schouten, Troje, Verfaillie, Blanke, & van der Zwan, 2008; Jackson & Blake, 2010; Stuart Jackson, Cummins, & Brady, 2008; Schouten, Troje, Brooks, van der Zwan, & Verfaillie, 2010; Schouten, Troje, & Verfaillie, in press; Schouten & Verfaillie, 2010; Vanrie, Dekeyser, & Verfaillie, 2004; Vanrie & Verfaillie, 2006, 2008). An orthographic 2D projection of a symmetrical point-light walker is perceptually bistable: It is either perceived to be 'facing the viewer' (FTV) or it is perceived as 'facing away' from the viewer (FA). Nevertheless, observers exhibit a 'facing bias', that is, a perceptual bias to interpret the figure as being oriented *towards* the viewer (Vanrie, et al., 2004).

[17] Thanks to Elke Moyens for her help in two pilot studies (Schouten, Moyens, Brooks, van der Zwan, & Verfaillie, 2008).

Here, we examined whether auditory information that was congruent either with an approaching or with a receding walker would influence the in-depth interpretation of a depth-ambiguous point-light walker. To this end, we presented to observers orthographically projected point-light walkers in four sound conditions. Walkers were either presented unimodally or they were paired with sounds synchronized to the footsteps. Sounds were either stationary or looming or receding. To induce apparent sound-depth (Rosenblum, Carello, & Pastore, 1987) we increased/decreased the intensity level of subsequent tones to create looming/receding sounds, respectively (Neuhoff, 1998). Stationary sounds had a constant intensity level. On each trial, observers indicated whether they perceived the point-light walker as facing towards them or facing away from them.

Point-light walkers usually elicit a strong facing bias. Therefore, to avoid ceiling effects in the looming condition, here we included, besides a point-light walker that consistently elicits a facing bias, also a point-light walker that consistently is perceived as rather ambiguous concerning its in-depth orientation (see methods). To minimize demand effects, observers were instructed to ignore the sound and to only respond to the visual information. Thus, if sounds had no effect, proportions of FTV responses should be similar in all sound conditions. Alternatively, if looming and receding sounds had an effect such that the perceived in-depth orientation was modified in the direction suggested by the sound, proportions of FTV responses should be higher in the looming condition and lower in the receding condition, compared to the no-sound condition. The proportions of FTV responses in the stationary condition should not differ from the no-sound condition because stationary sounds do not induce a particular in-depth interpretation.

EXPERIMENT

METHOD

Participants. Nineteen students from the University of Leuven participated for course credit. All observers had normal or corrected to normal vision and hearing and were naïve to the purpose of the experiment.

Stimuli and apparatus. There were two types of point-light walkers: A structure-only and a kinematics-only male point-light walker both derived from a gender continuum. A structure-only male point-light walker was composed of gender neutral kinematics (z-score: 0 std) and male structure (6 std). A kinematics-only male point-light walker was

composed of gender neutral structure (0 std) and male kinematics (6 std). See Troje (2002, 2008b) for details on stimulus creation. Note that in the present study the gender of the point-light walker is irrelevant. In Schouten, Troje, & Verfaillie (in press) it was shown that changes in the facing bias as a function of figure gender (see Brooks, et al., 2008; Schouten, et al., 2010) are not caused by changes in perceived gender but result from structural and kinematic stimulus properties. Here, we simply employ the empirical finding that particular stimulus properties consistently elicit a particular proportion of FTV responses. The choice of these two types of point-light walkers in the present study is thus solely based on their earlier observed property to either elicit a facing bias (the structure-only male point-light walker; about 81% of FTV responses) or to elicit a rather ambiguous in-depth orientation (the kinematics-only male point-light walker; about 54% of FTV responses). Each point-light walker was orthographically projected in frontal view in the centre of the screen (subtending a size of 8º of VA height) and each consisted of 15 black dots (radius = 15 arc mins of VA) on a grey background. Total trial duration always matched the duration of 4 walking cycles, that is, 8 steps (96 frames per cycle at 85 fps = 1130ms per cycle; 384 frames in total = 4518ms). Visually, trials consisted of a grey background presented for a duration of 3 step cycles (288 frames = 3388ms). Then, the point-light walker appeared for one step cycle (two steps; 96 frames = 1130ms). This was done to prevent observers to only rely on visual information near trial onset. The sound was played during the entire trial and a convincing looming or receding auditory percept was only obtained near the end of the trial. Auditory stimuli representing the footsteps consisted of eight 30ms sine waves of 200Hz, amplitude modulated depending on the required sound level. To avoid clicks, the first 15 ms of the sine waves were ramped. Note that the objective point of audiovisual simultaneity for these point-light walkers is unknown because the sounds of the footsteps were not recorded during the motion capture phase of the stimulus creation (Troje, 2002). Audiovisual synchrony was therefore set to match the subjective point of simultaneity as obtained in an earlier adjustment task (see e.g. Schouten & Verfaillie, 2006 or Chapter 8). The adjustment task provided the frame of the step cycle to which the sound had to be coupled to obtain maximum perceived synchrony. This point was then set to represent an audiovisual offset of 0 ms. In the looming condition the sounds linearly increased in intensity from 40 dB to 64 dB SPL. To mimic the decrease in auditory delay of an approaching sound source due to the slow travel time of sound (approximately 343 m/s), for each step (approximately 70 cm in natural circumstances) the audiovisual offset was decreased with 2ms (from 14 ms to 0 ms;

positive values indicate auditory delay). The reverse settings (linearly decreasing intensity from 64 dB to 40dB SPL and linearly increasing audiovisual offset from 0 ms to 14 ms) were applied to create receding sounds. In the stationary condition the sound level (55 dB SPL) and the audiovisual offset (0 ms) were constant. In the no-sound condition no sound was presented. On each trial the start position of the point-light walker was randomly chosen from the first 24 frames of the step cycle.

Procedure. All observers were tested individually in a dimly lit and sound attenuated room, sitting in front of a 21 inch CRT monitor at a viewing distance of 57 cm. Sounds were presented by three Dell speakers located 80 cm behind the screen: two satellite speakers (distance between left and right speaker was 70 cm) and one subwoofer, located centrally. Observers were asked to indicate on each trial by a key-press whether the point-light walker was perceived as facing towards them or facing away from them (arrow down for FTV, arrow up for FA). Observers were explicitly instructed to ignore the sound and to only report the visually perceived in-depth orientation. It was stressed that an equal distribution of both response alternatives was not necessary. After instructions, observers completed 32 practice trials (random selection of all possible conditions). Thereafter, the experiment commenced. Trials of all conditions were randomized and then divided in blocks of 32 trials between which participants could take a break. In total each participant completed 320 trials (2 types of point-light walkers * 4 sound conditions * 40 repetitions).

RESULTS

Figure 23 depicts the proportions of FTV responses for the two types of walkers in each of the four sound conditions. We analysed the data using a repeated measurement ANOVA and paired t-tests on probit transformed proportions. The average amount of FTV responses was 67%. Consistent with previous research (Brooks, et al., 2008; Schouten & Verfaillie, 2010; Vanrie, et al., 2004), we found a facing bias for point-light walkers. Also in line with previous research (Schouten, Troje, & Verfaillie, in press), there was a strong facing bias for the structure-only male point-light walker in the no-sound condition (81% of FTV responses). This bias was significantly smaller for the kinematics-only male point-light walker (56%, $t(18) = -4.39$, $p < 0.001$). Also across all sound conditions, the number of FTV responses was smaller for the kinematics-only male point-light walker (55%) than for the structure-only male point-light walker (79%). This was confirmed by a repeated measurement ANOVA with point-light walker type and sound condition as factors. There

was a main effect of point-light walker type ($F(1,18) = 22.25$, $p < 0.001$). This analysis also revealed a main effect of sound condition ($F(3,54) = 7.47$, $p < 0.001$). There was no interaction between sound condition and point-light walker type ($F(3,54) = 0.34$, $p = 0.796$).

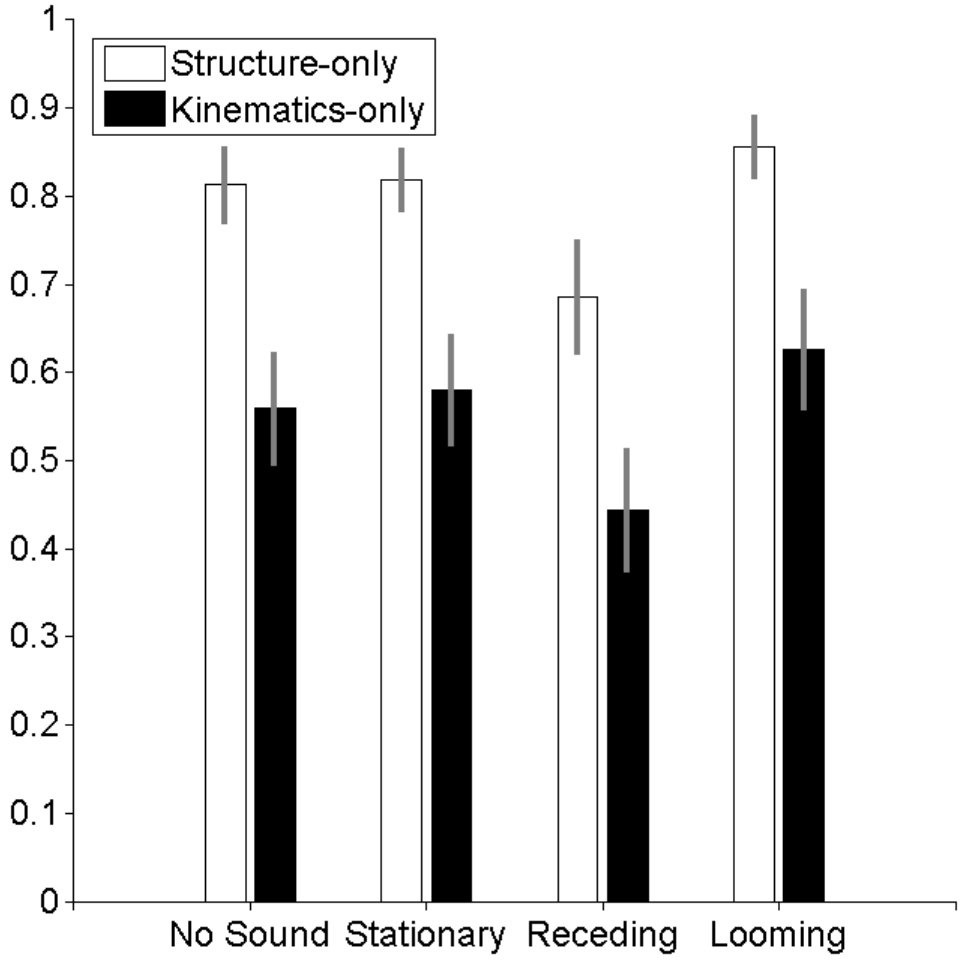


Figure 23. Mean proportions (across observers) of FTV responses for the structure-only male point-light walkers (white) and the kinematics-only male point-light walkers (black) in the no-sound, stationary, receding, and looming conditions. Error bars indicate +- 1 SE from the mean. Proportions of FTV responses are higher in the looming condition and lower in the receding condition compared to the no-sound and stationary condition. The effect of sound is similar for both types of point-light figures.

The crucial question in the current study was whether the amount of FTV responses in the looming condition was higher and in the receding condition was lower compared to the no-sound condition and the stationary condition. The amount of FTV responses in the

stationary condition and in the no-sound condition should not differ. To test this, we performed five paired t-tests with a familywise alfa of 0.05 (all t-tests were two-sided). The Bonferroni corrected alfa for each test was 0.01. The results showed that the number of FTV responses in the looming condition (74%) was significantly higher compared to the no-sound condition (69%; $t(37) = 3.33$, $p = 0.002$), and compared to the stationary condition (70%; $t(37)= 3.23$, $p = 0.003$). The number of FTV responses in the receding condition (56%) was significantly lower compared to the no-sound condition ($t(37) = -3.10$, $p = 0.004$) and compared to the stationary condition ($t(37) = -3.36$, $p = 0.002$). The amount of FTV responses in the stationary condition and in the no-sound condition did not differ ($t(37) = 0.04$, $p = 0.96$). Overall, these results indicate that, despite the instruction to ignore the sound, the visually perceived in-depth orientation of point-light walkers was reliably shifted in the direction suggested by the sound. Sound affected point-light walkers that elicit a facing bias and point-light walkers that are perceived as more depth-ambiguous, in the same way.

DISCUSSION

We tested whether looming or receding sounds, synchronized to the footsteps of depth-ambiguous point-light walkers, can alter their visually perceived in-depth orientation. Results show that looming sounds lead to more 'facing the viewer' responses whereas receding sounds have the opposite effect and lead to more 'facing away' responses.

The current data suggest that looming and receding sounds affect the in-depth interpretation of point-light walkers. Despite the instruction to ignore the sound and to only report the visually perceived in-depth orientation, observers were unable to do so. It may thus be possible that, in accordance with recent studies documenting on changes in the phenomenological experience of biological motion stimuli when a congruent sound is available (Arrighi et al., 2009; Brooks et al., 2007; Saygin et al., 2008; van der Zwan et al., 2009), here, in the looming and receding conditions, audiovisual integration led to phenomenological different perceptual experiences than in the stationary and the no-sound condition. After completion of the experiment but before debriefing, participants were asked some questions. One of the questions was to estimate how many *visually distinguishable* point-light walkers were presented during the experiment. We expected participants to be able to answer this question quite accurately, because only two easily

distinguishable types of point-light walkers were used. Indeed, all participants noticed there were at least two different point-light walkers. However, only 21% of the participants (4/19) gave the correct response (two) whereas the majority, 63% of the participants (12/19), indicated that they had seen three or four different point-light walkers (spontaneous responses were 'three', 'four', or 'three or four'). The rest of the observers (16% or 3/19) thought there were even more than four visually distinguishable point-light walkers. Although memory reports should always be interpreted with care, the observation that 79% of the participants spontaneously responded to have seen more than two visually distinguishable point-light walkers, may suggest that also in the current experiment the simultaneous presentation of sound led to phenomenological different perceptual experiences of the point-light walkers than when they were presented without sound.

Definitely, future research should concentrate on tasks that are more informative on the specific level at which these effects occur. Very recently, two studies were reported that may inspire the design of such tasks. Jackson and Blake (2010) demonstrated adaptation after-effects in the in-depth perception of point-light walkers: Observers who adapted to a point-light walker that was unambiguously facing the viewer (stereo depth cues) were more likely to perceive a depth-ambiguous test walker as facing away from the viewer, and vice versa. van der Zwan et al. (2009) used an audiovisual adaptation after-effect paradigm for the perception of gender from point-light walkers. Adaptation to a gender neutral point-light walker accompanied by female footstep sounds elicited male after-effects. Likewise, adaptation to a depth-ambiguous point-light walker accompanied by looming or receding sounds might elicit after-effects in the perceived in-depth orientation. Designing such a task is not trivial, however. For rising and falling tones to signal convincingly a looming or receding sound source, they need to be relatively short (4.5 seconds in the current experiment) while for adaptation after-effects in the in-depth perception, Jackson and Blake (2010) used adaptation times of minimum 45 seconds. It thus remains an empirical challenge to find out whether the present effects act at a perceptual or at a more cognitive level.

Concluding remarks and future research

Part I of this dissertation explored some factors that affect the perceived in-depth orientation of depth-ambiguous point-light walkers. Part II concentrated on the audiovisual perception of biological motion. To finalize this dissertation, here, for each part separately, we will summarize our findings, formulate conclusions, and provide some suggestions for future research.

PART I

To test whether behavioral or social relevance is an important variable in causing the facing bias for depth-ambiguous point-light walkers (Vanrie et al. 2004), Brooks et al. (2008; Chapter 3) explored whether the gender of the point-light figure had an effect on the facing bias. A strong effect of figure gender on the facing bias was observed: The facing bias was only found for male and gender neutral point-light figures. Extreme female figures were perceived as facing away.

In addition to a (weaker) figure gender effect, Schouten et al. (2010; Chapter 4) found an observer sex effect: The figure gender effect was found to interact with observer sex. More specifically, the facing bias for male point-light walkers was stronger for male observers than for female observers. While the observations of Brooks et al. (2008) and Schouten et al. (2010) are consistent with the account assuming that the gender that is perceived from a point-light walker biases its perceived in-depth orientation, they do not provide conclusive evidence for it. Indeed, crucial for the validity of such an account is the demonstration of a causal link between perceived gender and perceived in-depth orientation. It could be possible, for instance, that stimulus properties, regardless of the perceived gender they elicit, are responsible for biases in the perceived in-depth orientation.

Schouten, Troje, and Verfaillie (in press; Chapter 5) tested whether or not the relation between perceived gender and perceived in-depth orientation is causal. The results revealed that the relation between perceived gender and perceived in-depth orientation is not causal. Structural and kinematic stimulus changes that induced comparable changes in perceived gender were found to elicit opposite changes in perceived in-depth orientation. Thus, stimulus properties, irrespective of the perceived gender they elicit, were found to

play a crucial role in biasing the in-depth perception of depth-ambiguous point-light walkers.

Moreover, the results of Schouten et al. (in press) indicated that, to judge the in-depth orientation, observers preferably rely on the lower part of the point-light figure, and that the facing bias is very likely to arise from kinematic information in the lower part of the figure. As particular kinematic properties in the lower part of a point-light figure are also suspected to be involved in triggering the so-called 'life detector' (Chang & Troje, 2008, 2009a, 2009b; Saunders et al., 2009; Troje & Westhoff, 2006) it was speculated that the facing bias might be somehow linked to the life detector mechanism. Future research could further test this speculation by explicitly manipulating the kinematics of the point-light figure. For instance, it would be interesting to test whether the removal of accelerations (and other even higher order motions) from (the lower part of) the point-light figure, would diminish or even completely discard the facing bias. An alternative could be to record proportions of FTV responses for point-light walkers under limited-lifetime conditions. Shortening the lifetime of the dots, for example, might disrupt the local motion information that is necessary to trigger the life detector or the facing bias.

However, it would be too far of a stretch to claim that a mechanism like the life detector, yet still poorly understood and vaguely defined, is the only cause of biases in the perceived in-depth orientation. Indeed, contrary to the findings that the lower part of the point-light figure elicits the facing bias, in Schouten et al. (in press) the upper part of the point-light figure was found to elicit the facing away interpretation. At first sight, these findings were at odds with what was reported by Vanrie and Verfaillie (2006) who suspected the upper part of a point-light figure to be important for the facing bias. Ironically, however, this apparent contradiction pointed our attention to another process that potentially is of importance in the in-depth organisation of point-light figures, namely 'occlusion violation'.

As mentioned in Chapter 1, Vanrie et al. (2004) demonstrated that the presence of occlusion effectively disambiguates the in-depth interpretation of a point-light figure. Likewise, in situations in which occlusion could be expected only for one of two possible veridical interpretations (e.g., facing the viewer), the absence of occlusion might be a cue to bias the in-depth percept in the direction of the other interpretation (e.g., facing away) for which no occlusion would be expected. This might thus be another explanation why point-light walkers, or parts of it, that in principle contain no information as to their in-depth orientation, can still elicit a perceptual bias. It remains a challenge for future research

to find out whether indeed the absence of expected occlusion can bias the in-depth perception of point-light figures.

Quite some challenges in the research on the in-depth perception of point-light figures are still waiting to be addressed. To enable this research, Schouten and Verfaillie (2010; Chapter 2) provided a tool, the perspective technique. By gradually manipulating perspective cues, it is possible to obtain experimental control on the perceived in-depth orientation and to determine the point of subjective ambiguity.

In Chapter 6, the perspective technique was further refined and it was shown that merely the relative position and motion changes of the dots of the point-light figure are sufficient to elicit strong changes in perceived in-depth orientation. This refined version was employed to verify whether the observer sex effect, as reported in Schouten et al. (2010; Chapter 4), could also be observed when the change in perceived in-depth orientation was not elicited by changes in figure gender, but was induced by perspective cues instead. Results of this experiment indeed revealed an observer sex effect. The strength of the relation between perceived in-depth orientation and changes in perspective information differed between male and female observers. Thus, although their causes remain unclear, and the way in which they show up is sometimes inconsistent (Schouten et al., in press; Chapter 5), observer sex effects are often observed. Therefore, we think it is safe to conclude that, when it concerns judging the in-depth orientation of point-light figures, male observers respond quite differently to the same stimuli than female observers. Whether this intriguing difference in response behavior reflects a sex difference at the level of perceptual sensitivity to a particular kind of depth information, a different perceptual bias for males and females or rather an observer sex difference in decisional processes remains an important question for future research. We would like to address this question in a 2AFC task soon (see below).

As briefly mentioned in Chapter 6, one potential weakness of our research on the in-depth perception of point-light figures might be the task that is employed. Indeed, in all of our experiments, we used a naming task. That is, we simply asked observers to give a subjective binary response (‘facing the viewer?’ or ‘facing away?’). Proportions of ‘facing the viewer’ (FTV) responses were then assumed to reflect the observers’ perceptions. For instance, when proportions of FTV responses deviated from 0.50 we interpreted this as a perceptual bias. However, it remains unclear to what extent more cognitive decisional processes are involved. Future research on the in-depth perception would thus benefit from paradigms in which perceptual biases can be distinguished from response biases. The

perspective technique that we developed might aid in this, because it provides the opportunity to present point-light figures that are, as a result of an added physical property such as perspective, more consistent with one of both veridical in-depth interpretations. In other words, instead of presenting depth-ambiguous point-light figures for which there is, in fact, no unique correct response, now point-light figures can be presented that differ from each other on a known and controllable physical dimension. In combination with, for example, a 2AFC discrimination task, the perspective technique thus allows to measure discrimination thresholds.

Other paradigms might help as well to disentangle perceptual biases from response biases. For instance, it is possible to present to observers a point-light walker rotating around its vertical axis. Instead of asking to observers to report the perceived in-depth orientation, observers can be asked to indicate the perceived direction of rotation: 'Is the figure rotating clockwise (CW) or 'counterclockwise (CCW)?' (see Troje, 2010; also see Dekeyser, Vanrie, & Verfaillie, 1999; Jackson, Cummins, & Brady, 2008). Because each orientation of the point-light figure is consistent with two veridical interpretations, the perceived direction of rotation of a depth-ambiguous point-light walker is directly linked to the perceived in-depth orientation. Indeed, the 2D projection of a point-light figure who is facing the viewer and who is rotating clockwise is identical to a point-light figure who is facing away and rotating counterclockwise. Now, when a slight perspective cue is added, so that the true orientation is known, and when in half of the cases the figures rotate clockwise and in the other half of the cases the figures rotate counterclockwise, an unbiased measure of the proportion of FTV responses can be obtained. With this method, the perceived in-depth orientation can thus directly be derived from the perceived direction-of-rotation responses.

Many variables other than the ones reported and examined in the present dissertation might affect the perceived in-depth orientation of depth-ambiguous point-light figures. Variables such as the speed of walking, the presentation duration, the viewing angle, the aspect-ratio of the point-light figure, the ratio between horizontal and vertical motion, etc could be of influence. Note also that our findings showing that stimulus properties strongly affect the perceived in-depth orientation of point-light figures, does not exclude the possibility that more contextual or higher-level factors can be of influence on the process of depth assignment. Indeed, variables at different processing levels could influence the in-depth organisation (Vanrie, 2005).

It is a stimulating observation that, using our perspective technique, further exploration of other variables affecting the perceived in-depth orientation of point-light figures has already taken its first steps. First, for instance, one of our colleagues in the lab is currently investigating whether primes featuring a high emotional value, such as unpleasant pictures from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008), elicit a different perceived in-depth orientation compared to 'neutral' pictures from the same database.

Second, as it is known that motor process can affect the perception of biological motion (e.g., Bidet-Ildei & Coello, 2010; Casile & Giese, 2006; see Chapter 1), motor processes could also be of influence on the in-depth perception of point-light figures. In collaboration with Dr. Valeria Manera (University of Turin, Italy), we are currently setting up a study in which we investigate whether, compared to for example standing, the action of walking (e.g., on a treadmill) affects the point of subjective depth-ambiguity.

Third, also outside our lab, researchers are starting to investigate variables that affect the perceived in-depth orientation of point-light figures. Recently, Thornton and Garvey (2010) flanked a centrally presented point-light walker, whose in-depth orientation was manipulated using the perspective technique, on either side by a computer-animated flow-field. The flanking scene simulated camera movement in depth, either forward (expanding flow) or backwards (contracting flow) along a city street. Thornton and Garvey (2010) found a reliable decrease in FTV responses when flanking motion was consistent with movement into the scene, suggesting that scene context may be integrated before the assignment of figure depth.

In summary, as evidenced by the relatively small amount of literature that is currently available, and by the many questions that remain, the in-depth perception of point-light biological motion is a very young research domain. In the present dissertation, we explored only some of the many factors that might affect the perceived in-depth orientation of depth-ambiguous point-light walkers. We hope that our tool, the results presented in this dissertation, and our suggestions for future research might aid and inspire new work in this domain.

PART II

In Part II of this dissertation we investigated the audiovisual perception of biological motion. In Chapter 8 we presented an adjustment task to estimate the PSS. This

task, in which observers adjust the phase of the auditory stream until it subjectively matches the visual stream, seems to be well suited to estimate the PSS for audiovisual rhythms as, for example, a point-light walker paired with rhythmic footsteps.

In Chapter 9 we tested whether sound can facilitate the detection of a point-light figure. In two experiments we found no effect of sound at all. In sharp contrast, Arrighi, Marini, and Burr (2009) reported audiovisual enhancement that was even stronger than statistical optimal integration with a tap dance point-light action as a visual stimulus. As already mentioned in the General Discussion of Chapter 9, it remains unclear what exactly causes this discrepancy in results. Yet, a closer examination of the particular features of the stimuli that were used in these studies could be a first step in understanding and identifying necessary conditions to observe audiovisual interaction effects.

What seems *not* to be a necessary condition to observe strong audiovisual enhancement in the perception of biological motion, as shown by Arrighi et al. (2009), is the presence of an underlying human body structure. Indeed, the visual stimulus used by Arrighi et al. (2009) was not a classical point-light action in which the entire body is represented by point-lights. Instead, their point-light stimulus consisted of six dots that made up two rigid structures consisting of three dots each, representing merely the motions of the tap dancer's feet. In contrast, at first sight, our results (Chapter 9) and those of other studies (Petrini, Russel, & Pollick, 2009; Saygin, Driver, & De Sa, 2008) seem to suggest that degrading the structure of the point-light figure decreases audiovisual performance. First, performance in the SJ task (Experiment 2, Chapter 9) was quite poor when only a limited number of signal dots was presented. Second, Petrini, Russel, and Pollick (2009; also see Chapter 7) found that audiovisual synchrony judgements in novices decreased when particular dots of the point-light display were removed. Third, Saygin, Driver, and De Sa, (2008; also see Chapter 7) showed that when the upright gestalt of a point-light walker is destroyed (by scrambling or inverting the stimulus) audiovisual performance deteriorates.

However, closer examination of these examples of stimulus degradation points to a potentially relevant condition, other than the presence of an underlying human structure, that is affected by the stimulus degradation, namely, the presence of a clear visual event to which the sound can be bound (e.g., a visual impact or direction change). Indeed, first, in our experiments (Chapter 9), the limited lifetime technique may have destroyed the local motions of the foot dots that contained the impacts or direction changes to which the sound could be bound. Second, in Petrini, Russel, and Pollick (2009) the decreased performance

in novices was due to the removal of the drum stick point that signalled the visual impact. Third, inverting and scrambling the display in Saygin et al. (2008) may have deprived observers of the clear visual event (footstep) to which the sound should be bound. Crucial in the study of Saygin et al (2008) might be the fact that they used the Cutting algorithm to generate the point-light figure. Indeed, as shown earlier (Chang & Troje, 2008, 2009a, 2009b; Troje & Westhoff, 2006) even under scrambled conditions the feet of a point-light walker attract attention, but only when they feature the local invariants that are necessary to trigger the life detector. As suggested by Saunders, Suchan, and Troje (2009) these particular local invariants are not (or less) present in point-light walkers generated by the Cutting algorithm[18]. The effect found by Saygin et al. (2008) might thus be due to the fact that when observers are confronted with an upright walker it is clear to what the sound should be bound, namely the motions of the feet, which can be easily identified because of the structure of the point-light figure. When the walker is inverted (cfr. the body inversion effect; Reed, Stone, Bozova, & Tanaka, 2003; Reed, Stone, Grubb, & McGoldrick, 2006; also see Chapter 1) or scrambled, there is no more obvious and unique visual event to which the sound can be bound. Taken together, instead of the structure, the presence of a unique visual marker to which the sound can be bound seems to be a crucial condition for audiovisual binding to occur.

The fact that an underlying human structure or even the presence of biological motion is not a necessary condition to find strong audiovisual enhancement was recently also demonstrated by Van der Burg and colleagues (see Chapter 7). Van der Burg, Olivers, Bronkhorst, and Theeuwes (2008) found that an abrupt and non-spatial auditory pulse, synchronised with an abrupt color change of a visual target strongly facilitated the speed of target identification, independently of the number of distractor elements. Van der Burg, Cass, Olivers, Theeuwes, and Alais (2010) showed that a crucial condition to observe this 'pip pop' effect is that the synchronized audiovisual stimuli are transient, i.e. abrupt. According to Van der Burg et al. (2010) transient signals are necessary in synchrony driven binding to avoid spurious interactions with unrelated signals when these occur close together in time.

The question remains to what extent the results of Arrighi, Alais, and Burr (2009) differ from those of Van der Burg et al. (2008, 2010) or to what extent they are the result

[18] Assuming that the life detector indeed attracts attention this would imply that, if a motion-captured point-light walker was to be used in a task similar as the one employed by Saygin et al. (2009), also its scrambled version would lead to reasonable audiovisual performance, because the foot dots can still be easily identified.

of the same underlying mechanism. Or in other words, to what extent are the effects reported by Arrighi et al. (2009) specific to biological motion? Arrighi et al. (2009) themselves remain inconclusive on this issue. Although performance in the tasks of Arrighi et al. (2009) and Van der Burg et al. (2008) cannot be directly compared, because the tasks differ on several aspects such as, for example, the dependent variable (reaction times in a visual search task in Van der Burg et al., 2008 vs. detection thresholds in Arrighi et al., 2009), both tasks might require comparable perceptual processes. Indeed, there is some interesting overlap in neural substrates that, according to each of both groups of authors, may be involved in their task.

Arrighi et al. (2009) and Van der Burg et al. (2010) both refer to the Superior Temporal Sulcus (STS; see Chapter 1 and Chapter 7). As pointed out by Arrighi et al. (2009), STS shows particularly interesting complex audiovisual interactions. For instance, it responds to a stimulus such as an object striking a surface, but neither the image nor the sound of the event alone will elicit a response (Bruce, Desimone, & Gross, 1981). These kinds of 'purposeful audiovisual stimuli' seem to have similar characteristics to the tap dance routines. Van der Burg et al. (2010) also point to the STS as a likely substrate for their findings. According to these authors an auditory signal could, through feedback projections from STS, affect perceptual processing of a synchronized visual event at an early level of processing (see Calvert, 2001; Calvert & Thesen, 2004; Lewis & Noppeney, 2010). As a result, the synchronized visual event is perceived as being brighter and therefore pops out from its environment (Van der Burg et al., 2010).

Both groups of authors also mention the Superior Colliculus (SC; see Chapter 1 and Chapter 7). Yet, they remain doubtful about its involvement. According to Arrighi et al. (2009), SC does not seem to be directly involved with the perception of biological motion but it could be instrumental in the integration of auditory and visual signals prior to the further analysis of biological motion (also see Troje & Chang, in press; Chapter 1). As mentioned by Van der Burg et al. (2010), strong activation in SC is only observed when the signals from the different senses arrive from overlapping spatial locations. If the two signals are not spatially collocated activation in SC is greatly reduced or even inhibited (Meredith & Stein, 1996). Therefore, according to Van der Burg et al. (2010), SC is unlikely to be the underlying mechanism of the pip pop effect because the temporally aligned multisensory events in their study never shared the same spatial location.

The overlap in neural correlates that are assumed to be underlying the perceptual performance by Arrighi et al. (2009) and Van der Burg et al. (2010) in their tasks may thus

suggest that similar processes are at play. However, there is one neural structure that is mentioned by Arrighi et al. (2009) that is not mentioned by Van der Burg et al. (2010). As stated by Arrighi et al. (2009), the cortical structures most likely to be implicated in the recognition of tap dance patterns are those population of multisensory neurons specifically tuned for the perception (and/or production) of specific action patterns, namely audiovisual mirror neurons (Keysers et al., 2003; Kohler et al., 2002; Chapter 7). These neurons seem to be implicated in the recognition and interpretation of actions by matching the auditory and visual patterns of external actions onto the observer's internal motor coordinates. The response profiles of these neurons are strictly tuned to specific action patterns (Keysers et al., 2003). As evidence about the existence of cross-modal mirror neurons has been also collected in humans (Gazzola, Aziz-Zadeh, & Keysers, 2006) these cortical mechanisms could underlie the integration of auditory and visual information in the recognition of human actions, and thus play a key role in the perception of audiovisual motion profiles as tap dance routines (Arrighi et al., 2009).

Although, the existence of specifically tuned neural structures might suggest that the effect found by Arrighi et al. (2009) is unique for biological motion, so far there is no clear evidence that the strong integration observed by these authors occurred because the stimuli contained the motions generated by a human actor. Indeed, Arrighi et al. (2009) did not include a control condition with non-biological motion.

Hence, the question whether audiovisual integration for biological motion is 'special' remains. Whether this question can be solved strongly depends on whether biological motion is considered as 'special' and if so, whether it can be clearly defined what it precisely is that makes it so special. So far, however, a clear definition of 'biological motion' is lacking. Is it the motion that is generated by a living organism? Is it the presence of a structured point-light figure that mimics the motions of a human action such as a point-light walker generated by the Cutting algorithm? Is it the presence of motion-mediated structure that triggers the mirror neuron system? Or is it the presence of local invariants, accelerations, and decelerations that originate from the feet of terrestrial animals? Once it is defined what biological motion is, the challenge remains to find a fair control condition (Hiris, 2007). For instance, it would be interesting to compare performance in Arrighi et al.'s task with performance in a task in which the same display is inverted or in which the motions are generated by a computer program.

What can be concluded from our experiments in Chapter 9 is that sound does not help to reconstruct the human structure in point-light figures in which this structure is

degraded by masking or by the limitation of the number of dots. Taken together, kinematic (e.g., direction changes, sudden impacts) more than structural properties (e.g., the presence of a human body structure) seem to be important in audiovisual integration of biological motion stimuli. However, recognizing a meaningful human structure might aid audiovisual integration when it helps to identify relevant parts that can be bound to the sound (Saygin et al., 2008).

Sound seems to bind well to small rigid structures that feature the same temporal pattern as the auditory stream (Arrighi, Alais, & Burr, 2006; Arrighi et al., 2009; Fujisaki & Nishida, 2007; Guttman, Gilroy, & Blake, 2005; Van der Burg et al., 2008, 2010). In analogy to the attention mediated visual routines, or 'sprites' as proposed by Cavanagh, Labianca, and Thornton (2001) it could be interesting to speculate on the existence of 'audiovisual sprites'. Sprites are considered to be the set of routines that is responsible for detecting the presence of a specific characteristic motion in the input array, for modeling or animating the object's changing configuration as it makes this stereotypical motion, and for filling in the predictable details of the motion over time and in the face of noisy or absent image details (Cavanagh et al., 2001). Audiovisual sprites could be, for example, a set of routines responsible for detecting the presence of characteristic audiovisual patterns such as the simultaneous presence of a sound and a visual direction change (e.g., a bouncing object or a footstep). We further speculate that the STS might be a likely candidate to support such high-level audiovisual representations and that acquiring these representations could happen through repetitive experience with characteristic audiovisual events (Petrini, Holt, & Pollick, 2010; Petrini, Russel, & Pollick, 2009). Sound might help to recognize, or to direct attention to, a part of a stimulus that is specifically involved in the audiovisual interaction (e.g., an arm, leg, foot, etc.), which on its turn aids in the further analysis and the perceptual reconstruction of a meaningful human action (also see Neri, 2009). It is possible that audiovisual mirror neurons enrich these representations for audiovisual events that also could have been generated by the observer her or himself. Alternatively, or in addition, when invariants are present in particular stimulus parts that trigger the life detector, attention could be immediately directed to these aspects of the stimulus (Troje & Chang, in press) boosting the further audiovisual processing of these parts and, if necessary, the further perceptual reconstruction of the action.

In Chapter 10 we tested whether the visual and auditory information on the speed of a point-light walker is integrated in a statistically optimal way. Our data did not exclude statistical optimal integration, neither did they provide conclusive evidence for it. To

optimize this study, auditory and visual performance should be normalized. That is, ideally, the variance (and thus the reliability) of the unimodal perceptual estimates are matched to be more or less equal (see e.g., Arrighi et al., 2009; Roach, Heron, & McGraw, 2006). Love et al. (2007) reported a speed discrimination task with an audiovisual point-light drummer in which the unimodal variances were matched. The setup of this study was identical to our study. Interestingly, they reported a similar problem as the one that we encountered: Speed discrimination based on auditory cues alone is too good, leading to the ignorance of the visual cue in audiovisual presentations. To tackle this problem these authors added five different levels of Brown Noise (higher representation of lower frequencies than white noise) to the auditory stimuli to explicitly manipulate auditory performance. They only found statistical optimal integration for one out of five auditory noise levels, when the auditory performance was slightly worse than the visual performance. In the rest of the conditions audiovisual performance was similar or even worse than the best of the unimodal performances. This may suggest that the weight given to the auditory signal in such a task is relatively stable and is not proportional to its current reliability. For some perceptual estimations our system might thus be so predetermined to give a higher prior weight to one modality, because it is generally the most reliable for that type of information, that combining the information with information entering in another modality fails or even worsens performance.

Finally, bridging the research of Part I and Part II, in Chapter 11 we verified whether looming and receding sounds affect the perceived in-depth orientation of depth-ambiguous point-light figures. Results showed that looming sounds led to an increase and receding sounds to a decrease in the number of FTV responses, compared to when no sound or a stationary sound was available. Although this is an interesting observation on the audiovisual perception of biological motion, further research is necessary to test at what level these audiovisual interactions occur and whether these interactions are specific to biological motion. Indeed, the data reported in Chapter 11 do not exclude the possibility that the audiovisual interaction is due to more decisional rather than perceptual processes or that the interaction can also be found for non-biological motion.

To proceed the research presented in Chapter 11, our perspective technique might help to design a task in which perceptual performance is measured. For instance, in a pretest for a point-light walker proportions of FTV responses as a function of field of view angle could be measured so that, for example the 30%, 40%, 50%, 60% and 70% threshold can be derived. The 50% threshold is the point of subjective ambiguity (PSA) and the

point-light walker that matches this threshold could be taken as the reference stimulus. Point-light walkers that match the other thresholds could be taken as test stimuli. In a temporal 2AFC task then, on each trial observers are asked to compare the test stimulus (one interval) with the reference stimulus (other interval) and to indicate the interval of the point-light walker that is perceived as most convincingly facing the viewer (or least convincingly facing away). On each trial, observers receive feedback. Now, compared to when no sound is available, if looming and receding sounds are paired with the test stimuli it could be verified, for example, whether task performance increases/decreases when walkers matching the 60% and 70% thresholds are paired with a looming/receding sound or walkers matching the 30% and 40% thresholds are paired with a receding/looming sound. If sound indeed affects the perceived in-depth orientation at a perceptual level and in the direction suggested by the sound then performance in this 2AFC task should increase when the sound is congruent and decrease when the sound is incongruent with the test stimulus.

However, finding a fair control condition with non-biological motion is not trivial. Inverting the display would be the most obvious manipulation but a pilot experiment in our lab has already shown that with inverted displays observers perform very poorly when they have to judge the in-depth orientation.

In summary, the in-depth perception of point-light figures on which we focussed in Part I as well as the audiovisual perception of biological motion on which we concentred in Part II, are young research domains. So far, psychophysical evidence showing that audiovisual integration for biological motion is special or different from other kinds of audiovisual integration is sparse and not convincing.

It is clear that the perception of biological motion is not a single phenomenon but consists of a range of different aspects that have to be distinguished carefully, both conceptually and experimentally. The design of reasonable ‘non-biological motion’ control conditions that can be contrasted with ‘biological motion’ conditions remains an important challenge for future research. In fact, the validity of results of experiments that contrast the visual or the audiovisual perception of biological motion with non-biological-motion stimuli directly depends on how well researchers are able to gain control over those stimulus features that are hypothesized to be unique to biological motion.

References

Ahlström, V., Blake, R., & Ahlström, U. (1997). Perception of biological motion. *Perception, 26*(12), 1539-1548. doi: 10.1068/p261539

Akay, A., & Celebi, G. (2009). Depth perception differences between hemispheres, genders, and eye dominance A study on evoked potential. *Neural Regeneration Research, 4*(11), 947-950. doi: 10.3969/j.issn.1673-5374.2009.11.026

Alaerts, K., Swinnen, S. P., & Wenderoth, N. (2009). Interaction of sound and sight during action perception: Evidence for shared modality-dependent action representations. *Neuropsychologia, 47*(12), 2593-2599. doi: 10.1016/j.neuropsychologia.2009.05.006

Alais, D., & Burr, D. (2004). No direction-specific bimodal facilitation for audiovisual motion detection. *Cognitive Brain Research, 19*(2), 185-194. doi: 10.1016/j.cogbrainres.2003.11.011

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences, 4*(7), 267-278. doi: 10.1016/S1364-6613%2800%2901501-1

Arnold, D. H., Tear, M., Schindel, R., & Roseboom, W. (2010). Audio-visual speech cue combination. *Plos One, 5*(4), e10217. doi: 10.1371/journal.pone.0010217

Arrighi, R., Alais, D., & Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of Vision, 6*(3), 260-268. doi: 10.1167/6.3.6

Arrighi, R., Marini, F., & Burr, D. (2009). Meaningful auditory information enhances perception of visual biological motion. *Journal of Vision, 9*(4). doi: 10.1167/9.4.25

Aschersleben, G., & Bertelson, P. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension. 2. Evidence from sensory-motor synchronization. *International Journal of Psychophysiology, 50*, 157-163. doi: 10.1016/S0167-8760%2803%2900131-4

Azzalini, A., & DallaValle, A. (1996). The multivariate skew-normal distribution. *Biometrika, 83*(4), 715-726. doi: 10.1093/biomet/83.4.715

Bach, D. R., Schachinger, H., Neuhoff, J. G., Esposito, F., Di Salle, F., Lehmann, C., et al. (2008). Rising sound intensity: An intrinsic warning cue activating the amygdala. *Cerebral Cortex, 18*(1), 145-150. doi: 10.1093/cercor/bhm040

Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and spatial factors in gait perception that influence gender recognition. *Perception & Psychophysics, 23*(2), 145-152.

Barraclough, N. E., Xiao, D., Baker, C. I., Oram, M. W., & Perrett, D. I. (2005). Integration of visual and auditory information by STS neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience, 17*, 377-391. doi: 10.1162/0898929053279586

Battelli, L., Cavanagh, P., & Thornton, I. M. (2003). Perception of biological motion in parietal patients. *Neuropsychologia, 41*(13), 1808-1816. doi: 10.1016/S0028-3932%2803%2900182-9

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron, 41*, 809-823. doi: 10.1016/S0896-6273(04)00070-4

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron, 34*, 149-159. doi: 10.1016/S0896-6273(02)00642-6

Beintema, J. A., & Lappe, M. (2002). Perception of biological motion without local image motion. *Proceedings of the National Academy of Sciences, 99*(8), 5661-5663. doi: 10.1073/pnas.08248369

Bertelson, P. (1999). Ventriloquism: A case of crossmodal perceptual grouping. In G. Aschersleben, T. Bachmann & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 347-362). Amsterdam: Elsevier, North-Holland.

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin, 5*(3), 482-489.

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension. 1. Evidence from auditory-visual temporal order judgement. *International Journal of Psychophysiology, 50*, 147-155. doi: 10.1016/S0167-8760%2803%2900130-2

Bertelson, P., Pavani, F., Ladavas, E., Vroomen, J., & de Gelder, B. (2000). Ventriloquism in patients with unilateral visual neglect. *Neuropsychologia, 38*(12), 1634-1642. doi: 10.1016/S0028-3932%2800%2900067-1

Bertenthal, B. I., & Pinto, J. (1994). Global processing of biological motions. *Psychological Science, 5*, 221-225. doi: 10.1111/j.1467-9280.1994.tb00504.x

Bidet-Caulet, A., Voisin, J., Bertrand, O., & Fonlupt, P. (2005). Listening to a walking human activates the temporal biological motion area. *Neuroimage, 28*(1), 132-139. doi: 10.1016/j.neuroimage.2005.06.018

Bidet-Ildei, C., Chauvin, A., & Coello, Y. (2010). Observing or producing a motor action improves later perception of biological motion: Evidence for a gender effect. *Acta Psychologica, 134*(2), 215-224. doi: 10.1016/j.actpsy.2010.02.002

Blake, R. (1993). Cats perceive biological motion. *Psychological Science, 4*, 54-57. doi: 10.1111/j.1467-9280.1993.tb00557.x

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology, 58*, 47-73. doi: 10.1146/annurev.psych.57.102904.190152

Bonda, E., Petrides, M., Ostry, D., & Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *Journal of Neuroscience, 16*(11), 3737-3744.

Brooks, A., Schouten, B., Troje, N. F., Verfaillie, K., Blanke, O., & van der Zwan, R. (2008). Correlated changes in perceptions of the gender and orientation of ambiguous biological motion figures. *Current Biology, 18*(17), R728-R729. doi: 10.1016/j.cub.2008.06.054

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, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology, 46*(2), 369-384.

Buccino, G., Binkofski, F., & Riggio, L. (2004). The mirror neuron system and action recognition. *Brain and Language, 89*(2), 370-376. doi: 10.1016/s0093-934x(03)00356-0

Bülthoff, I., Bülthoff, H., & Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nature Neuroscience, 3*, 254-257. doi: 10.1038/699

Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral Cortex, 11*, 1110-1123.

Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D., & David, A. S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *NeuroReport, 10*, 2619-2623. doi: 10.1097/00001756-199908200-00033

Calvert, G. A., Brammer, M. J., & Iversen, S. D. (1998). Crossmodal identification. *Trends in Cognitive Sciences, 2*(7), 247-253.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., et al. (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*, 649-657. doi: 10.1016/S0960-9822(00)00513-3

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2000). Crossmodal integration of non-speech audio-visual stimuli. *NeuroImage, 11*(5), 727. doi: 10.1016/S1053-8119(00)91657-4

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audiovisual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage, 14*, 427-438. doi: 10.1006/nimg.2001.0812

Calvert, G. A., Spence, C., & Stein, B. E. (2004). *The handbook of multisensory processes*. Cambridge: MIT Press.

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: Methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris, 98*(1-3), 191-205. doi: 10.1016/j.jphysparis.2004.03.018

Casile, A., & Giese, M. A. (2005). Critical features for the recognition of biological motion. *Journal of Vision, 5*(4), 348-360. doi: 10.1167/5.4.6

Casile, A., & Giese, M. A. (2006). Nonvisual motor training influences biological motion perception. *Current Biology, 16*(1), 69-74. doi: 10.1016/j.cub.2005.10.071

Cavanagh, P., Labianca, A. T., & Thornton, I. M. (2001). Attention-based visual routines: sprites. *Cognition, 80*, 47-60.

Chandrasekaran, C., Turner, L., Bulthoff, H. H., & Thornton, I. M. (2010). Attentional networks and biological Motion. *Psihologija, 43*(1), 5-20. doi: 10.2298/psi1001005c

Chang, D. H. F., & Troje, N. F. (2008). Perception of animacy and direction from local biological motion signals. *Journal of Vision, 8*(5). doi: 10.1167/8.5.3

Chang, D. H. F., & Troje, N. F. (2009a). Acceleration carries the local inversion effect in biological motion perception. *Journal of Vision, 9*(1). doi: 10.1167/9.1.19

Chang, D. H. F., & Troje, N. F. (2009b). Characterizing global and local mechanisms in biological motion perception. *Journal of Vision, 9*(5). doi: 10.1167/9.5.8

Colonius, H., & Diederich, A. (2004). Why aren't all deep superior colliculus neurons multisensory? A bayes' ratio analyses. *Cognitive, Affective, and Behavioral Neuroscience, 4*(3), 344-353. doi: 10.3758/CABN.4.3.344

Cooke-Simpson, A., & Voyer, D. (2007). Confidence and gender differences on the mental rotations test. *Learning and Individual Differences, 17*(2), 181-186. doi: 10.1016/j.lindif.2007.03.009

Cutting, J. E. (1978a). Generation of synthetic male and female walkers through manipulation of a biomechanical invariant. *Perception, 7*(4), 393-405. doi: 10.1068/p070393

Cutting, J. E. (1978b). Program to generate synthetic walkers as dynamic point-light displays. *Behavior Research Methods & Instrumentation, 10*(1), 91-94.

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk - Gait perception without familiarity cues. *Bulletin of the Psychonomic Society, 9*(5), 353-356.

Cutting, J. E., Moore, C., & Morrison, R. (1988). Masking the motions of human gait. *Perception & Psychophysics, 44*, 339-347.

de Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences, 7*(10), 460-467. doi: 10.1016/j.tics.2003.08.014

Dekeyser, M., Vanrie, J., & Verfaillie, K. (1999, September). *Exploring the oscillation effect with point-light figures*. Poster presented at the XIth Conference of the European Society for Cognitive Psychology. Gent, Belgium.

Dekeyser, M., Verfaillie, K., & Vanrie, J. (2002). Creating stimuli for the study of biological-motion perception. *Behavior Research Methods, 34*, 375-382.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General, 115*, 107-117. doi: 10.1037/0096-3445.115.2.107

Diederich, A., & Colonius, H. (2004). Modeling the time course of multisensory interaction in manual and saccadic responses. In G. Calvert, C. Spence & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 395-408). Cambridge: MIT Press.

Dittrich, W. H. (1993). Action categories and the perception of biological motion. *Perception, 22*, 15-22. doi: 10.1068/p220015

Dittrich, W. H., Troscianko, T., Lea, S. E. G., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception, 25*, 727-738. doi: 10.1068/p250727

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9*(6), 719-721. doi: 10.1068/p090719

Downing, P. E., Jiang, Y. H., 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

Dufour, A., Touzalin, P., Moessinger, M., Brochard, R., & Despres, O. (2008). Visual motion disambiguation by a subliminal sound. *Consciousness and Cognition, 17*(3), 790-797. doi: 10.1016/j.concog.2007.09.001

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429-433. doi: 10.1038/415429a

Ernst, M. O., & Bulthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*(4), 162-169. doi: 10.1016/j.tics.2004.02.002

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics, 63*(4), 719-725.

Freeman, E., & Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound. *Current Biology, 18*(16), 1262-1266. doi: 10.1016/j.cub.2008.07.066

Fujisaki, W., Koene, A., Arnold, D., Johnston, A., & Nishida, S. (2006). Visual search for a target changing in synchrony with an auditory signal. *Proceedings of the Royal Society B-Biological Sciences, 273*(1588), 865-874. doi: 10.1098/rspb.2005.3327

Fujisaki, W., & Nishida, S. (2005). Temporal frequency characteristics of synchrony-asynchrony discrimination of audio-visual signals. *Experimental Brain Research, 166*(3-4), 455-464. doi: 10.1007/s00221-005-2385-8

Fujisaki, W., & Nishida, S. (2007). Feature-based processing of audio-visual synchrony perception revealed by random pulse trains. *Vision Research, 47*(8), 1075-1093. doi: 10.1016/j.visres.2007.01.021

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn Sci, 8*(9), 396-403. doi: 10.1016/j.tics.2004.07.002

Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Current Biology, 16*(18), 1824-1829. doi: 10.1016/j.cub.2006.07.072

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*(5), 473-490. doi: 10.1162/089892999563544

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Miffin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. London: Erlbaum.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience, 4*, 179-192. doi: 10.1038/nrn1057

Grossman, E. D., Battelli, L., & Pascual-Leone, A. (2005). Repetitive TMS over posterior STS disrupts perception of biological motion. *Vision Research, 45*(22), 2847-2853. doi: 10.1016/j.visres.2005.05.027

Grossman, E. D., & Blake, R. (1999). Perception of coherent motion, biological motion and form-from-motion under dim-light conditions. *Vision Research, 39*(22), 3721-3727. doi: 10.1016/S0042-6989%2899%2900084-X

Grossman, E. D., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron, 35*, 1167-1175. doi: 10.1016/S0896-6273(02)00897-8

Grove, P. M., & Sakurai, K. (2009). Auditory induced bounce perception persists as the probability of a motion reversal is reduced. *Perception, 38*(7), 951-965. doi: 10.1068/p5860

Gunns, R. E., Johnston, L., & Hudson, S. M. (2002). Victim selection and kinematics: A point-light investigation of vulnerability to attack. *Journal of Nonverbal Behavior, 26*(3), 129-158. doi: 10.1023/A:1020744915533

Guttman, S. E., Gilroy, L. A., & Blake, R. (2005). Hearing what the eyes see - auditory encoding of visual temporal sequences. *Psychological Science, 16*(3), 228-235. doi: 10.1111/j.0956-7976.2005.00808.x

Hari, R., Avikainen, S., Kirveskari, E., Salenius, S., & Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: a neuromagnetic study. *Proceedings of the National Academy of Sciences, 15061-15065*. doi: 10.1073/pnas.95.25.15061

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

Herman, L. M., Morrelsamuels, P., & Pack, A. A. (1990). Bottlenosed dolphin and human recognition of veridical and degraded video displays of an artificial gestural language. *Journal of Experimental Psychology: General, 119*(2), 215-230. doi: 10.1037/0096-3445.119.2.215

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology, 63*(3), 289-293. doi: 10.1037/h0039516

Hewig, J., Trippe, R. H., Hecht, H., Straube, T., & Miltner, W. H. R. (2008). Gender differences for specific body regions when looking at men and women. *Journal of Nonverbal Behavior, 32*(2), 67-78. doi: 10.1007/s10919-007-0043-5

Hill, H., & Pollick, F. E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light displays. *Psychological Science, 11*(3), 223-228. doi: 10.1111/1467-9280.00245

Hiris, E. (2007). Detection of biological and nonbiological motion. *Journal of vision, 7*(12). doi: 10.1167/7.12.4

Hiris, E., & Ewing, K. (2010). The perceived sex of biological motion displays is influenced by adaptation to biological motion but not adaptation to static faces. [Abstract]. *Journal of Vision, 10*(7), 781. doi: 10.1167/10.7.781

Hiris, E., Humphrey, D., & Stout, A. (2005). Temporal properties in masking biological motion. *Perception & Psychophysics, 67*(3), 435-443.

Hirnstein, M., Bayer, U., & Hausmann, M. (2009). Sex-specific response strategies in mental rotation. *Learning and Individual Differences, 19*(2), 225-228. doi: 10.1016/j.lindif.2008.11.006

Hogervorst, M. A., & Eagle, R. A. (1998). Biases in three-dimensional structure-from-motion arise from noise in the early visual system. *Proceedings of the Royal Society of London Series B-Biological Sciences, 265*(1406), 1587-1593.

Holmes, N. P. (2007). The law of inverse effectiveness in neurons and behaviour: Multisensory integration versus normal variability. *Neuropsychologia, 45*(14), 3340-3345. doi: 10.1016/j.neuropsychologia.2007.05.025

Holmes, N. P. (2009). The principle of inverse effectiveness in multisensory integration: some statistical considerations. *Brain Topography, 21*(3-4), 168-176. doi: 10.1007/s10548-009-0097-2

Holmes, N. P., Calvert, G. A., & Spence, C. (2009). Multimodal integration. In M. D. Binder, N. Hirokawa & U. Windhorst (Eds.), *Encyclopedia of Neuroscience*. Berlin Heidelberg: Springer.

Hommel, B., Musseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24*(5), 849-878.

Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation*. New York: Wiley.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science, 286*(5449), 2526-2528. doi: 10.1126/science.286.5449.2526

Ikeda, H., Blake, R., & Watanabe, K. (2005). Eccentric perception of biological motion is unscalably poor. *Vision Research, 45*(15), 1935-1943. doi: 10.1016/j.visres.2005.02.001

Jackson, S., & Blake, R. (2010). Neural integration of information specifying human structure from form, motion, and depth. *Journal of Neuroscience, 30*(3), 838-848. doi: 10.1523/jneurosci.3116-09.2010

Jackson, S., Cummins, F., & Brady, N. (2008). Rapid perceptual switching of a reversible biological figure. *Plos One, 3*(12), e3982. doi: 10.1371/journal.pone.0003982

Jacobs, A., & Shiffrar, M. (2005). Walking perception by walking observers. *Journal of Experimental Psychology: Human Perception and Performance, 31*(1), 157-169. doi: 10.1037/0096-1523.31.1.157

Jastorff, J., & Orban, G. A. (2009). Human functional magnetic resonance imaging reveals separation and integration of shape and motion cues in biological motion processing. *Journal of Neuroscience, 29*(22), 7315-7329. doi: 10.1523/jneurosci.4870-08.2009

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*(2), 201-211.

Johnson, K. L., & Tassinary, L. G. (2005). Perceiving sex directly and indirectly - Meaning in motion and morphology. *Psychological Science, 16*(11), 890-897. doi: 10.1111/j.1467-9280.2005.01633.x

Jordan, H., Fallah, M., & Stoner, G. R. (2006). Adaptation of gender derived from biological motion. *Nature Neuroscience, 9*(6), 738-739. doi: 10.1038/nn1710

Jordan, H., Reiss, J. E., Hoffman, J. E., & Landau, B. (2002). Intact perception of biological motion in the face of profound spatial deficits: Williams Syndrome. *Psychological Science, 13*(2), 162-167. doi: 10.1111/1467-9280.00429

Kaplan, J., & Iacoboni, M. (2007). Multimodal action representation in human left ventral premotor cortex. *Cognitive Processing, 8*(2), 103-113. doi: 10.1007/s10339-007-0165-z

Keetels, M., & Vroomen, J. (in press). Perception of synchrony between the senses. In M. M. Murray & M. T. Wallace (Eds.), *Frontiers in the neural basis of multisensory processes*.

Keysers, C., Kohler, E., Umilta, M. A., Nanetti, L., Fogassi, L., & Gallese, V. (2003). Audiovisual mirror neurons and action recognition. *Experimental Brain Research, 153*, 628-636. doi: 10.1007/s00221-003-1603-5

Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An interference effect of observed biological movement on action. *Current Biology, 13*(6), 522-525. doi: 10.1016/s0960-9822(03)00165-9

Klopfer, D. S. (1991). Apparent reversals of a rotating mask - a new demonstration of cognition in perception. *Perception & Psychophysics, 49*(6), 522-530.

Knoblich, G., Thornton, I. M., Grosjean, M., & Shiffrar, M. (2006). *The human body perception from the inside out*. New York, NY, USA: Oxford University Press.

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica, 134*(3), 372-384. doi: 10.1016/j.actpsy.2010.03.010

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science, 297*, 846-848. doi: 10.1126/science.1070311

Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics, 21*, 575-580.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* Technical Report A-8. University of Florida. Gainesville.

Lange, J., Georg, K., & Lappe, M. (2006). Visual perception of biological motion by form: A template-matching analysis. *Journal of Vision, 6*(8), 836-849. doi: 10.1167/5.4.6

Lange, J., & Lappe, M. (2006). A model of biological motion perception from configural form cues. *Journal of Neuroscience, 26*(11), 2894-2906. doi: 10.1523/JNEUROSCI.4915-05.2006

Leder, H., Candrian, G., Huber, O., & Bruce, V. (2001). Configural features in the context of upright and inverted faces. *Perception, 30*(1), 73-83. doi: 10.1068/p2911

Lewis, R., & Noppeney, U. (2010). Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *Journal of Neuroscience, 30*(37), 12329-12339. doi: 10.1523/jneurosci.5745-09.2010

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex-differences in spatial ability - a meta-analysis. *Child Development, 56*(6), 1479-1498. doi: 10.2307/1130467

Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movement. *Journal of Experimental Psychology: Human Perception and Performance, 31*(1), 210-220. doi: 10.1037/0096-1523.31.1.210

Love, S., Hillis, J. M., Waadeland, C., Rocchesso, D., Avanzini, F., Dahl, S., et al. (2007). How audio and visual cues combine to discriminate tempo of swing groove drumming. *Journal of vision, 7*(9), 870. doi: 10.1167/7.9.870

Maier, J. X., Neuhoff, J. G., Logothetis, N. K., & Ghazanfar, A. A. (2004). Multisensory integration of looming signals by rhesus monkeys. *Neuron, 43*(2), 177-181. doi: 10.1016/j.neuron.2004.06.027

Mamassian, P., & Landy, M. S. (1998). Observer biases in the 3D interpretation of line drawings. *Vision Research, 38*(18), 2817-2832. doi: 10.1016/S0042-6989%2897%2900438-0

Manera, V., Schouten, B., Becchio, C., Bara, B. G., & Verfaillie, K. (2010). Inferring intentions from biological motion: A stimulus set of point-light communicative interactions. *Behavior Research Methods, 42*, 168-178. doi: 10.3758/BRM.42.1.168

Mather, G., & Murdoch, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London Series B-Biological Sciences, 258*(1353), 273-279. doi: 10.1098/rspb.1994.0173

Mather, G., Radford, K., & West, S. (1992). Low-level visual processing of biological motion. *Proceedings of the Royal Society of London Series B-Biological Sciences, 249*(1325), 149-155.

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%2802%2901903-4

McGurk, H., & McDonald, J. W. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748. doi: 10.1038/264746a0

Mendonça, A. C., & Santos, J. A. (2008, July). *Auditory footsteps affect visual biological motion orientation detection.* Poster presented at the International Multisensory Research Forum. Hamburg, Germany.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons .1. temporal factors. *Journal of Neuroscience, 7*(10), 3215-3229.

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science, 221*(4608), 389-391. doi: 10.1126/science.6867718

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology, 56*, 640-662.

Meredith, M. A., & Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *Journal of Neurophysiology, 75*(5), 1843-1857.

Metzger, W. (1934). Beobachtungen über phänomenale Identität. *Psychologische Forschung, 19*, 1-60.

Meyer, G. F., & Wuerger, S. M. (2001). Cross-modal integration of auditory and visual motion signals. *Neuroreport, 12*(11), 2557-2560. doi: 10.1097/00001756-200108080-00053

Molenberghs, P., Brander, C., Mattingley, J. B., & Cunnington, R. (2010). The role of the superior temporal sulcus and the mirror neuron system in imitation. *Human Brain Mapping, 31*(9), 1316-1326. doi: 10.1002/hbm.20938

Montepare, J. M., Goldstein, S. B., & Clausen, A. (1987). The identification of emotions from gait information. *Journal of Nonverbal Behavior, 11*(1), 33-42. doi: 10.1007/BF00999605

Montepare, J. M., & Zebrowitz-McArthur, L. (1988). Impressions of people created by age-related qualities of their gaits. *Journal of Personality and Social Psychology, 55*(4), 547-556. doi: 10.1037/0022-3514.55.4.547

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research, 17*(1), 154-163. doi: 10.1016/S0926-6410%2803%2900089-2

Morrell, L. K. (1968). Temporal characteristics of sensory interaction in choice reaction times. *Journal of Experimental Psychology, 77*(1), 14-18. doi: 10.1037/h0021289

Nelissen, K., Luppino, G., Vanduffel, W., Rizzolatti, G., & Orban, G. A. (2005). Observing others: Multiple action representation in the frontal lobe. *Science, 310*(5746), 332-336. doi: 10.1126/science.1115593

Neri, P. (2009). Wholes and subparts in visual processing of human agency. *Proceedings of the Royal Society B-Biological Sciences, 276*(1658), 861-869. doi: 10.1098/rspb.2008.1363

Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience, 9*(9), 1186-1192. doi: 10.1038/nn1759

Neri, P., Morrone, M. C., & Burr, D. C. (1998). Seeing biological motion. *Nature, 395*(6705), 894-896. doi: 10.1038/27661

Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature, 395*(6698), 123-124. doi: 10.1038/25862

O'Leary, A., & Rhodes, G. (1984). Crossmodal effects on visual and auditory object perception. *Perception & Psychophysics, 35*(6), 565-569.

Oh, S. (2009). The reversible limbs of a stick walker. *Perception, 38*(1), 149-152. doi: 10.1068/p6275

Oram, M. W., & Perrett, D. I. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to biological motion stimuli. *Journal of Cognitive Neuroscience, 6*, 99-116. doi: 10.1162/jocn.1994.6.2.99

Pavlova, M., Krägeloh-Mann, I., Birbaumer, N., & Sokolov, A. (2002). Biological motion shown backwards: the apparent-facing effect. *Perception, 31*, 435-443. doi: 10.1068/p3262

Pavlova, M., & Sokolov, A. (2000). Orientation specificity in biological motion perception. *Perception & Psychophysics, 62*(5), 889-899.

Peelen, M. V., Wiggett, A. J., & Downing, P. E. (2006). Patterns of fmri activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron, 49*(6), 815-822. doi: 10.1016/j.neuron.2006.02.004

Pena, D., Contreras, M. J., Shih, P. C., & Santacreu, J. (2008). Solution strategies as possible explanations of individual and sex differences in a dynamic spatial task. *Acta Psychologica, 128*(1), 1-14. doi: 10.1016/j.actpsy.2007.09.005

Perrett, D. I., Smith, P. A. J., Mistlin, A. J., Chitty, A. J., Head, A. S., Potter, D. D., et al. (1985). Visual analysis of body movements by neurons in the temporal cortex of

the macaque monkey - a preliminary report. *Behavioural Brain Research, 16*(2-3), 153-170. doi: doi:10.1016/0166-4328(85)90089-0

Petrini, K., Dahl, S., Rocchesso, D., Waadeland, C. H., Avanzini, F., Puce, A., et al. (2009). Multisensory integration of drumming actions: musical expertise affects perceived audiovisual asynchrony. *Experimental Brain Research, 198*(2-3), 339-352. doi: 10.1007/s00221-009-1817-2

Petrini, K., Holt, S. P., & Pollick, F. (2010). Expertise with multisensory events eliminates the effect of biological motion rotation on audiovisual synchrony perception. *Journal of Vision, 10*(5). doi: 10.1167/10.5.2

Petrini, K., Russell, M., & Pollick, F. (2009). When knowing can replace seeing in audiovisual integration of actions. *Cognition, 110*(3), 432-439. doi: 10.1016/j.cognition.2008.11.015

Peuskens, H., Vanrie, J., Verfaillie, K., & Orban, G. A. (2005). Specificity of regions processing biological motion. *European Journal of Neuroscience, 21*(10), 2864-2875. doi: 10.1111/j.1460-9568.2005.04106.x

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

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition, 82*(2), B51-B61. doi: 10.1016/S0010-0277(01)00147-0

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology, 9*(2), 129-154. doi: 10.1080/713752551

Puce, A., & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical Transactions of the Royal Society of London Series B - Biological Sciences, 358*(1431), 435-445. doi: 10.1098/rstb.2002.1221

Reed, C. L., & Farah, M. J. (1995). The psychological reality of the body schema: A test with normal participants. *Journal of Experimental Psychology: Human Perception and Performance, 21*(2), 334-343. doi: 10.1037/0096-1523.21.2.334

Reed, C. L., Stone, V., Bozova, S., & Tanaka, J. (2003). The body inversion effect. *Psychological Science, 14*(4), 302-308. doi: 10.1111/1467-9280.14431

Reed, C. L., Stone, V. E., Grubb, J. D., & McGoldrick, J. E. (2006). Turning configural processing upside down: part and whole body postures. *Journal of Experimental*

Psychology: Human Perception and Performance, 32(1), 73-87. doi: 10.1037/0096-1523.32.1.73

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169-192. doi: 10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor action. *Cognitive Brain Research, 3*(2), 131-141. doi: 10.1016/0926-6410(95)00038-0

Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proceedings of the Royal Society B-Biological Sciences, 273*(1598), 2159-2168. doi: 10.1098/rspb.2006.3578.

Rosenblum, L., Carello, C., & Pastore, R. (1987). Relative effectiveness of three stimulus variables for locating a moving sound source. *Perception, 16*(2), 175-186. doi: 10.1068/p160175

Rosenthal, O., Shimojo, S., & Shams, L. (2009). Sound-induced flash illusion is resistant to feedback training. *Brain Topography, 21*(3-4), 185-192. doi: 10.1007/s10548-009-0090-9

Runeson, S., & Frykholm, G. (1981). Visual perception of lifted weight. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 733-740. doi: 10.1037/0096-1523.7.4.733

Safford, A. S., Hussey, E. A., Parasuraman, R., & Thompson, J. C. (2010). Object-based attentional modulation of biological motion processing: Spatiotemporal dynamics using functional magnetic resonance imaging and electroencephalography. *Journal of Neuroscience, 30*(27), 9064-9073. doi: 10.1523/jneurosci.1779-10.2010

Saunders, D. R., Suchan, J., & Troje, N. F. (2009). Off on the wrong foot: Local features in biological motion. *Perception, 38*(4), 522-532. doi: 10.1068/p6140

Saygin, A. P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain, 130*, 2452-2461. doi: 10.1093/brain/awm162

Saygin, A. P., Driver, J., & De Sa, V. R. (2008). In the footsteps of biological motion and multisensory perception: judgments of audiovisual temporal relations are enhanced for upright walkers. *Psychological Science, 19*(5), 469-475. doi: 10.1111/j.1467-9280.2008.02111.x

Saygin, A. P., Wilson, S. M., Hagler, D. J., Bates, E., & Sereno, M. I. (2004). Point-light biological motion perception activates human premotor cortex. *The Journal of Neuroscience, 24*(27), 6181-6188. doi: 10.1523/jneurosci.0504-04.2004

Schiff, W., Caviness, J. A., & Gibson, J. J. (1962). Persistent fear responses in rhesus monkeys to the optical stimulus of looming. *Science, 136*, 982-983. doi: 10.1126/science.136.3520.982

Schouten, B., Moyens, E., Brooks, A., van der Zwan, & Verfaillie, K. (2008, July). *The effect of looming and receding sounds on the in-depth perception of point-light figures.* Poster presented at the International Multisensory Research Forum. Hamburg, Germany.

Schouten, B., Troje, N. F., Brooks, A., van der Zwan, R., & Verfaillie, K. (2010). The facing bias in biological motion perception: Effects of stimulus gender and observer sex. *Attention, Perception & Psychophysics, 72*(5), 1256-1260. doi: 10.3758/APP.72.5.1256

Schouten, B., Troje, N. F., & Verfaillie, K. (in press). The facing bias in biological motion perception: Structure, kinematics, and body parts. *Attention, Perception & Psychophysics*. doi: 10.3758/s13414-010-0018-1

Schouten, B., Troje, N. F., Vroomen, J., & Verfaillie, K. (2010). Looming and receding sounds affect the perceived in-depth orientation of depth-ambiguous biological motion figures. *Plos One*, Manuscript under revision.

Schouten, B., & Verfaillie, K. (2006). The audiovisual perception of biological motion. [Abstract]. *Perception, 35*, 205-205.

Schouten, B., & Verfaillie, K. (2010). Determining the point of subjective ambiguity of ambiguous biological motion figures with perspective cues. *Behavior Research Methods, 42*, 161-167. doi: 10.3758/BRM.42.1.161

Schroeder, C. E., & Foxe, J. (2005). Multisensory contributions to low-level, 'unisensory' processing. *Current Opinion in Neurobiology, 15*(4), 454-458. doi: 10.1016/j.conb.2005.06.008

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature, 385*, 308. doi: 10.1038/385308a0

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature, 408*, 788. doi: 10.1038/35048669

Shams, L., Kamitani, Y., & Shimojo, S. (2002). A visual illusion induced by sound. *Cognitive Brain Research, 14*, 147-152. doi: 10.1016/S0926-6410%2802%2900069-1

Shams, L., Kamitani, Y., & Shimojo, S. (2004). Modulations of visual perception by sound. In G. Calvert, C. Spence & B. E. Stein (Eds.), *The handbook of multisensory processes.* (pp. 27-33). Cambridge: MIT Press.

Shams, L., Kamitani, Y., Thompson, S., & Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *NeuroReport, 12*(17), 3849-3852. doi: 10.1097/00001756-200112040-00049

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology, 28*(2), 61-70. doi: 10.1250/ast.28.61

Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics, 59*(1), 1-22.

Spence, C., & Driver, J. (2002). *Crossmodal space and crossmodal attention*. Oxford: University Press.

Spence, C., & Parise, C. (2010). Prior-entry: A review. *Consciousness and Cognition, 19*(1), 364-379. doi: 10.1016/j.concog.2009.12.001

Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: A review. *Hearing Research, 258*(1-2), 134-142. doi: 10.1016/j.heares.2009.04.015

Spence, C., Senkowski, D., & Roder, B. (2009). Crossmodal processing. [Editorial Material]. *Experimental Brain Research, 198*(2-3), 107-111. doi: 10.1007/s00221-009-1973-4

Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Current Biology, 13*(13), 519-521. doi: 10.1016/S0960-9822(03)00445-7

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience, 8*, 497-506. doi: 10.1162/jocn.1996.8.6.497

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, Massachusetts: MIT Press.

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience, 19*(12), 1964-1973. doi: 10.1162/jocn.2007.19.12.1964

Sumi, S. (1984). Upside-down presentation of the Johansson moving light-spot pattern. *Perception, 13*(3), 283-286. doi: 10.1068/p130283

Thirkettle, M., Benton, C. P., & Scott-Samuel, N. E. (2009). Contributions of form, motion and task to biological motion perception. *Journal of Vision, 9*(3). doi: 10.1167/9.3.28

Thompson, B., Hansen, B. C., Hess, R. F., & Troje, N. F. (2007). Peripheral vision: Good for biological motion, bad for signal noise segregation? *Journal of Vision, 7*(10). doi: 10.1167/7.10.12

Thompson, J. C., Clarke, M., Stewart, T., & Puce, A. (2005). Configural processing of biological motion in human superior temporal sulcus. *Journal of Neuroscience, 25*(39), 9059-9066. doi: 10.1523/jneurosci.2129-05.2005

Thornton, I. M., & Garvey, M. (2010, August). *Disambiguating biological motion*. Poster presented at the European conference on Visual Processing. Lausanne, Switserland.

Thornton, I. M., Pinto, J., & Shiffrar, M. (1998). The visual perception of human locomotion. *Cognitive Neuropsychology, 15*, 535-552. doi: 10.1080/026432998381014

Thornton, I. M., Rensink, R. A., & Shiffrar, M. (2002). Active versus passive processing of biological motion. *Perception, 31*, 837-853. doi: 10.1068/p3072

Thornton, I. M., & Vuong, Q. C. (2004). Incidental processing of biological motion. *Current Biology, 14*(12), 1084-1089. doi: 10.1016/j.cub.2004.06.025

Thornton, I. M., Vuong, Q. C., & Bulthoff, H. H. (2003). A chimeric point-light walker. *Perception, 32*(3), 377-383. doi: 10.1068/p5010

Thurman, S. M., & Grossman, E. D. (2008). Temporal "Bubbles" reveal key features for point-light biological motion perception. *Journal of Vision, 8*(3). doi: 10.1167/8.3.28

Troje, N. F. (2002). Decomposing biological motion: A framework for analyses and synthesis of human gait patterns. *Journal of vision, 2*, 371-387. doi: 10:1167/2.5.2

Troje, N. F. (2008a). Biological motion perception. In A. I. Basbaum, A. Kaneko, G. M. Shepard & G. Westheimer (Eds.), *The Senses: A Comprehensive Reference, Vol 2, Vision II.* (pp. 231-238). San Diego: Academic Press.

Troje, N. F. (2008b). Retrieving information from human movement patterns. In T. F. Shipley & J. M. Zacks (Eds.), *Understanding events: How humans see, represent, and act on events* (pp. 308-334): Oxford University Press.

Troje, N. F. (2010). Perceptual biases in biological motion perception and other depth-ambiguous stimuli. [Abstract]. *Journal of Vision, 10*(7), 792. doi: 10.1167/10.7.792

Troje, N. F., & Chang, D. H. F. (in press). Shape-invariant processing of biological motion. In K. Johnson & M. Shiffrar (Eds.), *Visual Perception of the Human Body in Motion: Findings,*
Theory, and Practice. Oxford: 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., & Szabo, S. (2006). Why is the average walker male? [Abstract]. *Journal of Vision, 6*(6), 1034. doi: 10.1167/6.6.1034.

Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a "life detector"? *Current Biology, 16*(8), 821-824. doi: 10.1016/j.cub.2006.03.022

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.

Tuomainen, J., Andersen, T. S., Tiippana, K., & Sams, M. (2005). Audio-visual speech perception is special. *Cognition, 96*(1), B13-B22. doi: 10.1016/j.cognition.2004.10.004

Van Belle, G., De Graef, P., Verfaillie, K., Rossion, B., & Lefevre, P. (2010). Face inversion impairs holistic perception: Evidence from gaze-contingent stimulation. *Journal of Vision, 10*(5). doi: 10.1167/10.5.10

Van Belle, G., De Smet, M., De Graef, P., Van Gool, L., & Verfaillie, K. (2009). Configural and featural processing during face perception: A new stimulus set. *Behavior Research Methods, 41*(2), 279-283. doi: 10.3758/brm.41.2.279

Van der Burg, E., Cass, J., Olivers, C. N. L., Theeuwes, J., & Alais, D. (2010). Efficient visual search from synchronized auditory signals requires tansient audiovisual events. *Plos One, 5*(5), e10664. doi: 10.1371/journal.pone.0010664

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology-Human Perception and Performance, 34*(5), 1053-1065. doi: 10.1037/0096-1523.34.5.1053

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 Eijk, R. L. J. (2008). *Audiovisual synchrony perception*. Unpublished doctoral dissertation. Technische Universiteit Eindhoven. Eindhoven.

van Eijk, R. L. J., Kohlrausch, A., Juola, J. F., & van de Par, S. (2008). Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception & Psychophysics, 70*(6), 955-968. doi: 10.3758/pp.70.6.955

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of 3-dimensional spatial visualization. *Perceptual and Motor Skills, 47*(2), 599-604.

Vangeneugden, J. (2010). *Action representation by macaque visual temporal cortical neurons.* Unpublished doctoral dissertation. K.U.Leuven. Leuven.

Vangeneugden, J., Pollick, F., & Vogels, R. (2009). Functional differentiation of macaque visual temporal cortical neurons using a parametric action space. *Cerebral Cortex, 19*(3), 593-611. doi: 10.1093/cercor/bhn109

Vanrie, J. (2005). *Bistability and the perception of depth in point-light actions*. Unpublished doctoral dissertation. KULeuven. Leuven.

Vanrie, J., Dekeyser, M., & Verfaillie, K. (2004). Bistability and biasing effects in the perception of ambiguous point-light walkers. *Perception, 33*(5), 547-560. doi: 10.1068/p5004

Vanrie, J., & Verfaillie, K. (2004). Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, 36*(4), 625-629.

Vanrie, J., & Verfaillie, K. (2006). Perceiving depth in point-light actions. *Perception & Psychophysics, 68*(4), 601-612.

Vanrie, J., & Verfaillie, K. (2008). Perceptual coupling of multiple point-light figures. *Visual Cognition, 16*(5), 585-615. doi: 10.1080/13506280701269334

Vanrie, J., & Verfaillie, K. (2010). On the reversibility of point-light actions. *Manuscript submitted for publication*.

Verfaillie, K. (1993). Orientation-dependent priming effects in the perception of biological motion. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 992-1013. doi: 10.1037/0096-1523.19.5.992

Verfaillie, K. (2000). Perceiving human locomotion: Priming effects in direction discrimination. *Brain and Cognition, 44*, 192-213. doi: 10.1006/brcg.2000.1228

Verfaillie, K., & De Graef, P. (2000). Transsaccadic memory for position and orientation of saccadic source and target. *Journal of Experimental Psychology: Human Perception and Performance, 26*(4), 1243-1259. doi: 10.1037/0096-1523.26.4.1243

Verfaillie, K., De Troy, A., & Van Rensbergen, J. (1994). Transsaccadic integration of biological motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(3), 649-670. doi: 10.1037/0278-7393.20.3.649

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex-differences in spatial abilities - A metaanalysis and consideration of critical variables. *Psychological Bulletin, 117*(2), 250-270. doi: 10.1037/0033-2909.117.2.250

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: Crossmodal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance, 26*(5), 1583-1590. doi: 10.1037/0096-1523.26.5.1583

Vroomen, J., & de Gelder, B. (2004a). Perceptual effects of cross-modal stimulation: Ventriloquism and the freezing phenomenon. In G. Calvert, C. Spence & B. E. Stein (Eds.), *The handbook of multisensory processes.* (pp. 141-152). Cambridge: MIT Press.

Vroomen, J., & de Gelder, B. (2004b). Temporal ventriloquism: Sound modulates the flash-lag effect. *Journal of Experimental Psychology: Human Perception and Performance, 30*(3), 513-518. doi: 10.1037/0096-1523.30.3.513

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics, 72*, 871-884. doi: 10.3758/APP.72.4.871

Wallace, M. T., Meredith, M. A., & Stein, B. E. (1998). Multisensory integration in the superior colliculus of the alert cat. *Journal of Neurophysiology, 80*(2), 1006-1010.

Watanabe, K., & Shimojo, S. (2001). When sound affects vision: Effects of auditory grouping on visual motion perception. *Psychological Science, 12*(2), 109-116. doi: 10.1111/1467-9280.00319

Watkins, S., Shams, L., Josephs, O., & Rees, G. (2007). Activity in human V1 follows multisensory perception. *Neuroimage, 37*(2), 572-578. doi: 10.1016/j.neuroimage.2007.05.027

Watkins, S., Shams, L., Tanaka, S., Haynes, J. D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *Neuroimage, 31*(3), 1247-1256. doi: 10.1016/j.neuroimage.2006.01.016

Weiss, E. M., Kemmler, G., Deisenhammer, E. A., Fleischhacker, W. W., & Delazer, M. (2003). Sex differences in cognitive functions. *Personality and Individual Differences, 35*(4), 863-875. doi: 10.1016/S0191-8869%2802%2900288-X

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*(3), 638-667. doi: 10.1037/0033-2909.88.3.638

Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of Fit. *Perception & Psychophysics, 63*(8), 1293-1313.

Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics, 63*(8), 1314-1329.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin, 131*(3), 460-473. doi: 10.1037/0033-2909.131.3.460

Movies[19]

Movie 1: Contemporary example of Johansson's point-light technique.

Movie 2: Limited lifetime point-light walker by Beintema and Lappe (2002).

Movie 3: Orthographic frontal/back projection of a point-light walker.

Movie 4: Perspective back projection of a point-light walker (FVA = 90°).

Movie 5: Illustration of the perspective technique

Movie 6: Superposition of the gender figures that only differ in structural information (neutral kinematics). The more red the circles, the more female the figure. The more blue the circles the more male the figure. Note that this is not how point-light figures were presented to observers. This animation illustrates the structural differences between the seven *structure-only stimuli*.

Movie 7: Superposition of the gender figures that only differ in kinematic information (neutral structure). The more red the circles, the more female the figure. The more blue the circles the more male the figure. Note that this is not how point-light figures were presented to observers. This animation illustrates the kinematic differences between the seven *kinematics-only stimuli*.

Movie 8: Illustration of the McGurk effect.

Movie 9: Example of a signal trial in the synchronized condition of Experiment 2 of Arrighi, Marini, and Burr (2009).

Movie 10: Illustration of the stimulus creation of the audiovisual point-light dancer.

Movie 11: Example of a trial in the synchronized condition in the visual search task of Experiment 2 in Chapter 9 (3 signal dots).

[19] Movies available at https://perswww.kuleuven.be/~u0050370/PhD