Uminho | 2017

Dominic Noy

A MULTIMODAL APPROACH TO INTERPERSONAL GAIT SYNCHRONIZATION

Universidade do Minho
Escola de Psicologia

Dominic Noy

A MULTIMODAL APPROACH TO INTERPERSONAL GAIT SYNCHRONIZATION

novembro 2017

Universidade do Minho
Escola de Psicologia

Dominic Noy

A MULTIMODAL APPROACH TO INTERPERSONAL GAIT SYNCHRONIZATION

A thesis submitted for the degree of
Doctor of Philosophy

Supervised by:
Dr. Jorge A. Santos & Dr. Sandra Mouta

novembro 2017

Declaration

Email:
dominic.noy@gmail.com

Title:
A Multimodal Approach to Interpersonal Gait Synchronization

Supervisors:
Dr. Jorge A. Santos & Dr. Sandra Mouta

Year of Conclusion:
2017

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA TESE APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

University of Minho, ______ de ________________ de 2017.

The author: ____

State of Integrity

I hereby declare having conducted my thesis with integrity. I confirm that I have not used plagiarism or any form of falsification of results in the process of the thesis elaboration. I further declare that I have fully acknowledged the Code of the Ethical Conduct of the University of Minho.

University of Minho, ______ de ________________ de 2017.

Full name: ____

Signature: ____

Dedication

To Inês and Jorge.

Acknowledgements

I would like to thank my supervisors Jorge Santos and Sandra Mouta for their guidance and patience; my friends and colleagues João Lamas and Daniel Basso, without their help all these fancy studies would have not been possible; Carlos Coelho, who was besides Inês and Jorge the most important emotional support; Catarina Mendoça and Marko Hiipakka, who were really important for me at the start of my PhD; Isabel Lisboa, Emanuel Sousa, and Alfredo Perreira for all the smart advices and help; and of course, all the other members of the Lab, who made the last five years so comfortable and enjoyable. Finally and most importantly, I want to thank my family, spread over Portugal, Germany, Hungary, and Australia.

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

This thesis was funded with a doctoral scholarship (bolso de doutoramento: SFRH/BD/88396/2012) by the Fundação para Ciência e Tecnologia

Abstract

When people walk side-by-side, they often synchronize their movements. First, we investigated in three experiments whether audiovisual signals from the walking partner are integrated according to a mechanism operating as a Maximum Likelihood Estimator (MLE). Sensory cues from a walking partner were virtually simulated. In Experiment 1, seven participants synchronized with human-sized Point Light Walkers and/or footstep sounds. Results revealed highest performance when auditory cues were presented, regardless of the visual ones. This auditory dominance effect might have been due to artifacts of the setup. Therefore, in Experiment 2, human-sized virtual mannequins were implemented and the audiovisual stimuli were rendered in real-time in order to guarantee cross-modal congruence, co-localization, and synchrony. All four participants synchronized best with audiovisual cues and for three participants the results are consistent with the MLE model. Finally, Experiment 3 yielded performance decrements for three participants when the cues were temporally incongruent. These findings suggest that the integration of congruent audiovisual cues increase the intentional step synchronization of side-by-side walkers. In a fourth experiment, we tested whether synchronization is achieved by matching global body motion rather than single segments like the feet. Eight pairs of participants walked side-by-side in a large field. Results revealed that asynchronies between signals obtained from the principal components of co-variation of several body segments vary less than the asynchronies computed from individual body segments suggesting a synchronization of the global body motions of the walkers. The overall findings are partially consistent with the information processing approach and the dynamical system approach. The work's output also highlights that it requires a very high spatiotemporal alignment of the stimuli when using such virtual environment techniques in contexts like rehabilitation or sports.

Resumo

Quando as pessoas caminham lado a lado, muitas vezes sincronizam os seus movimentos. Realizamos três estudos de forma a saber se os sinais audiovisuais oriundos do parceiro são integrados de acordo com um mecanismo que opera como um estimador de Máxima Verossimilhança (MLE). A informação sensorial de um parceiro foi simulada virtualmente. Na Experiência 1, sete participantes sincronizaram a sua passada com um Point Light Walker e/ou com sons de passos. Os resultados revelaram um melhor desempenho durante a presença de sinais auditivos. O efeito de dominância auditiva pode ter sido devido a artefactos da experiência. Portanto, na Experiência 2 foram implementados humanoides virtuais e os estímulos foram renderizados em tempo real para garantir a congruência entre modalidades. Os quatro participantes sincronizaram melhor com informação audiovisual e dos quatro, três são consistentes com o MLE. Posteriormente, na Experiência 3 o desempenho de três participantes diminuiu quando as pistas estavam temporalmente incongruentes. Isto mostra que a integração dos sinais audiovisuais congruentes aumenta a sincronização intencional da marcha. Desta forma, numa quarta experiência, testamos se esta sincronização é conseguida através dos movimentos globais do corpo de ambos os intervenientes. Oito pares de participantes caminharam lado ao lado. Os resultados revelaram que quando a medida de sincronia combina os sinais obtidos através da componente de covariação de vários segmentos do corpo, esta medida é mais precisa do que se for realizada a partir de um qualquer segmento, sugerindo que os humanos se sincronizam com os movimentos globais. Os resultados são consistentes com uma abordagem que assume um controlo intencional de nível cortical superior, bem como com a abordagem de sistemas dinâmicos. O trabalho tem implicações para o uso de ambientes virtuais em contextos como a reabilitação ou o desporto e salienta que os estímulos requerem um alinhamento espaço-temporal muito elevado.

Contents

List of Figures

List of Tables

Abbreviations

CNS	Central Nervous System
COM	Centre of Body Mass
CPG	Central Pattern Generator
DRP	Discrete Relative Phase
DT	De-synchronization Threshold
HRTF	Head-Related Transfer Function
LPC	Linear Phase Correction Model
MLE	Maximum Likelihood Estimator
MN	Mirrow Neurons
PCA	Principle Component Analysis
PC1	First Principle Component
PLW	Point Light Walker
SMS	Sensorimotor Synchronization
T	Synchronisation Threshold
WTI	Window of Temporal Integration

0.1 Introduction

0.1.1 Synchrony of Side-by-Side Walkers

When couples walk together, they frequently synchronize their movements without being aware of it. This phenomenon is called spontaneous synchronization (Zivotofsky, Gruendlinger, & Hausdorff, 2012; Zivotofsky & Hausdorff, 2007). There is much empirical evidence for spontaneous synchronization. For instance, two interacting individuals synchronize the involuntary oscillations of pendulums holding in their hands without that they intend to do so. Other examples are the movement synchronization of voluntary rhythmic forearm or finger movements and also when sitting next to each other in rocking chairs, rocking movements become synchronized (e.g., Coey, Varlet, Schmidt, & Richardson, 2011; Demos, Chaffin, Begosh, Daniels, & Marsh, 2012; Issartel, Marin, & Cadopi, 2007; Oullier, De Guzman, Jantzen, Lagarde, & Kelso, 2008; Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007; Richardson, Marsh, & Schmidt, 2005; Schmidt & O'Brien, 1997).

Spontaneous synchronization. In a typical spontaneous synchronization paradigm, two individuals are positioned close to each other and are asked to simultaneously execute a rhythmic movement for a period of time. Subsequently, the amount of synchronized movement periods is quantified. Although the individuals are not instructed to synchronize and often perform a secondary task so that they do not guess the purpose of the study, synchronization periods with values above chance can be observed.

The inquiry of spontaneous movement synchronization received much attention by the scientific community because it is a type of nonverbal interpersonal coordination (Schmidt & Richardson, 2008; Repp, 2005). Nonverbal coordination is a fundamental requirement for successful interactions of individuals in team sports, factory work, and also in simple everyday tasks. It requires a highly precise, mutual, and accurate spatiotemporal displacement of the body and individual segments.

Yet, synchronization differs from natural every-day coordination tasks. "Synchronous" comes from the Greek words "chronos", which means time, and "syn", which means "the same" or "common". Possible translation are "sharing the common time" and “occurring [at] the same time” (Pikovsky, Rosenblum, & Kurths, 2001, p. 2). It is often defined as a bounded temporal relationship (Mörtl et al., 2012). We adopted this definition of synchrony throughout this manuscript. Also, coordination can be understood as the timed displacement of movements, but in contrast to synchronization, movements do not need to be temporally bounded (see e.g., Latash, 2012). For this reason, synchronization can be understood as a constrained type of coordination where the possible degrees of freedom of the effector systems are spatiotemporally reduced. In this way, it provides a more controlled setup to study the cognitive timing capacities (see e.g., Repp, 2005) that are supposed to underlie synchronization; and findings still might be generalizable to the “more flexible” interpersonal coordination.

Traditionally, the study of interpersonal movement synchronization was restricted to movements that are not naturally periodic and are not performed as a typical behavioral routine. For instances, participants were asked to repetitively tap a finger or a forearm on a surface, swing a pendulum, or rock in a rocking chair (e.g., Coey et al., 2011; Demos et al., 2012; Issartel et al., 2007; Oullier et al., 2008; Richardson et al., 2007, 2005; Schmidt & O’Brien, 1997). The spontaneous synchronization of movements during side-by-side walking became only recently a scope of inquiry. A caveat of studying synchronization in walking is that the underlying control mechanisms seem far more complex than the control of the previously mentioned activities, like finger or forearm tapping. Its inquiry requires an unambiguous definition and understanding of the kinematics and its mechanisms.

Walking is one type of natural human gait. Others are jogging, running, and sprinting. Gait is defined as the coordination of body segments for locomotion. Locomotion is the displacement of the body. Thus, walking can be defined as particular coordination of body segments in order to displace the whole body.

A gait cycle is defined as the time interval between successive repetitive events forming a loop. As an event, it is usually chosen the initial contact of the ground with the foot of one leg. This is called the heel strike. The entire cycle is classified into seven events, being 1) initial contact, 2) opposite toe off, 3) heel rise, 4) opposite initial contact, toe off, feet adjacent, tibia vertical. The other foot follows the same sequence of events. The distance between two successive heel strikes is defined as a step length. The distance between heel strikes of the same foot is defined as a stride length. The number of steps per unit time is defined as cadence. The walking speed is the displacement per unit time. Cadence, cycle, stride length, and speed are relatively stable across adults (see Table 1).

Table 1: Gait Parameters. Retrieved from Levine et al. (2012).

Sex	Age	Cadence(steps/min)	Cycle(s)	Stride Length(m)	Speed(m/s)
F	18 - 49	98 -138	0.87 - 1.22	1.06 - 1.58	0.94 - 1.66
M	18 - 49	91 - 135	0.89 - 1.32	1.25 - 1.85	1.10 - 1.82

The seven events subdivide the gait cycle into seven periods: During stance phase, these are loading response, mid-stance, terminal stance, pre-swing. During swing phase, these are initial swing, mid-swing, terminal swing. During stance phase, which is approximately 60% of a cycle, the foot is in contact with the ground. During swing phase, which is approximately 40%, the leg leaves the ground and swings the foot forward like a pendulum(Levine, Richards, & Whittle, 2012). Then, at stance phase, the heel touches the ground and rolls over through to the toe. Like an inverted pendulum, the center of body mass passes over the top of the foot and the leg holds the body weight. Thus, walking can be formally described as a double pendulum (Garcia, Chatterjee, Ruina, Coleman, et al., 1998). The center of body mass exchanges gravitational potential energy and forward kinetic energy due to properties of the leg motion that are similar to a spring (Lacquaniti, Grasso, & Zago, 1999). Different to an ideal pendulum, energy must be invested–which is usually done by muscles–to compensate for energy lost in stance phase. Nerve signals activate appropriate muscle groups moving the limbs while maintaining balance, support, and stability.

Studying synchronization in walking is particularly attractive because the rhythmic movements emerge naturally and therefore rhythmicity does not need to be induced by the particular experimental task (Nessler, De Leone, & Gilliland, 2009; Nessler & Gilliland, 2009, 2010; van Ulzen, Lamoth, Daffertshofer, Semin, & Beek, 2008). The first study demonstrating the spontaneous synchronization of the movements of side-by-side walkers was conducted in 2007 (Zivotofsky & Hausdorff, 2007). Subsequently, it was demonstrated that it even emerges when gait characteristics—like the natural cadence or the step length—largely differ. This was shown in an experiment in which two individuals were walking together while their gait was manipulated by the attachment of weights to the ankles (Nessler, Gutierrez, Werner, & Punsalan, 2015). Also, when walking next to a confederate who tried to "desynchronize", large alterations were observed in the individuals' natural gait dynamics (Nessler et al., 2013). To provide a compelling example: In 2009, Usain Bolt won the 12th IAAF World Championship in Athletics. Video assessment of the final revealed that his movements were frequently synchronized with his opponent Tyson Gay (Varlet & Richardson, 2015). Step frequency in running is optimal when maximal efficiency and minimal oxygen consumption are balanced. From an optimal control theoretical point of view, their synchronization is surprising (Pikovsky et al., 2001) because the morphology of both runners differs (stature: Usain Bolt, 1.95m; Tyson Gay, 1.78m). Consequently, the spontaneous synchronization of walking remains an interesting subject of study, and its naturally inductive properties can be extended to high-performance training or rehabilitation (Willems et al., 2007; Hausdorff et al., 2007), as we discuss in Section 0.6.4.

Intentional synchronization. In nature, there are also examples in which synchronization is requested by the particular activity or it is directly demanded. Then, it is called intentional synchronization. Intentional synchronization can be observed during activities such as dancing and during artistic performances such as synchronized swimming or cheerleading. Also, soldiers are drilled in boot camps to synchronize their steps during the military march (Bohannon, 2017), and in team rowing, the crew members must synchronize their movements in order to efficiently dislocate the shell (de Brouwer, de Poel, & Hofmijster, 2013), or during musical performance.

Intentional movement synchronization is studied in tasks in which the oscillating movements have to co-occur with repetitive external events or with the oscillating movements of an interaction partner, called Sensorimotor Synchronization (SMS). Event and movement may have clearly identifiable discrete endpoints on which synchrony measures are based—like the moment when the finger contacts a surface or when the heel strikes the ground—or it is assessed the synchrony between the continuous motion trajectories.

A fundamental question is here how the different sensory systems are interacting to precisely time the appropriate movements (see Repp & Su, 2013; Repp, 2005). In the particular case of intentional movement synchronization during side-by-side walking, the individuals must perceive the spatiotemporal properties of their own movements and adapt to the movements of their partners while maintaining balance and heading direction. For adequate limb coordination, this must be timed within milliseconds (Duysens & Van de Crommert, 1998). It requires sensory cues from the environment and the body to be continuously integrated. In this work, we evaluate the contribution of different sensory conditions—visual, auditory and audio-visual—for the synchronization of side-by-side walkers.

Multiple sensory systems may be involved. For instances, in order to identify where an individual is within her/his gait cycle, it may be sufficient to watch the continuous displacement of the feet. Yet, during walking, the gaze is usually directed to future foot contact locations (Lappe, Bremmer, & Van den Berg, 1999). Nonetheless, the visual system provides cues for perceiving self-motion from which the current gait cycle position can be retrieved (Campos & Bülthoff, 2012). The most relevant visual cue for the perception of self-motion is optic flow. However, because the movements of different body parts (eye movements and head movement) are superimposed on global body displacement, optic flow cues have to be combined with other signals to robustly disambiguate the flow patterns (Lappe et al., 1999).

It was suggested that neural ensembles within the vestibular nucleus code self-motion by integrating the input from multiple afferent information (vestibular, visual, proprioceptive, somatosensory) and efference copies of the motor commands (Cullen, 2012; Fitzpatrick, Wardman, & Taylor, 1999). Their integration disambiguates cue information provided by single modalities (Cullen, 2012; de Winkel, Weesie, Werkhoven, & Groen, 2010). Moreover, besides visual self-motion inputs, proprioceptive and vestibular signals are integrated constituting the kinesthetic sense. In addition, proprioceptive and cutaneous signals are integrated for the perception of the position of segments relative to each other and relative to the surface (see Kaya, 2014).

If not holding hands, estimations about the current gait cycle phase of the walking partner may also be retrieved from auditory and visual cues. Detailed visual cues are provided, should the individual focus the movements of the partner. It is assumed that this does not happen constantly due to navigation and self-motion control demands

(Warren, Kay, & Yilmaz, 1996). Therefore, peripheral visual cues are more likely to be used for perceiving the motion of the partner. Also, salient auditory cues are provided by footstep sounds produced by a short-lasting large upward force on the foot at heel strike (Pastore, Flint, Gaston, & Solomon, 2008). Consequently, gait synchronization may be understood as the cross-modal matching of audiovisual signals from the partner with kinesthetic, cutaneous, visual, and auditory signals from one's own movements. An exhaustive description of the cues that are available to estimate the spatiotemporal position of one's own movements and the movements of the walking partner is presented in Section 0.1.3 and Section 0.1.5.

Yet, a caveat might be that signals from different sensory systems arrive asynchronously at different information processing levels due to different propagation, transduction, transmission, and processing times (Repp, 2005; Vroomen & Keetels, 2010) (see Section 0.1.6). Furthermore, environmental cues might be ambiguous, more or less regular and accessible (see e.g., Cullen, 2012; Hartmann, 1983; Kopčo & Shinn-Cunningham, 2011; Kolarik, Moore, Zahorik, Cirstea, & Pardhan, 2016), and there is noise within each sensory system (Ernst & Bülthoff, 2004) (see Section 0.1.3 and Section 0.1.4).

There are different theories about the way the CNS deals with the conflicting information in order to construct a robust percept and/or precisely time movements with environmental events (see Section 0.1.4). Currently, an account receiving most empirical support suggests that multiple information may be integrated according to a mechanism called Maximum Likelihood Estimator (see Bayesian Optimal Integration Theory Alais & Burr, 2004; Ernst & Bülthoff, 2004; Ernst, 2006) (see Section 0.1.4).

While there are many studies providing evidence for the presence of such a mechanism underlying perceptual judgments and sensorimotor synchronization with simple rhythmic stimuli (Elliott, Wing, & Welchman, 2010; Wing, Doumas, & Welchman, 2010; Wright & Elliott, 2014; Sejdić, Fu, Pak, Fairley, & Chau, 2012), the contribution of different sensory systems for the synchronization of side-by-side walkers is scarce and ambiguous (Nessler & Gilliland, 2009; Zivotofsky et al., 2012). This can be partly related to insufficient systematic manipulation of the sensory channels. For treadmill walking, many available studies did not control for the impact of sensory input (van Ulzen et al., 2008; van Ulzen, Lamoth, Daffertshofer, Semin, & Beek, 2010) and others provided shortcomings in their manipulation. For instances, in one study, participant pairs walked on individual treadmills positioned next to each other, and they were instructed to synchronize with the partner while auditory and visual sensory signals were masked (Nessler & Gilliland, 2009). The results did not reveal clear differences between the sensory conditions. According to the authors, this is because vibrations from the surface floor might have been sensed and earplugs did not prevent entirely sounds produced by the walking partner. Moreover, findings in treadmill walking cannot be generalized to natural walking. Gait dynamics in treadmill walking are different, reflected by, for example, less variable stride time intervals, reduced stride time, stride length, speed, arm-leg coordination (Frenkel-Toledo et al., 2005; Carpinella, Crenna, Rabuffetti, & Ferrarin, 2010). Also, treadmill walking reduces optic flow and provides additional pacing cues (Zivotofsky et al., 2012).

The first study investigating synchronization in over-ground walking used qualitative methods (i.e., video analysis & observer ratings) and showed that hand-holding rather than visual or auditory cues from the walking partner leads to spontaneous synchronization (Zivotofsky & Hausdorff, 2007). Other studies showed spontaneous gait synchronization when sensing synthesized vibrations during cell phone conversations (Murray-Smith, Ramsay, Garrod, Jackson, & Musizza, 2007) and when walking while being fixed to each other or while the leading walker could be seen (Harrison & Richardson, 2009; Marmelat, Delignières, Torre, Beek, & Daffertshofer, 2014). To our knowledge, there is only one study that controlled for different sensory cues and determined their contribution for the spontaneous synchronization during overground side-by-side walking. In this study participants walked next to each other for 70m while visual feedback was prevented by side-blinders and auditory feedback was prevented by white noise displayed through headphones. Results revealed that tactile (i.e., hand-holding) and auditory information leads in some couples to spontaneous synchronization (Zivotofsky et al., 2012). Also, side-blinders constrain the visual field and therefore visual cues like optical flow.

Thus, in former studies, single modalities such as visual, auditory and haptic were masked but it is often not clear how effective the masking technique was or if other spatiotemporal cues were available. Moreover, sensory manipulations and treadmill walking changed the natural gait dynamics. In addition, most frequently the exact specification of different sensory cues was not the main goal of the study. Some studies were interested in identifying attractor layouts and testing whether spontaneous synchronization emerged through self-organization principles (Dynamical System Approach) (Harrison & Richardson, 2009; van Ulzen et al., 2008, 2010) (see Section 0.1.2). Sensory information was mainly treated as coupling medium without giving much importance to their specific contributions. In addition, while some studies focused on spontaneous synchronization, others focused on intentional synchronization. Intentional synchronization was not investigated by the studies using over-ground side-by-side walking. The underlying control mechanisms might be different in spontaneous and intentional synchronization. Moreover, some studies focused on head movements, angle displacement, discrete or continuous signals (van Ulzen et al., 2010), and the decision to study a particular effector was not explicitly clarified (van Ulzen et al., 2010; Murray-Smith et al., 2007; Nessler & Gilliland, 2009; Nessler, Kephart, Cowell, & De Leone, 2011; Nessler & Gilliland, 2010, 2009; Varlet & Richardson, 2015).

Behavioral studies revealed a strong enhancement when complex, moving, and biological stimuli were presented through multiple modalities (Brooks et al., 2007; Wuerger, Meyer, Hofbauer, Zetzsche, & Schill, 2010; Arrighi, Marini, & Burr, 2009; Thomas & Shiffrar, 2013; Saygin, Driver, & de Sa, 2008; Bidet-Caulet, Voisin, Bertrand, & Fonlupt, 2005). This might be due to an improved disambiguation when using cues from multiple modalities to represent the object in question. Furthermore, neuroimaging studies showed that there are overlapping areas for the processing of biological stimuli and multimodal signals (see Thomas & Shiffrar, 2010; Grossman, Blake, & Kim, 2004; Barraclough, Xiao, Baker, Oram, & Perrett, 2005; Beauchamp, Lee, Haxby, & Martin, 2002; Peuskens, Vanrie, Verfaillie, & Orban, 2005; Saygin et al., 2008) (see Section 0.1.3).

Therefore, it is quite surprising that there seem to be no attempts to test for the effects of multimodal integration when synchronizing with complex biological stimuli in routines like side-by-side walking. This thesis is a first attempt to understand the sensory integration in movement synchronization of side-by-side walkers. We focused here on the intentional synchronization. Our main question is whether sensory cues are integrated in an optimal fashion (according to the MLE model, see Section 0.1.4) when trying to synchronize during side-by-side walking. To answer this question, we discuss how each of the sensory systems might contribute to the synchronization process (see Section 0.1.3 & Section 0.1.5), and subsequently, we describe four experiments in which it was tested for the cue integration mechanism by manipulating auditory and visual cues from the walking partner.

The previous deliberation is our principal argument for the motivation of this thesis and the remainder of the introduction provides a detailed description related to each of the individual argumentation points. One may read this thesis by complementing the arguments above with additional information from each section (as referenced).

The remainder of the introduction is structured as followed: First, we introduce the basic synchronization approaches. This provides a general framework that helps to position our approach between the existing theories and models of sensorimotor synchronization (see Section 0.1.2).

Second, we discuss which sensory cues the individual can use to estimate the temporal position of the partner. We illustrate that perception is ambiguous when based on individual sensory cues and that their combination reduces this ambiguity (see Section 0.1.3). Here, it is also discussed how different cues could be combined (see Section 0.1.4).

Third, we discuss which sensory cues an individual receives about one's movements and how these could be matched with the signals from the movements of the walking partner. In order to do so, we refer to the control mechanism enabling a stable walking in a dynamic environment. We introduce the relevant sensory systems, which also should be employed during the synchronization task (see Section 0.1.5). Then, it is described in detail which sensory cues are available during side-by-side walking (see Section 0.1.5) and which complications arise when these cues have to be matched with the signals of the partner (see Section 0.1.6). Finally, we derive our experimental hypothesis and explain how we could test for the validity of the MLE model during the intentional synchronization of side-by-side walkers (see Section 0.1.6).

0.1.2 Synchronization Models

In this section, we describe the main frameworks explaining interpersonal synchronization.

Interpersonal movement synchronization is basically approached within two frameworks. While dynamical system models refer to the general principles underlying the systems of any kind, models based on information processing principles make assumptions about the specific mechanisms of each system. In the following, both frameworks are shortly introduced and then we explain why the present thesis takes an information processing approach.

Dynamical System Approach

In 1665, Christian Huygens observed that by placing two clocks closely together, their pendulums enter in synchronous swinging (Strogatz, 2003). He assumed that the clocks exchanged mechanical energy through vibrations of a common support surface that were caused by the mechanical clockwork. The marine alga Gonyaulax is called “living lantern of the sea”, its name coming from the visible red glowing that illuminates seabed and coral reefs. This bioluminescent property is detectable because many thousands of cells are synchronously active; they follow the same circadian rhythm (see Allison & Draghici, 2013). In North America and Southeast Asia, fireflies communicate through rhythmic bioluminescent flashes. At night, some species illuminate entire mangrove trees by a pulsating yellow light established through synchronized flashing (Moiseff & Copeland, 2010).

According to Pikovsky, an essential feature of synchronization is that individual rhythms become adjusted via their interaction. A rhythm can be formally described as an oscillator. This is a model describing the change of a system that periodically returns to its initial state. The time between two initial states defines a cycle. The choice of the initial state is arbitrary. Oscillators that are coupled mutually influence their cycles. If coupling between oscillators is strong enough, states in the cycle actually “occur” at “the same time” (Pikovsky et al., 2001, p. 2).

Simple biological and mechanical systems are not the only systems that synchronize. It appears to be a fundamental phenomenon that can emerge between oscillators of all living and non-living objects. Within the human nervous system, heart cells fire in synchrony for establishing a regular heartbeat. The simultaneous discharging of billion of cells causes epilepsy. Several processes of the respiratory system synchronize in order to transport oxygen to the cells for energy production. Humans can synchronize oscillating finger movements. Depending on the cycling frequency, in-phase and antiphase synchronization is more stable. The emergence of these particular stable states is described by the coupled oscillator model (see the Haken-Kelso-Bunz Model of Haken, Kelso, & Bunz, 1985).

Similar models have been used to describe the synchronization of variables established by the nervous systems of two individuals. Women living together may synchronize their menstrual cycles (Strogatz & Stewart, 1993) and the audience of a concert enters into synchronized clapping. This seems to occur spontaneously without any intention of the individuals to coordinate with each other (see e.g., Schmidt & O'Brien, 1997). Further findings are—as mentioned in Section 0.1.1—that two interacting individuals synchronize oscillating movements. The fact that the particular effector seems widely arbitrary (e.g., Coey et al., 2011; Demos et al., 2012; Issartel et al., 2007; Oullier et al., 2008; Richardson et al., 2007, 2005; Schmidt & O'Brien, 1997; Zivotofsky et al., 2012; Zivotofsky & Hausdorff, 2007) and that the synchronization dynamics of such dissimilar systems such as pendulum clocks, simple biological organisms, and humans are very similar (see e.g., Schmidt & Richardson, 2008; Kelso, 1997) is used as evidence for the Dynamical System Theory. In order to elucidate how it explains spontaneous and intentional synchronization, we first introduce its basic concepts.

Individuals and environment are considered as a complex interacting and dynamic system that imposes constraints on the possible behavior repertoire. Behavior is a consequence of these constraints and random fluctuations not controlled by the individual. The principle behind this mechanism is called "Self-Organization" (Warren, 2006; Kelso, 1997; Schmidt, Fitzpatrick, Caron, & Mergeche, 2011). Kelso (1997) illustrated Self-Organization by the motion of fluids: When putting oil into a cold pan, oil molecules remain "chaotically" located next to each other. When heating the pan, cooler molecules drop and warmer molecules rise owing to a changing density gradient. It leads to a rolling motion of the oil bubble.

This motion can be modeled by a dynamical system. A dynamical system describes the change of the state of a system over time. They are often nonlinear implying that sudden qualitative changes can occur depending on the particular parameter values of the system. To describe the behavior of dynamical systems, one usually employs differential equations. These relate functions of variables to its changes. By plotting different solution of a differential equation, it can be observed that all solutions tend toward particular regions in the variable space. These are called stable states or attractors.

When heating a pan, the constant rolling motion in particular directions is an attractor. It is the result of a simple principle: molecules with higher density drop and molecules with lower density rise depending on the environmental condition. The particular environmental condition in the present example is here the temperature. The great number of different trajectories on which the oil bubble can move illustrates how simple principles can lead to behavioral complexity. This rolling motion illustrates the Self-Organization principle (Kelso, 1997).

The motion of a damped pendulum is an example for an attractor in a mechanical system. After perturbing the pendulum, it returns to a halt (i.e., fixed-point attractor) due to the restoring force of gravity. Another example is the pendulum motion of a grandfather clock. It has a restoring force proportional to displacement. When it is perturbed, it returns to the limit cycle (i.e., limit cycle attractor). This is a specific trajectory in the space of all possible states (i.e., state space).

When placing two grandfather clocks closely together, their pendulums enter after some time into synchronous swinging. The pendulums normally swing in relative antiphase. The relative phase is the difference between the phases of each oscillator within their limit cycle. When being simultaneously at the same position within the cycle, this is in-phase synchrony; relative phase = 0 degrees. When being 180 degrees apart, this is anti-phase synchrony; relative phase = 180 degrees. The relative phase can be used as a variable that describes the dynamical state established by the motion of the pendulums (Warren, 2006).

The relative phase of the pendulums of two closely located grandfather clocks can be modeled by two harmonic oscillators that are coupled (see Appendix A.1 for a detailed description of the model). They might synchronize when energy is transferred. Always when a pendulum motion direction changes, kinetic energy is transferred from the surface to the other pendulum. When time increases, the system–established by the coupled grandfather clocks–converges to the relative phase of 180 degrees.

The dynamical system approach assumes that interpersonal movement synchronization emerges because two interacting persons establish a complex system that behaves according to the same general laws as the above-described mechanical systems (i.e., grandfather clocks) (Schmidt & Richardson, 2008). Mostly, the dynamical system approach proposes deterministic mathematical models and does not specify the mechanisms of the particular system in question (see Colling & Williamson, 2014). It models rhythmic processes of different matters by the dynamics of a pair of coupled oscillators but does not specify the particular coupling medium. Yet, already Huygens referred to the physical facts of the system in order to explain pendulum synchronization as an exchange of mechanical energy through a common support surface. Similarly, this thesis attempts to provide mechanistic explanations about the particular system. The system is composed of two interacting humans that synchronize repetitive movements because they are coupled via their sensory systems. The particular processes that underlie the coupling of the sensory systems being the main focus of inquiry.

The Information Processing Approach

The Information Processing Approach provides mechanistic explanations about interpersonal movement synchronization. In a mechanistic explanation, the system is decomposed into interacting processes and its effects are specified. We here provide firstly a simplified overview of general action control models in order to introduce the Information Processing Approach of movement synchronization.

General action control

It is assumed that interacting with the environment is possible because the individual has cognitive representations of the environment and of the effects actions have on the environment. These representations are established through a complex interplay of perceptual and memory processes. Actions arise from the computation of motor commands. In order to compute a motor command, the CNS makes use of these representations

(Grush, 2004; Hommel, Elsner, Morsella, Bargh, & Gollwitzer, 2009; Hommel, 2009).

The underlying mechanism is called “Inverse Model”. This is a function that maps a state vector, such as the coordinates of the spatial system, on the necessary neural commands. Then, commands activate the neuromuscular system. The particular activation pattern of the muscles finally moves the joints. These mechanisms enable that an effector can be moved to the desired position in space. In order to correct movements, the model is extended with a feedback loop. In addition, because sensory feedback from movements and action effects arrives with delay and the sensory systems measures with noise, two mechanisms were suggested. A) A “Forward Model” uses an efference copy of the motor command in order to simulate sensory effects. This is based on action-effect representations developed through experience. B) Signals registered by the sensory systems are continuously integrated with the predicted signals of the Forward Model (Grush, 2004).

The Corollary Discharge Theory of von Holst, 1954 (see Goldstein, 2008) may provide an illustration of this mechanism. For stable perception during movement, signals from the retina are integrated with signals from the movements of the eye in order to distinguish between moving environment and moving eyes (see Kandel, Schwartz, & Jessell, 2000). Due to delay, this compensation may be based on incoming sensory signals and efference copies of the motor commands for the eye movement (see Goldstein, 2008). The corollary discharge is also coupled with proprioceptive-vestibular interactions, which allows the brain to distinguish actively generated from passive head movements (Angelaki & Cullen, 2008). In a complex and quickly changing environment, the organism may place more weight on efference copies of motor commands. According to this approach, for action control, noisy and delayed sensory signals can be partially compensated by predictions enabled through action-effect representations (Grush, 2004; Berniker & Kording, 2011).

Spontaneous interpersonal synchronization. In accordance with this view, the "Common Coding Approach" (Hommel, Müsseler, Aschersleben, & Prinz, 2001) suggests that spontaneous interpersonal movement synchronization emerges because two individuals simultaneously execute the same or similar motor programs mediated by action-effect representations. The main argument is that these representations are activated when perceiving action effects or performing the action (Sebanz, Bekkering, & Knoblich, 2006). Rise to these assumptions is given by neurophysiological studies showing that observing actions of others automatically activates motor areas that are responsible for producing similar actions. The Mirror Neuron (MN) system is suggested to be the underlying neural correlate (Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992). MNs are located in the premotor cortex and the posterior parietal cortex and fire during action execution and when observing somebody else performing similar actions (Iacoboni, 2009). Most studies found MNs that respond to goal-directed hand movements like the grasping of objects (Cattaneo & Rizzolatti, 2009). Nevertheless, neurons also have been found that respond to complex biological stimuli (see Gallese & Goldman, 1998). It indicates that MNs could code far more complex motion than previously thought. MNs were identified in the animal cortex. For humans, direct evidence is lacking. However, behavioral studies provided much indirect evidence (Morsella, Bargh, & Gollwitzer, 2009). It was for instances demonstrated that the observation of actions interferes with the execution of incongruent actions (i.e., Congruency Effect). This was attributed to MNs activity (see e.g., Kilner, Paulignan, & Blakemore, 2003; Gazzola & Keysers, 2009).

Intentional interpersonal synchronization. As mentioned in Section 0.1.1, the ability to intentionally synchronize movements is investigated with the Sensorimotor Synchronization paradigm (SMS). In SMS, the individual is required to synchronize movements, which are often finger taps, with events, which are usually delivered periodically as metronome sounds or light flashes (see Repp, 2005).

Most studies implementing SMS are based on the Information Processing Approach of action control. Synchronization is supposed to be achieved by a strategy, called event-based timing. When the motor response and the onset of a rhythmic external event are perceived as asynchronous, error correction commands can be computed that reduce the asynchrony in the succeeding event-response pairs. It is assumed that this requires a cognitive representation of the temporal structure of the events and the movement generating process (Wing & Kristofferson, 1973; Schulze & Vorberg, 2002; Repp, 2005).

In a seminal study, Schulze and Vorberg developed the Linear Phase Correction

Model (LPC) for the SMS of finger taps with metronomes (see Equation 1). This model accounts for the serial dependence of asynchronies between responses and metronome events (Schulze & Vorberg, 2002). LPC is a cognitive model that makes assumptions about the cognitive mechanism underlying synchronization. It models the asynchrony between response and metronome event as a linear function of a) the preceding asynchronies, b) the internal (cognitive) representation of the metronome cycle called "Time Keeper", c) motor delays and perceptual delays, and d) the actual metronome cycle.

Equation 2: the Linear Phase Correction Model (LPC) (Schulze & Vorberg, 2002).

$$A_{n+1} = (1-\alpha)A_n + T_n + M_{n+1} - M_n - C_{n-1}, \tag{1}$$

where A_{n+1} is the asynchrony at cycle $n+1$, α is the correction gain, T_n is the Time Keeper at cycle n, M_{n+1} is the perceptual and motor delay at cycle $n+1$, and C_n is the metronome event at cycle n. T and M are considered random variables ($T \sim NV[\mu_T, \sigma_T], M \sim NV[\mu_M, \sigma_M]$). As in Wing et al. we disregarded here the motor noise (Wing, Endo, Bradbury, & Vorberg, 2014).

Support for the model is provided by experiments using tapping and intermittent circle drawing (K. Torre & Delignières, 2008; Zelaznik, Spencer, & Ivry, 2002). Usually, successive movement cycles show weak to absent positive autocorrelations over several lags (long-range) and a large negative lag-1 autocorrelation resulting from the error correction and motor variability (Schulze & Vorberg, 2002). The model parameters can be estimated based on the serial dependence of the asynchronies. The model was extended by a period correction mechanisms for frequency changes, adaptive metronomes (Repp & Keller, 2008), and interpersonal coordination (see Wing et al., 2014; Honisch, Elliott, Jacoby, & Wing, 2016).

An advantage of this approach is that synchronization variability is decomposed into different components. It may originate from uncertainty in a) perceiving the external event and the own movements, b) cognitively representing the cycle, and c) planning and executing appropriate movements. Thus, within-subject variability arises from noise inherent to the CNS. We presented here the LPC model in order to illustrate how the particular information processing mechanisms can be specified. This distinguishes it from the Dynamical System Approach, which assumes similar equations for the synchronization dynamics of e.g. mechanical pendulum clocks, fireflies, and human movements.

As far as we know, there are no attempts to specify information processing mechanisms underlying synchronization during walking (see e.g., Nessler et al., 2009; Nessler & Gilliland, 2009, 2010; van Ulzen et al., 2008; Zivotofsky & Hausdorff, 2007; Zivotofsky et al., 2012). We identified the system in question as two humans who are coupled via their sensory systems. We evaluate the synchrony of repetitive movements during side-by-side walking, to investigate the involved sensory integration mechanisms. As such, the present dissertation mainly takes an Information Processing Approach.

0.1.3 Sensory Cues From the Walking Partner

In this section, we show which cues are provided by the walking partner; that individual cues often provide ambiguous information; those perceptual ambiguities are modality dependent; and that the integration of multiple cues from different modalities reduces these ambiguities.

As mentioned in Section 0.1.1, multiple sensory systems seem to be involved during side-by-side walking. To synchronize movements with an interaction partner, the individual has to estimate the temporal position of the body segments of the partner. If there is no direct physical contact, this can be based on information from the auditory system and/or the visual system. While, for instances, auditory cues provide regular and discrete information about the duration of the movement cycles, indicated mostly by the sound of the heel strikes, visual cues provide less regular but continuous information about the body position and displacement (e.g., shoulders or head up-down cycles). For this reason, we identified the features that seem most relevant as the a) relative distance to the partner and the dimensions of the partner's body segments (i.e., spatial dimension), b) discrete cues that signalize the duration of the partner's movement cycle (i.e., temporal dimension), and c) the continuous displacement (i.e., motion) of the body and its segments (spatiotemporal dimension). We do not claim that these features are exhaustive. There might be other features that could be used to achieve movement synchronization. The main goal of the presentation of the selected three features is to introduce the concept of cue integration.

Cue Combination Reduces Perceptual Ambiguities

Each sensory system might capture a particular feature more or less precisely and accurately. In order to estimate these spatiotemporal features by cues delivered through a single modality (auditory or visual), the CNS confronts obstacles that depend on the structure of the visual and the auditory system and on the physical properties of the visual and the auditory cues. We here illustrate that both sensory systems are better suited for the estimation of particular stimulus features and that the combination of multiple sensory cues reduces the perceptual ambiguities that arise when estimating stimulus features based on single cues and sensory systems.

Spatial Perception. The photoreceptors in the retina are sensitive to a certain spectrum of electromagnetic energy that radiates as a wave. Electromagnetic energy with a wavelength between 400 to 700nm is considered as visible light. It stimulates the retina at certain positions depending on the place where light is reflected from the partner's body. This monocular retinal cue has to be combined with further cues to unambiguously perceive one's distance to the partner and the distance between referential points on the partner's body. For instances, through deletion and accretion—limbs that are farther away, gradually become occluded and reappear depending on one's own movements and the movements of the partner—cues are provided about the relative distance of the limbs to the observer. This is complemented by motion parallax: the part of the body of the partner that is more proximate moves faster. Relevant cues about distance may also be provided by the ciliary muscle that stretches the eye for focal vision and by binocular disparity, which is the difference between the projections on both eyes. Nevertheless, most spatial relations can be retrieved relatively directly by computing the distance of the stimulated positions on the retina (Goldstein, 2008).

In contrast, an auditory spatial perception only exists when multiple cues are combined. This owes to several factors: First, sound is the change of pressure in a medium. This is for human hearing usually air. It propagates as a mechanical pressure wave that is hearable at wavelengths between 20Hz and 20.000Hz. Similar to light, sound is reflected by the walking partner's body. Different to light, sound cannot travel through vacuum and it reverberates. This means that sound signals give rise to multiple sensory inputs before decay. Although reverberation serves as a cue to perceive distance, it degrades azimuth judgments (Hartmann, 1983). Azimuth is one of the three directions of sound localization dimensions. Azimuth is the horizontal displacement direction, elevation is the vertical direction, and distance is the direction from close to far (Goldstein, 2008).

Second, the sound intensity level registered at the eardrum depends on the physical distance of the sound source (e.g., where the partner's heel strike occurs) relative to the ear. It can be used as auditory distance cue. However, because it also depends on the acoustic power of the sound source (e.g., how strong the ground is hit by the heel) (Zahorik, Brungart, & Bronkhorst, 2005), it provides ambiguous distance information.

Third, sound spectral cues can be used for the localization of sound sources. Spectral cues are the differences in the distribution of frequencies within the outer ear. It was shown that by removing some frequencies of the sound, perception was degraded for frontal and lateral sounds (Kopčo & Shinn-Cunningham, 2011). Because the walking surface quality might change the spectral properties of the sound, spatial information delivered by this cue is vulnerable to structural differences of the environment (e.g., treadmill affect most of these cues).

Fourth, auditory spatial information can be derived from the combination of binaural cues. When presenting sound on the lateral side of the head, the ear that is farther away from the sound source receives the signals delayed and attenuated. It causes Interaural Time Differences (ITD) and Interaural Level Differences (ILD). The temporal difference between the left and the right ear (ITD) informs about the location of the source, mainly for low-frequency sound. The sound level difference between the left and the right ear (ILD) informs about the location of the source, mainly for high-frequency sound. In addition, ILD is affected by the distance of the sound source (Brungart, Durlach, & Rabinowitz, 1999). In spite of the fact that both cues combined provide reliable sound information for azimuth, they are ambiguous for elevation. The cone of confusion is the axis parallel to both ears. Sounds presented on this axis cannot be localized adequately. Even though there are monaural spectral cues that provide information for elevation (Blauert, 1997; D. M. Thompson, 2005), these cues do not provide reliable information for sound sources when located close to the individual (1-15m) (Kolarik et al., 2016).

There are also dynamic cues for spatial perception. Acoustic tau is the rate of change in sound level. Motion Parallax is here the change in angular direction of sound. Its exact effects are yet not clarified. Nonetheless, it was claimed that for distances within the peripersonal space, these cues rather impair perceptual judgments (Kolarik et al., 2016). Peripersonal space is the space immediately surrounding our bodies (Rizzolatti, Fadiga, Fogassi, & Gallese, 1997).

Concluding, there are several visual and auditory cues for estimating the position of the partner's body and segments. Individuals are usually more accurate and precise when using visual cues compared to auditory cues. At distances within peripersonal space—what is usually the case during synchronization tasks—many of the above-presented cues seem to provide ambiguous information. In order to reduce this ambiguity, multiple cues have to be integrated.

Temporal perception. In order to explain how humans perceive stimulus duration (i.e, the perceived time interval between two succeeding events), many authors refer to the underlying information processing mechanism (i.e., pacemaker-accumulator model). For instances, it was suggested that the CNS contains an internal time-keeper (see Section 0.1.2) (Schulze & Vorberg, 2002). The "pacemaker"—which is a specific component of this time-keeper identified within the basal ganglia—is supposed to emit pulses at a regular rate during stimulation (presentation of stimulus). Counting its accumulated pulses gives rise to the elapsed time of stimulus presentation. The perception of temporal intervals might differ depending on which cues are presented and it is also affected by the cognitive and emotional state of the individual (Grondin, 2010). Although the exact functioning of the pacemaker mechanisms is not clarified (Grahn, 2012), such mechanisms suggest that time perception is a cognitive construction rather than a direct representation of physical stimulus durations. Then, temporal sensitivity might increase through the accumulation of evidence from any sensory cue. Consistently, it was shown that the combination of both auditory and visual cues increased the temporal sensitivity (K.-M. Chen & Yeh, 2009).

Although time is considered amodal because it is distributed over the cerebral cortex and comprises large overlapping visual and auditory cortical areas, it was shown that the temporal sensitivity of the auditory system is greater than that of the visual system. The auditory system could involve pacemakers that emit pulses at higher rates than the visual system (Zélanti & Droit-Volet, 2012; Grahn, 2012). Another account proposed that the timekeeping mechanism functions by employing cognitive resources and that the stimulation of the visual system requires more cognitive resources than the auditory system (Zélanti & Droit-Volet, 2012).

Another cognitive capacity related to time perception is rhythm and beat perception (Su, 2014; Burger, Thompson, Luck, Saarikallio, & Toiviainen, 2013; Drake, Jones, & Baruch, 2000; Merker, Madison, & Eckerdal, 2009; Su & Pöppel, 2012). Rhythm is defined as the pattern of temporal intervals in a sequence of events. Beat is defined as the perception of regular events in this sequence (Grahn, 2012). Rhythm and beat perception requires that the individual makes predictions of the upcoming event onsets based on the time differences of the preceding events. Thus, it requires temporal sensitivity because temporal intervals must be estimated. Consistent with the previous findings, it was demonstrated that interval perception is superior when the events are provided through the auditory modality and that it improves with multimodal cues (Su, 2014; Burger et al., 2013; Drake et al., 2000; Merker et al., 2009; Su & Pöppel, 2012). There are similar findings in sensorimotor synchronization. Finger tapping is more precisely synchronized with rhythmic and discrete events when they are provided through the auditory modality. Moreover, when synchronizing with discrete visual events, such as light flashes and an auditory distractor sequence is presented simultaneously but has to be ignored, the distraction is stronger than vice versa (see e.g., Repp & Su, 2013).

Motion perception. Another essential feature for estimating the position of the partner is to perceive the displacement of the body and its segments over time, called motion. Motion can be perceived through the auditory and the visual modality.

In order to perceive motion through the auditory modality, the CNS must extract dynamic cues from the time difference between the left and the right ear (ITD) (Kandel et al., 2000). Yet, for a less ambiguous motion perception, this requires a combination with spatial cues. For perceiving motion through the visual modality, information has to be extracted from the dynamics of retinal stimulation. The retinal projection must then be integrated with local motion cues into a global motion representation (Burr & Santoro, 2001). Moreover, retinal stimulation over time has to be combined with efference motor command signals of eye movements and vestibular signals to dissociate the moving environment (the moving partner) from one's own eye and head movements (Goldstein, 2008).

Moving visual stimuli seem to affect auditory motion perception to a greater extent than vice versa (Alink et al., 2011), indicating a superior motion sensitivity of the visual system. Moreover, these mechanisms highlight that cues from different modalities could be combined to improve motion perception. For instances, step sounds give rise to cadence, which could improve velocity and motion direction perception of continuously visible motion of the body. Consistently, studies demonstrated that the judgments of the time to arrival of a moving object to a target position (Wuerger et al., 2010) and velocity judgments (Bentvelzen, Leung, & Alais, 2009) were improved through the combined presentation of audiovisual cues.

Here, we illustrated that features on the spatial dimension are less ambiguously represented by visual cues and features on the temporal dimension are less ambiguously represented by auditory cues. When an event is presented on both modalities but in conflict, individuals tend to shift their estimates towards the event that stimulates the better suited sensory system. These effects are known as the (spatial and temporal) ventriloquism effects and were more recently called asymmetric cross-modal biases. When presented with temporally or spatially incongruent stimuli on different modalities, one modality affects the perceptual judgments to a greater extent than the other modality (Bertelson & Aschersleben, 2003). Traditionally, the asymmetric cross-modal bias was attributed to the resolutions of the sensory systems (see e.g., Welch, DutionHurt, & Warren, 1986). But as we illustrated above and is further discussed in Section 0.1.4 and Section 0.1.4, the interactions causing cross-modal biases seem far more complex. Most importantly, we here illustrated that the perception of spatial, temporal, and spatiotemporal environmental features is ambiguous when single cues are not combined across modalities. In the next section, we further specify these multimodal combination mechanisms.

Multimodal Perception

Indeed, the presentation of additional cues cannot only disambiguate motion perception but may be “the norm rather than an exception" for our daily perceptual experiences. An example is the bouncing ball illusion (Metzger, 1934). When visually displaying two balls that follow linear motion pathways that cross each other on half the way, an ambiguous motion perception is created: Observers alternately judge the motion trajectory of the balls as crossing each other or as bouncing against each other. When adding a sound to the moment when the balls cross each other, the observer experiences an unambiguous perception of two bouncing balls (Shimojo & Shams, 2001). Another example is the McGurk Effect: The presentation of incongruent visual cues and auditory cues of a spoken syllable results in the auditory perception of a new syllable that was not presented on each individual modality alone (Massaro, 1998). These are examples of how visual motion perception is modulated by the presence of auditory cues (bouncing ball illusion) and how speech sounds are modulated by the presence of visual motion cues (McGurk Effect). Both examples illustrate that the combination of multimodal cues may underpin most of our perceptual experiences.

What are the physiological mechanisms behind multimodal integration? By stimulating individual sensory receptors, receptor potentials are transduced into APs and travel through nerves to higher processing levels. Signals converge over time and/or space at each level of the processing hierarchy. First, the receptive fields of a single modality become larger. At later processing stages, different features are combined through the synchronous firing of neurons within and between different cortical areas. If a stimulus leads to the depolarization of a group of neurons at a higher cortical area coding a particular environmental feature, this should be perceived. However, due to naturally occurring noise, the CNS might interpret weak APs as noise. Moreover, some single cues lead to the depolarization of the same neuronal groups. In this way, different cues can lead to similar perceptual experiences (ambiguities). Stimulating the receptors of two modalities leads to APs that travel through two sensory pathways. The signals become combined at higher brain layers. This disambiguates noise from sensory experience. Yet, higher brain processing centers also may receive signals from other cortical areas such as the hippocampus. So, the experience of a particular event/or environmental feature is not a sum of individual sensory signals but a reconstruction of a cognitive representation through the integration of multiple afferent and efferent signals. Consequently, multimodal integration can lead to qualitatively different responses to multimodal modal stimuli (behavioral or neural) compared to each unimodal stimulus (Kandel et al., 2000) and it can change with experience.

Neural correlates of multisensory integration. There is plenty of neuroimaging and neurophysiological evidence for multimodal integration. Many sensory signals of the same modality may converge within the relay nuclei in the thalamus. The within-modality integrated signals project then on the primary cortical areas. At cortical level, signals from the different primary cortices are integrated. These multisensory areas are distributed over the cortex (Stein & Stanford, 2008). In addition, multisensory neurons were found in the basal ganglia (putamen) for visual and somatosensory signals and between the occipito-paretal space and the occipito-temporal space. But their interaction and function are yet not well understood (Keetels & Jean, 2012). Another relevant structure seems to be the vestibular nucleus that integrates vestibular signals, proprioceptive signals, and visual signals. It plays besides others a crucial role in the perception of self-motion (Cullen, 2012, 2016). There are also areas for each modality and integration areas for different sub-modalities. For example, proprioception is processed at Broadman area 3a and haptic perception at Broadman area 3b. Its integration occurs at Broadman area 2 (Kandel et al., 2000). The most prominent neural structures that were identified to process multisensory information are a) the superior colliculus, b) the superior temporal gyrus and sulcus, and c) the visual and auditory association areas (e.g. for SC Meredith & Stein, 1983; Meredith, Nemitz, & Stein, 1987; Meredith & Stein, 1996; Calvert, 2001; Barraclough et al., 2005).

Multimodal integration of biological motion stimuli. The previous deliberation (about the integration of multimodal cues) was based on the findings of studies implementing simple stimuli (i.e., bouncing balls, metronome clicks, light flashes). Our current question is whether the integration of multimodal cues is required when synchronizing movements with a partner during side-by-side walking. We assume that the interaction with a moving partner is perceived through the integration of multiple auditory and visual cues (combined with top-down processes) rather than processed separately.

More robust conclusions may be obtained by considering findings of studies investigating the multimodal integration of more plausible biological motion stimuli: A popular paradigm to study biological motion perception implements point light displays. In 1973, Gunnar Johansson discovered that humans recognize with great ease human motion from the visual cues of the main joints of the body. Point light displays were then constructed by attaching visible dots to the joints. When the body cannot be seen, as in a dark room, this technique dissociates contour from motion. The moving dots provide dynamic (i.e., translation, velocity, and acceleration) and structural information because the distance between the dots is fixed. It was shown that humans are highly sensitive to this information and that they can derive complex actions from it (Dittrich, 1993; Norman, Payton, Long, & Hawkes, 2004). The ease of recognition, even when experimentally generating visual noise (i.e., additional independently and randomly moving dots), indicates that the visual system is very effective in extracting biological motion (Cutting, Moore, & Morrison, 1988; Thornton, 1998). Recognition declined when the point light displays were inverted (Troje & Westhoff, 2006) or when the dots were presented at random positions (i.e., scrambled). It was proposed that in these conditions the CNS does not integrate the spatiotemporal properties of the point light displays (Thornton, Rensink, & Shiffrar, 2002). There are also studies demonstrating the distinctive features of biological motion perception for the auditory system. Based on the sound properties of human movement, participants could judge walking direction, identified postures and gender, and distinguished between drumming beats when produced by a human or synthesized artificially (see Pizzera & Hohmann, 2015).

Importantly, there is evidence that signals of biological motion stimuli become integrated across different sensory modalities. In some situations, the perception of Point Light Walkers (PLW) is ambiguous. A PLW is a point light display of a human walker. The ambiguous perception is illustrated by the frontal bias. The frontal bias means that the walking direction of a PLW cannot be judged adequately and when presenting a PLW in the frontal plane, it results in a “bistable” perception of the walking direction. That is, the PLW is sometimes judged walking towards the observer and sometimes judged walking away from the observer (Vanrie, Dekeyser, & Verfaillie, 2004). It was shown that additional auditory cues reduce this ambiguous perception (Mendonça, 2012). Also, point light displays were recognized faster when presented through both the auditory and the visual modality (Arrighi et al., 2009) and additional auditory cues affected the gender judgments based on visual PLWs (van der Zwan et al., 2009).

There is also evidence that cues of different sensory modalities are integrated when judging the motion of biological stimuli. In one study it was assessed whether audiovisual cues from moving biological stimuli improves velocity judgments (Mendonça, Santos, & López-Moliner, 2011). Participants judged if a test stimulus was faster than a standard stimulus. As standard stimuli, it were used PLWs (visual) combined with footstep sounds (auditory) representing a human walking at constant velocity. As test stimuli, the PLWs were presented alone, the foots steps sounds were presented alone, or both were combined. When presented in combination, the visual and the auditory stimulus signalized the same walking velocity (congruent condition) or they signalized different velocities (incongruent condition). The degree of incongruence was systematically manipulated. Results revealed that discrimination performance was highest with bimodal congruent stimuli. Because the global motion of the PLWs and the step frequency provide cues for walking velocity, it suggests that the CNS also combines qualitatively different cues across modalities (Mendonça et al., 2011; Silva et al., 2013).

It must be highlighted that the mechanism underlying the multimodal integration of biological motion cues could differ from the mechanisms integrating cues of simple stimuli: Sounds increased visual PLW detection when being spatiotemporal congruent and/or meaningful. When a) the sounds and the visual cues delivered incongruent distance cues, b) the PLW was scrambled (Brooks et al., 2007; Wuerger et al., 2010; Arrighi et al., 2009) or c) when the sounds were not meaningfully associated with the visual stimulus—i.e. pure tones instead of footstep sounds were paired with PLW—,perceptual judgements did not improve (Thomas & Shiffrar, 2013, 2010). Moreover, multisensory effects were stronger for upright walkers (Saygin et al., 2008). Consequently, the perception of multisensory biological motion, like a human walker, may be processed in a different manner than non-biological moving objects (Bidet-Caulet et al., 2005).

Further evidence is provided by neuroimaging and neurophysiological studies.

Cortical areas that were identified for the multimodal integration, like the superior temporal sulcus and the premotor cortex, are also active during the perception of actions of point light displays (see Thomas & Shiffrar, 2010). The posterior superior temporal sulcus increased activity when observing upright PLWs compared to inverted PLWs (Grossman et al., 2004). In addition, neurons in the superior temporal sulcus were found to be sensitive to audiovisual motion (Barraclough et al., 2005) and they did not respond to the presentation of rigid objects (Beauchamp et al., 2002; Peuskens et al., 2005). Furthermore, listening to footsteps activated the superior colliculus (Bidet-Caulet et al., 2005), auditory cortex, frontal, and parietal areas, which are associated with visual attention, and the posterior superior temporal sulcus (Pizzera & Hohmann, 2015). Some of these areas might serve as a multimodal area with specialized neurons for biological motion.

Another line of research investigated the multimodal integration of biological motion cues for beat perception. It was shown that point light displays can induce the perception of beat and that this does not occur with more simple visual stimuli such as light flashes. It is assumed that beat perception employs the motor system (Su, 2014), and the observation of biological motion stimuli engages the observer's motor system, possibly through MNs activity. Listening to footstep sounds seems to have similar effects (Murata et al., 1997). The additional activity of the motor system during the perception of biologically plausible movements may amplify the differences between nonbiological motion and biological motion perception for the auditory and the visual modality (see Saygin et al., 2008).

Multimodal integration also seems to underly the superior performance when synchronizing movements with repetitive multimodal events (SMS). Synchronization variability was usually lower (higher synchronization performance) when the stimuli were multimodal. Studies implementing an SMS paradigm often used finger tapping as response (Repp, 2005; Elliott et al., 2010). Nevertheless, similar effects were found using stepping (Wright & Elliott, 2014) and during walking to simple uni- and multimodal stimuli (Sejdić et al., 2012).

In short, there is behavioral and neurophysiological evidence that multimodal cues are integrated in perceptual judgment and synchronization tasks. Moreover, there might exist essential differences between the multimodal processing of complex, moving and biological stimuli compared to simple, rigid, and non-biological stimuli. The beneficial effects of cue integration should be greater for the former than for the latter.

Bearing the previous discussion in mind, visual and auditory cues from the walking partner should reduce an ambiguous perception of the temporal position of the walking partner. When considering each sensory system individually, they ambiguously provide spatiotemporal cues. For example: consider the case in which the individual focuses on the heel strike onsets of the partner (heel strike synchronization was typically the subject of most studies). In order to estimate the heel strike onsets, watching the feet may provide relatively reliable information about the position of the feet in the walking direction and orthogonal to the walking direction. Yet, due to the particular position of the eyes, it is difficult to perceive the vertical distance between the heel and the surface. This creates an imprecise temporal perception of the exact heel strike onsets. Moreover, if not constantly focusing the partner, synchronization might rather be achieved by predicting the upcoming heel strike onsets based on the preceding gait cycle durations.

Footsteps sounds provide more reliable temporal cues about the upcoming heel strike onsets. This owes to the sharp energy rise time of the sounds produced by the heel strikes. Moreover, the auditory system has a greater sensitivity for perceiving temporal intervals, like the cycle durations (and beat/cadence perception), as discussed in Section 0.1.3. But in order to estimate the exact onsets, the temporal auditory cues should be combined with spatial and continuous motion cues. These can be monaural, like the sound spectrum and reverberations, and binaural cues like Interaural Temporal Differences (ITD) and Interaural Level Differences (ILD). However, these cues provide spatial information indirectly and are prone to error (Brungart et al., 1999; Kolarik et al., 2016; Goldstein, 2008). Therefore, auditory spatiotemporal cues should be combined with visual cues (e.g., motion parallax, occlusion, etc.). For this reason, the discrete and more regular footstep sounds could reduce the uncertainties in estimating heel strike onsets of the walking partner provided by, mostly, peripheral visual cues. Similarly, visual cues should reduce uncertainties in estimating the position and motion of the walking partner.

Consequently, we assume that during side-by-side walking, individuals could integrate visual and auditory cues to most efficiently retrieve the spatiotemporal features of the walking partner for synchronization. But, as illustrated above, auditory and visual cues represent the different features (i.e., position, distance, temporal intervals, & motion) with different precision and accuracy. If the walking partner is perceived through multimodal channels, it must be clarified how (much) each modality contributes to the final percept. This is particularly relevant when both sensory systems provide redundant information about the same feature. In the next section, we approach the relative contribution of each sensory system for the final percept.

0.1.4 Multimodal Integration Strategies

In this section, we describe cue integration strategies. There are different theories of multimodal cue integration that suggest how cues from different modalities are combined by the CNS. A popular and more traditional account is the Modality Appropriateness Hypothesis (see e.g., Welch et al., 1986). A theory that receives currently most empirical support is the Optimal Sensory Integration Theory (or MLE) (see e.g., Ernst & Bülthoff, 2004). Both theories are based on findings in perceptual judgment tasks and were more recently approached with an SMS paradigm to also account for findings in sensorimotor control and timing. We present here both theories and discuss their validity.

Modality Appropriateness

As discussed in Section 0.1.3, the visual system seems to dominate the auditory system on the spatial domain and the auditory system seems to dominate the visual system on the temporal domain. The traditional view is that these asymmetries exist because one sensory modality is more appropriate for delivering spatial or temporal properties of the event (Modality Appropriateness Hypothesis, see e.g., Welch et al., 1986). Thus, when multiple sensory cues are available, a possible strategy to estimate properties of an external event is to give priority to the cues that are naturally more appropriate for the particular task (see e.g., Welch et al., 1986). According to this theory, when an individual tries to synchronize one's steps with the steps of the partner and there are auditory and visual cues for the temporal onsets of the partner's steps, the individual would give more weight to cues from the modality that is naturally more appropriate to represent the temporal onsets. These would be the auditory cues in this case.

However, more recently the Appropriateness Hypothesis was challenged: a possible test for modality appropriateness is to manipulate the reliability with which each modality represents the external event in question. For instances, in a SMS task, participants may synchronize finger tapping with a target sequence while presenting a distractor stimulus simultaneously through another modality. Auditory distractors have usually stronger effects than visual distractors (in finger tapping to metronome events). This is consistent with the previously postulated auditory appropriateness (Bertelson & Aschersleben, 2003; Repp & Penel, 2002). However, when synchronizing with a continuous visual stimulus (i.e. a bouncing ball), synchronization performance becomes almost as good as with the auditory stimulus (Iversen, Patel, Nicodemus, & Emmorey, 2015). Moreover, it was shown that the strength of the distractor effect depends on the degree of expertise with the particular modality (i.e., visual expertise - video gamers; auditory expertise - musicians). Similar findings were made for the inquiry of beat perception (Su, 2014). Overall, asymmetric distractor effects disappear when controlling for the relative precision with which participants can perform with visual and the auditory stimuli (Hove, Iversen, Zhang, & Repp, 2013; Varlet, Coey, Schmidt, & Richardson, 2012; Alais & Burr, 2004).

We conclude that the asymmetric cross-modal bias, which was traditionally thought to be a result of a more appropriate modality, could be experimentally manipulated by changing the properties of the environmental cues (Alais & Burr, 2004; Ernst & Banks, 2002; Andersen, Tiippana, & Sams, 2005; Shams, Ma, & Beierholm, 2005; Elliott et al., 2010; Wuerger et al., 2010). Therefore, the Appropriateness Hypothesis can be abandoned in favor of other cue integration strategies, which look more promising.

The Optimal Integration Theory

The (Bayesian) Optimal Integration theory is currently the empirically best-supported theory accounting for multimodal integration in perceptual judgments and SMS tasks (Stein & Stanford, 2008; Alais & Burr, 2004; Ernst & Bülthoff, 2004; Hove et al., 2013; Mendonça et al., 2011). How does it work? Consider the task of estimating the asynchrony between two events provided through the auditory and the visual modality. This could be, for example, the heel strike of the individual and the partner. To know the asynchrony size, the individual event onsets have to be estimated. This estimation process, based on unimodal sensory cues, can be captured by a likelihood function (Figure 1b).

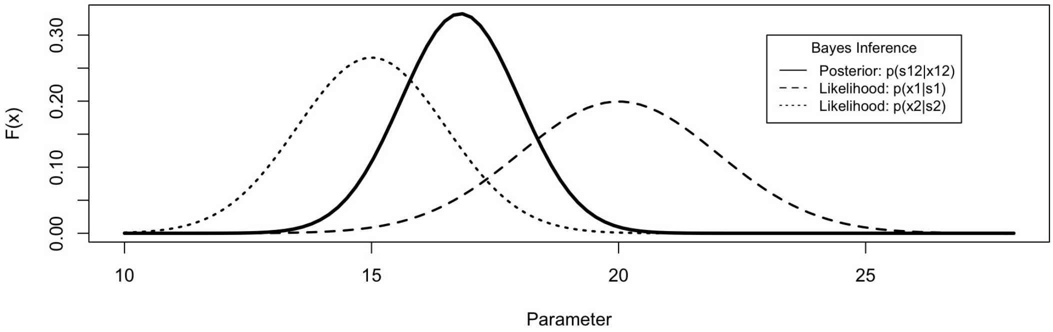
Formally, a likelihood function $L(s \mid x)$ relates the likelihood of an unobserved parameter s (or a vector of parameters) to its assumed measure x (Figure 1a). In order to make assumptions about the underlying probability distribution that generated x (i.e., statistical inference), we search for the s, among a possible range of parameters, that most likely has generated x. It is the parameter where the likelihood function has its maximum. The maximum is obtained by finding the parameter value where the derivation of the likelihood function equals zero. This provides the best estimator for the location parameter of the underlying theoretical distribution. The estimator is therefore called the Maximum Likelihood Estimator (MLE). According to Bayesian inference, two likelihood functions can be integrated. Their product is proportional to a posterior distribution.

$$p(x_{post} \mid s_{post}) \propto p(s_1 \mid x_1) p(s_1 \mid x_2)$$

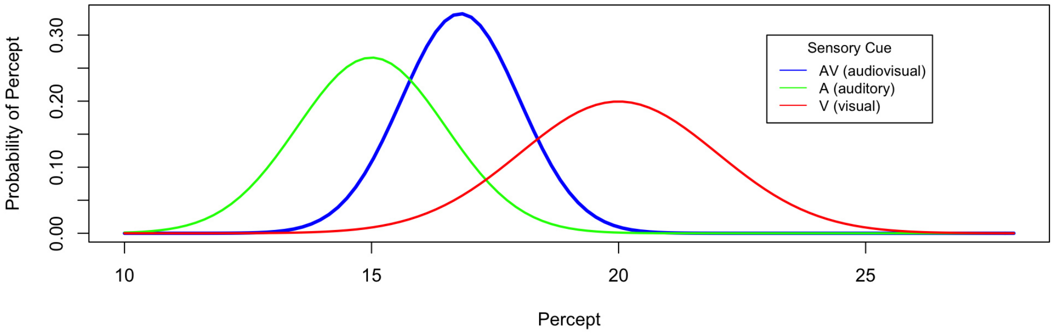
The posterior distribution is the probability of making the measurements x_{post} given the parameter s_{post}. This distribution has the maximum at an onset that lies between the maxima of the individual likelihood functions (Figure 1a).

Consider now that there are measures of an event onset made by two sensory systems. Perceptual estimates are usually modeled by normal distributions ($NV[\mu, \sigma]$) (see e.g., Ernst, 2006). An optimal way of combining them is through their integration. Integrating two normal distributions leads to another normal distribution. This is why the parameters of the posterior distribution (μ_{post} and σ_{post}) depend on the μ and σ of the individual distributions.

The μ_{post} of the posterior distributions is determined by



(a) Bayesian Inference. $p(x1 \mid s1)$ is the likelihood function of measure 1 and $p(s2 \mid x2)$ is the likelihood function of measure 2. The product of two likelihoods is proportional to the posterior distribution $p(x12 \mid s12) = p(s1 \mid x2)p(s2 \mid x2)/p(x1, x2)$. The normalization factor $p(s1, s2)$ can be disregarded because the maxima remain at the same place. For simplicity, we do not assume prior distributions here. These are distributions representing former beliefs about the stimulus that are independent from the actual observation.



(b) Sensory Integration. The estimates based on the auditory cue and the visual cue are integrated according to Bayesian Inference. The combined estimate is based on audiovisual cues and lies between the individual distribution functions. Its distribution is narrower (lower s.d.).

Figure 1: The Maximum Likelihood Integration Theory.

$$\mu_{post} = w_1\mu_1 + w_2\mu_2, \tag{2}$$

$$w_1 = \frac{1/{\sigma_1}^2}{1/{\sigma_1}^2 + 1/{\sigma_2}^2}, \tag{3}$$

$$w_2 = \frac{1/{\sigma_2}^2}{1/{\sigma_1}^2 + 1/{\sigma_2}^2}, \tag{4}$$

where w_1 and w_2 are the weights and $_1$ and $_2$ indicate the likelihood functions 1 and 2. The σ^2_{post} of the posterior distribution is determined by

$$\sigma^2_{post} = \frac{{\sigma_1}^2{\sigma_2}^2}{{\sigma_1}^2 + {\sigma_2}^2}. \tag{5}$$

As can be seen by w_1 and w_2, when integrating two normal distributions, a narrower distributions (i.e., smaller σ) contributes more to the location parameter μ_{post} than a wider distribution (i.e., larger σ^2). Thus, when integrating two normal distributions of the onset estimates of the auditory and the visual modality, the distribution of the integrated onset estimates (audiovisual) lies between these distributions (Figure 1b).

Its exact location depends on the reliability of the onset estimates based on signals from the particular modality. The reliability is defined as the inverse of σ^2, which is represented by w in the formal description. Considering our previous assumptions, the reliability of estimating the auditory event onsets is expected to be higher than the reliability of estimating visual event onsets. However, according to the model, the optimal way of estimating the actual onset is to combine both cue estimates (Figure 1b). In addition, the MLE model describes that σ^2_{post} is always lower than the σ^2 of the individual distributions, which were integrated.

As a consequence, integrating individual distributions of sensory estimates should lead to a narrower distribution of the combined estimate. If perception is underpinned by such a mechanism, then the combined estimate is optimal. There have been several studies showing that the MLE model can describe the multimodal integration effects in perceptual judgment tasks (see e.g., Alais & Burr, 2004; Ernst, 2006; Ernst & Bülthoff, 2004; Mendonça et al., 2011). Yet, in SMS tasks, the estimation of event onsets is only one of several interacting perceptual and motor processes. Variability may arise from sensory registration, timekeeping, and motor planning and implementation (Wing et al., 2010; Elliott et al., 2010). However, considering that the sensory registration is essential in order to minimize asynchrony, it could affect synchronization precision.

Consistently, research has shown that synchronization performance of finger tapping to rhythmic cues could also be modeled by MLE (Elliott et al., 2010; Wing et al., 2010). In one study, participants were instructed to synchronize their tapping with a metronome presented aurally, visually, as a tactile cue, or bimodal (i.e., audio-visual, audio-tactile, tactile-visual). In order to systematically vary the reliability of the cues, the temporal regularity of the event sequences was disrupted by different variability levels. For the low variability condition, the results revealed auditory dominance. When decreasing the reliability of the auditory event, participants seemed to integrate haptic and visual cues in order to reduce asynchrony variability. The reduction was close to the prediction made by the MLE (Elliott et al., 2010).

In short, there is much support that multiple cues become integrated during perceptual judgment and motor synchronization tasks when properties of the environmental events can be retrieved through multiple modalities. The MLE model accounts for most findings. This suggests that individual cues might be integrated based on their reliability with which they represent the event in question, and the Appropriateness Hypothesis seems to reflect a bias created by experience with a particular task, more than multimodal integration itself.

Such a model may also underly sensory cue integration during side-by-side walking. A caveat is that walking is far more complex than most of the presented studies so that the generalization of previously presented findings and conclusions is difficult. Moreover, there are multiple cues from different modalities that the individual receives about one's own movements. These then have to be cross-modality matched with the audiovisual cues from the partner, which might be integrated according to an MLE model. This is discussed in the next section.

0.1.5 Sensory Cues From one's own Movements

Here we discuss how the signals from the movements of the walking partner could be matched with signals that an individual receives about one's own movements. We therefore firstly introduce the different sensory systems of the individual that are employed to control walking in a dynamic environment. The second goal of this section is to highlight its complexity and the obstacles one confronts when experimentally studying synchronization during such complex action tasks.

The control of walking

The control of walking may be described on different levels, whether it is the exchange of signals between cells, the muscles activity, the exertion of forces, or the segments kinematics. A more detailed overview suited to the goals of the present work is given in the Appendix A.2.

For normal walking, limbs must be spatiotemporally coordinated. This may be mainly achieved by the automatic activity of networks of neurons in the spinal cord. These circuits are called Central Pattern Generator (CPG). CPG is a self-oscillating network that generates the basic pattern of alternating activity of extensors and flexors during walking (Pearson & Gordon, 2000).

CPG activity is not sufficient to achieve a stable walking. Walking in a dynamic environment requires a very quick limb coordination and adjustment to stability and environmental demands. This depends on the availability and processing of multiple sensory cues (Cullen, 2012, 2016) via several pathways. CPG activity can be altered by supra-spinal and afferent input. Afferent input mainly comes from the somatosensory system. It involves proprioception, via receptors within the muscles, and haptic perception, via cutaneous receptors. Afferent input allows quick adjustment of walking pattern, via direct reflex pathways and via indirect pathways projecting on neural circuits within the spinal cord (Kurtzer, 2014; Ghez & Krakauer, 2000; Loeb, 1989).

Moreover, input from several sensory systems is processed in supra-spinal areas, i.e., the motor cortex, cerebellum, and the brain stem. In these areas movements are planned based on the integrated signals from sensory systems (e.g., visual, auditory system, vestibular system), intentions and memory processes (Pearson & Gordon, 2000). Supra-spinal signals seem to be relevant when an individual attempts (intentionally) to synchronize one's movements when interacting with a partner. For instances, based on inverse feedback and feedforward mechanisms (Ghez & Krakauer, 2000), which were presented in Section 0.1.2, the CNS might plan actively where and when the next heel strike should occur by employing supra-spinal signals from multiple sensory systems. Note that alternatively, another possible synchronization strategy of the CNS could be to adjust some parameters of a widely automatic perception-action system (e.g., the CPG activity) so that movement synchronization emerges (Latash, 2012). This hypothesis is discussed later, in Section 0.6.3. Most importantly, the mechanisms underpinning walking require that multiple sensory cues (i.e., visual, cutaneous, proprioceptive, & vestibular) are integrated. In the next section, we specify these sensory systems that seem most relevant to control one's own body motion for the synchronization during walking.

Synchronization-relevant cues

In order to intentionally achieve synchronization during side-by-side walking, the individual must match the movements with the movements of the walking partner. For this reason, her or his current gait cycle position must be estimated. There are multiple cues that could be used for this estimation. Those cues stimulate receptors of different sensory systems. There are exteroceptors, which are receptors that sense external stimuli like vision or sound and b) interoceptors, which are receptors for somatosensory and kinesthetic perception. Moreover, information may be retrieved by predictions based on feedforward models (Grush, 2004; Steinicke, Visell, Campos, & Lécuyer, 2013).

As mentioned in Section 0.1.1, in order to identify the position of her/his gait cycle, it may be sufficient when the individual watches the displacement of the feet. Yet, during walking, the gaze is usually directed to future foot contact locations (Lappe et al., 1999). Therefore, the feet are not observed during normal walking. Nonetheless, the visual system provides cues for perceiving self-motion. Self-motion is, besides others, essential for estimating one's heading direction, traveled distance, speed, and time to contact (Campos & Bülthoff, 2012). It may also be used to retrieve the current gait cycle position.

The most relevant visual cue for the perception of self-motion is optic flow, what consists in radial pattern with a focus of expansion (FOE) in the heading direction. According to Gibson (Gibson, 1958), the FOE serves as single control signal for navigation (Warren et al., 1996). However, because the movements of different body parts (eye movements, head movement) are superimposed on global body displacement, self-motion perception is ambiguous when relying only on optic flow cues. In order to robustly disambiguate retinal flow produced by the globally moving body from the eye and head movements, optic flow cues have to be combined with other cues (Lappe et al., 1999), such as vestibular cues and neck proprioception (Crowell, Banks, Shenoy, & Andersen, 1998).

The vestibular system detects the rotational head velocity and linear head acceleration (Cullen, 2012; Fitzpatrick et al., 1999). Based on the activity of the vestibular system, the gait cycle may be estimated by the sinusoidal acceleration pattern, since it is maximal at heel strike. The vestibular system is likely to detect this maximal acceleration allowing for a precise and accurate estimation of heel strike onsets (Kaya, 2014). For more information about the vestibular system, see Appendix A.3.

Yet, although it was assumed that the activity of the vestibular system is sufficient for a representation of self-motion, vestibular input alone is ambiguous due to superimposed body segments (head and body). It was suggested that neurons within the vestibular nucleus code self-motion by integrating the input from the multiple afferent information (vestibular, visual, proprioceptive, & somatosensory) and efference copies of the motor commands. Their integration disambiguates the cue information provided by single modalities (Cullen, 2012; de Winkel et al., 2010). Moreover, self-motion is affected by activity in the auditory system (Larsson, Västfjäll, & Kleiner, 2004; Suzuki, Sakamoto, & Gyoba, 2004) and the somatosensory system (Lécuyer, Vidal, Joly, Mégard, & Berthoz, 2004).

Signals from the proprioceptive system are integrated with signals from the vestibular system constituting the kinesthetic sense (Gelfan & Carter, 1967). It provides a sense of the dynamics and kinematics of the body segments. Proprioceptive and cutaneous signals are integrated for the perception of the position of segments relative to each other and relative to the surface (see Kaya, 2014). The proprioceptive system informs about continuous and discrete events. For instance, muscle spindles signal the contraction and muscle length changes before and when the heel contacts the ground and when muscles in the leg are compressed and stretched. In contrast, the cutaneous system informs mostly about the discrete, temporal pattern of the heel strikes. This is so because the responsible cutaneous receptors, Merkel and Meissner cells, are located on the plantar foot. At heel strike, weight is imposed on the leg followed by weight transfer when the foot contacts the ground. This temporal pattern of pressure signalizes heel strike onsets (Kaya, 2014; M. Patel, Magnusson, Kristinsdottir, & Fransson, 2009).

The auditory system is a structure of hair cells that respond to the frequency and amplitude of a sound wave. At later processing stage, the highest peak and location is processed/identified. The heel strike leads to a sudden and large upward force on the foot. Although much energy is absorbed by the elastic tissue of the heel, it produces acoustic peaks of a few milliseconds on the ground. Its spectrum and amplitude vary owing to footwear, speed, and ground properties. In normal walking, it usually has the highest amplitude of all sounds (Pastore et al., 2008). It is discrete and relatively regular depending on the cadence.

Efference copies of the motor command are involved when receptor stimulation has to be canceled or compensated. This mechanism enables the perception of non-moving aspects. For instance, a stable environment can be perceived although the retina is stimulated at different locations during eye movements; the body is perceived as rigid although the vestibular receptors are stimulated during head rotation. Compensation may be achieved by a mechanism comparing the magnitude of the stimulation with the predictions based on a cognitive representation of the action effects. These representations might also be used to predicting upcoming events based on feedforward mechanism (Steinicke et al., 2013).

When repetitive auditory sequences are presented like, for example, the stepping sequence of a walker, the temporal components, which are the pattern and tempo, might be coded in several areas across the cortex (Thaut, Trimarchi, & Parsons, 2014). Thus areas might be re-activated (simulated) for anticipating upcoming steps (A. D. Patel & Iversen, 2014; Repp, 2005). The perceived onset could be a combination of the sensory signals and the predictions based on signals like the efference copies of the motor commands (Grush, 2004).

In short, there are multiple signals from one's own movements that inform about the current phase of the gait cycle via self-motion. Self-motion is provided unambiguously by integrating visual, auditory, vestibular, proprioceptive, haptic, and efference signals. Moreover, there are continuous vestibular, proprioceptive, and visual cues that are not directly related to self-motion. Cutaneous and auditory cues are mainly provided discretely but they are regularly available when the pace is regular. Visual cues are provided irregularly because the individual seems to focus also heading direction and future foot contact locations. Temporal information about heel strike onsets is delivered by all involved sensory systems. It may be the event of the gait cycle that is represented most reliably. On the other hand, signals based on the continuously available self-motion, visual optic flow cues, and kinesthetic cues are used to control navigation and stability. These integrated signals may also be used to control synchronization.

As mentioned in Section 0.1.3, the gait cycle phase of a walking partner can be estimated by using auditory and visual cues. For visual cues to be provided, the individual should focus on the movements of the partner. It is assumed that this does not happen constantly due to variable head movements and gaze behavior (see e.g., Grasso, Prévost, Ivanenko, & Berthoz, 1998; Hollands, Patla, & Vickers, 2002). In contrast, auditory cues, mostly provided by heel strikes, are more discrete but its access is regular. While auditory cues provide mainly temporal information about cadence, continuous visual cues provide spatiotemporal information about cadence and displacement.

It might also be that surface vibrations produced by the walking partner could be perceived through haptic sensors on the foot sole and proprioceptive sensors on the joints and the muscles. It was shown that individuals can identify surface structures during walking based on haptic and proprioceptive information (Giordano et al., 2012) and individuals can synchronize walking tempo with rhythmically displayed haptic cues (Maculewicz, Erkut, & Serafin, 2016). Moreover, surface vibrations should also stimulate the receptors of the vestibular and the auditory system. Such findings point towards that the perceptual system is highly sensitive to surface vibrations and that an individual may be able to integrate such information for perceiving the heel strike onsets of the partner.

If an individual intentionally attempts to synchronize, the current phase of the gait cycle of the individual and the walking partner must be estimated. Based on these estimations, temporal asynchronies can be computed. These can then be used to compute appropriate correction commands. If synchronization emerges spontaneously, these mechanisms might be slightly different and might operate on lower processing levels. Lower-level processes are discussed in Section 0.6.3. However, in general, for both performances, gait synchronization could be understood as the cross-modal matching of audiovisual signals from the partner with kinesthetic, somatosensory, visual, and auditory signals from one's own movements.

On the one hand, there is multiple redundant information about the same spatial and temporal properties. For instances, both visual and auditory cues provide information about the heel strike onsets. On the other hand, there is non-redundant information delivered by single modalities that disambiguate perception. For example, the velocity of the moving body could be estimated based on auditory cues signalizing cadence combined with continuous visual motion and distance cues. Cadence alone does not provide sufficient information about velocity. All these cues might contribute to the temporal matching of one's own movements and the movements of the partner, possibly through their integration according to an MLE model. A caveat is that cues of different sensory signals have different travel times leading to asynchronies of different magnitudes at several stages of the information processing pathways. The origin and extent of these asynchronies and their consequences for the synchronization task are discussed in the next Section 0.1.6.

0.1.6 Caveats and Goals

In this section, we describe the difficulties of integrating multiple sensory cues due to naturally occurring signal delays and variability within each sensory system. Subsequently, we suggest a solution to this problem and formulate our experimental hypothesis.

In previous experiments, it was shown that when participants have to judge the simultaneity of a visual and an auditory event, auditory cues were judged as occurring slightly earlier (a few milliseconds) than the visual event, when both are elicited simultaneously from by the same source near the observer. The exact size depends on the distance to the cues (e.g., Vroomen & Keetels, 2010). When participants attempt to synchronize finger taps with metronome events, finger taps usually precede stimulus onsets by a few tens of seconds (i.e., 20-80ms) in order to be perceived as synchronous (see e.g., Mates, Müller, Radil, & Pöppel, 1994; Repp, 2005).

Although these modality-leading biases could be attributed to artifacts induced by the experimental methods, since the authors did not control for the co-location of the audiovisual stimuli (Silva et al., 2013), such findings highlight that the temporal matching in perceptual and SMS tasks takes place on a central processing level and that different sensory cues could require different times to be processed.

All signals pass through complex pathways before being integrated at higher levels. First, information (for exteroceptors) is transported within a particular physical medium (Propagation). Then, it stimulates the receptor of the corresponding sensory system (for extero- and interoceptors). Subsequently, the receptor potential is converted into Action Potentials (AP) (Transduction). Next, the APs travel along the sensory nerves (Transmission). Finally, the multiple APs are integrated and interpreted over space and time, mostly within the cortex (Processing). The propagation and processing time of the different signals differ across sensory pathways. Consequently, signals from different systems must occur at different temporal onsets in the environment to be perceived as synchronous; or are adapted in such way that requires a small lag or delay between both to be perceived as synchronous by the CNS. This explains why a finger tap must precede auditory metronome events (in particular conditions) so that both are perceived as synchronous. In this case, the haptic signals might need more time than the acoustic signals to be processed. Next, are provided some examples of the temporal differences for each of the processing steps.

Propagation: While propagation time has to be taken into account for exteroceptors (visual or auditory), it is absent for interoceptors (somatosensory perception, kinesthetic perception). External sensory signals travel with different velocities. While light travels at almost 300.000.000m/s, sound travels at approximately 343m/s. Thus, when the stimulus source is far away, sound arrives late compared to light. Because the CNS adapts to these naturally occurring asynchronies (Keetels & Vroomen, 2012), light reflected by events at large distances usually has to arrive earlier than sound to be perceived as synchronous with the sound.

Transduction: Audiovisual stimuli that are elicited from objects at a distance of 10 to 15m are perceived as synchronous (or simultaneous). This distance is therefore called the "horizon of simultaneity". When the stimulus source is closer to the individual than the horizon of simultaneity, sound has to precede vision (Keetels & Vroomen, 2012). This may be due to a faster transduction time of the auditory system. The visual system's phototransduction is the transduction from electromagnetic waves into action potentials at the photoreceptors through an "indirect" modulation of ion channels by a second messenger involving G protein activation. It takes tens of seconds ($\sim$50ms) in cones (see Poepple 1990 in Atmanspacher & Ruhnau, 2012; Arshavsky, Lamb, & Pugh Jr, 2002).

In contrast, most other sensory systems consist of mechanoreceptors. In the auditory system, hair cells are displaced by an acoustic pressure wave entering the ear. In the vestibular system, ciliary cells displacement is caused by the deformation of the otolith membrane (for linear acceleration). The displacement of the ciliary cells directly changes ion channels without the activation of a second messenger. Also, the somatosensory system consists of mechanoreceptors. In the cutaneous system, the receptors respond to pressure. In the proprioceptive system, the receptors respond to distortion (stretch, compression). This mechano-transduction is faster (1-10ms) (V. Torre, Ashmore, Lamb, & Menini, 1995; Atmanspacher & Ruhnau, 2012) than phototransduction.

Transmission: After transduction, the AP travels along afferent neurons to the peripheral nerve where it enters in the posterior root of the spinal cord (in somatosensory information). Because receptors are located at different places along the body, the signals have to travel different distances. For the visual, auditory, and vestibular system, APs travel through differently complex paths within subcortical structures. The time it takes until the signal reaches the primary sensory cortices depends on many factors, like the length and thickness of the nerve fibers, its myelination, and the number of synaptic junctions.

Signals for proprioception travel through type 1a, type 1b, and type 2 nerves. The neural transmission velocity of type 1 nerves is 80-120m/s. The neural transmission velocity of type 2 nerves is 33-75m/s. Signals for haptic perception also travel through type 2 nerves and through type 3 nerves. The neural transmission velocity of type 3 nerves is 3-30m/s (Siegel & Sapru, 2006). This implies that the transmission times of somatosensory signals from foot sole to central processing requires about 10ms to 20ms (type 1), 21ms to 48ms (type 2), and 53ms to 533ms (type 3).

According to Keetels and Vroomen, the neural transmission times of visual signals are about 50ms and for auditory signals about 10ms (Keetels & Vroomen, 2012). The length of the pathways of the visual, auditory, and vestibular system should be widely similar for different individuals. However, the pathways of the somatosensory system vary with the stature of the individual. A pressure signal at the receptors of the foot sole reaches the primary somatosensory cortex within 53.3ms for an individual with a stature of 1.60m and within 60ms for an individual with a stature of 1.80m. This is a difference of nearly 7ms, being almost as large as the transmissions times of signals of the auditory system (Keetels & Vroomen, 2012) and on its own might be significant in interpersonal gait synchronization.

The velocity of signals traveling from receptor to central processing level is also called nerve conduction velocity. This includes transduction and transmission times. It is determined by measuring the time difference between stimulus presentations and evoked potential in the respective cortical area (Neetens, 1984). For example, a signal that is generated by stimulation of receptors in the retina requires approximately 100ms to reach the occipital visual area, determined by visual evoked potential (see also Thorpe, Fize, & Marlot, 1996). Brainstem auditory evoked potentials are registered in the cochlea, superior olivary complex, and superior colliculus in the midbrain within 10ms after receptor stimulation. These delays match the sum of the transduction and transmission times provided by the previously presented studies (Keetels & Vroomen, 2012).

The fastest signal transmission is probably in the vestibular system. Rise to this assumption provides the vestibular-ocular reflex (VOR). The VOR is a reflex that serves the stabilization of the retinal images during head movements by generating compensatory eye movements. In order to do this effectively, eye movements have to be generated within 5-6ms (Cullen, 2012).

Concluding, there are different sensory systems providing information about the same event. Depending on the particular medium and sensory system, such information travels via different information processing pathways. At each level, this may require more or less time. Most of such asynchronies are externally compensated, through different distances and propagation times, and internally compensated, through different transduction times, transmission times, processing times, and adaption processes. This compensation leads to the processing of synchronous signals (see e.g., Silva et al., 2013). However, information might still be temporally incongruent to some extent so that the CNS requires some degree of tolerance to these incongruences (Vroomen & Keetels, 2010).

Thus, although it is here only provided an incomplete picture of all pathways, delays, and processes, it demonstrates that signals from different sensory systems may be incongruent at different levels of these above-mentioned information-processing pathways. For most of the perceptual experiences that are important for gait control (self-motion, optic flow, localization of objects), different uni- and multimodal cues have to be combined based on these incongruent signals. As for synchronization, an exact timing is required, and the sensory signals from the movements of the walking partner must be matched with one's own movements. It is curious how precise and accurate matching can be achieved based on these multiple signals.

When two signals represent independent events (e.g., arising from different origins), signals are not integrated and the perceptual experience may be based on the information from single cues. The evidence may be obtained from relatively simple heuristic rules implemented in the CNS: Signals are not integrated, a) when they most probably come from different locations, b) when the temporal asynchronies are large, and c) when single cues alone elicit strong stimulus effects (see Stein & Meredith, 1993). The maximal temporal separation at which signals are integrated is called Window of Temporal Integration (WTI) (Vroomen & Keetels, 2010). Its size varies largely between the type of task and stimulus. For the perception of complex biological stimuli, auditory and visual cues seem to be integrated up to large asynchronies (∼250ms, see Vatakis & Spence, 2006). Moreover, when an unambiguous perception is established based on information from a single cue, there is no need to integrate further ambiguous cues (see the Principle of Inverse Effectiveness Holmes, 2009).

However, in healthy participants walking side-by-side (interpersonal distance of 30cm to 100cm), the cues from one's movements and the cues from the walking partner should meet the above mentioned “heuristic requirements” and therefore should be integrated. Then, the ambiguous and slightly asynchronous cues might be integrated optimally, according to the previously proposed MLE model. In this way, the perceived features would be a result of the combination of multiple cues, whereby their weights of contribution are specified by the reliability with which each cue represents, on its own, the particular feature in question. For heel strike synchronization, this would mean that the perceived heel strike onsets of the individual are perceived as lying between the onsets signalized by vestibular cues, visual cues, auditory cues, and somatosensory cues (and more complex, already combined cues from kinesthetic and self-motion signals). The perceived heel strike onset of the walking partner would then lie between the onset signalized by auditory cues and visual cues. Based on these estimations, asynchronies could subsequently be computed and reduced.

Testing such assumptions is not a straightforward procedure. The control of all these relevant variables is quite difficult when merging multimodal and biological stimuli in a complex perceptual-motor task. In addition, the manipulation of sensory channels may cause unwanted side effects on relevant processes during walking. For example, manipulating the vestibular system, the proprioceptive system, or the somatosensory system affects functions involved in maintaining stability and inter-limb coordination (Cullen, 2012, 2016; Prochazka & Ellaway, 2012; Prochazka, Gritsenko, & Yakovenko, 2002; Ghez & Krakauer, 2000; Kaya, 2014). Also, the use of side-blinders constraints optic flow and therefore the perception of self-motion (Campos & Bülthoff, 2012; Lappe et al., 1999). For this reason, this is a first attempt to test for cue integration in movement synchronization during side-by-side walking by focusing on auditory and visual cues from the walking partner.

In the first three experiments of the present thesis, the perceptual cues (visual and auditory) from a walking partner were virtually simulated. Given that auditory and visual cues inform about the time of the upcoming heel strike of a walking partner, it seems likely that the CNS integrates signals of both modalities to obtain the best estimate of the temporal onsets. We hypothesized that this should then improve synchronization according to the MLE model. In Experiment 1, human real-sized Point Light Walkers (PLWs) obtained from the motion capture of walking individuals were used. Standard

PLWs contain the spatiotemporal components of human motion and their implementation allows a controlled manipulation of these components. Previously, it was shown that PLWs are an adequate mean for the study of intermodal perceptual processes as for instances recognition (see e.g., Thomas & Shiffrar, 2013), velocity (Mendonça et al., 2011) or simultaneity judgments (Silva et al., 2013). In this experiment, participants walked next to PLWs and were instructed to synchronize. Auditory, visual, or audiovisual cues about the PLW were provided. In Experiment 2, the virtual environment was improved by using a human-sized virtual mannequin as visual stimulus and stimuli positions were rendered in real-time. Thus, the spatiotemporal congruency of audiovisual cues was increased. In Experiment 3, the temporal onsets of the auditory and the visual cues were manipulated and its effect on motor synchronization performance was analyzed. Due to convenience, in the first three experiments, synchronization performance was quantified by considering the discrete heel strike onsets, only.

Subsequently, in a fourth experiment, possible shortcomings of the preceding three experiments were scrutinized by assessing synchronization performance of several body segments of real human pairs walking side-by-side. We further experimentally manipulated the direction of information exchange and the available visual and auditory cues in order to generalize our findings. We further speculated whether synchronization during side-by-side walking might be achieved by matching global motion signals rather than single segments.

0.2 Experiment 1

0.2.1 Methods

Participants

Eight participants (7 naïve, 2 female, 6 male, age: $M = 28$, $SD = 3$, all right dominant hand) without gait disabilities took part in the experiment. All individuals gave informed consent for their participation.

Stimuli & Experimental Design

The experiment was programmed in Python (*Python 2.0*, 2016), using OpenGL for graphics presentation and OpenAL for audio playback.

Visual Stimulus. Body kinematics during walking with different velocities (0.7-1.5m/s) of six males and eight females were previously captured with a Vicon motion capture system at 240Hz. From eight gait cycles of these models, PLWs, with 13 dots generated in 2D coordinates and rendered as black dots, served as the visual stimulus. The 13 dots signalized the spatiotemporal positions during walking of the head, shoulders, elbows, hands, hips, knees, and ankles. The PLWs were presented by 3 DLP projectors Christie Mirage with a resolution per channel of 1400x1050 at 60Hz. The three images were retro-projected side-by-side with blending areas between images resulting in a 7.20m(H) x 2.10m(V) flat screen in a dark room. The PLW dots were black (4 cd/m2) and the background was light gray (70 cd/m2). The PLW was projected in real size in the sagittal plane walking from one side of the screen to the other.

Auditory Stimulus. The auditory stimulus was a footstep recorded from an individual with average stature in Portugal—1.62m, including male and female population (Arezes, Barroso, Cordeiro, Costa, & Miguel, 2006)—walking on a wooden floor barefoot at a velocity of 1.3m/s. From these records, two footsteps were auralized by a MATLAB routine with head-related transfer functions (HRTF) with similar acoustic properties of the sound that reaches the ear produced by an individual walking next to the participant at 0.5m. (Left foot: azimuth 90 degrees, elevation -72 degrees, relative to the right ear; right foot: azimuth 90 degrees, elevation -62 degrees). The intensity was matched to the recorded sound intensity at an average ear height (1.53m), being for the closer left foot 63 (dB LAeq) and for the right foot slightly lower. The footsteps were presented through wireless headphones (Sennheiser RS 120 II).

Audiovisual Stimulus. The audiovisual stimulus was the PLW presented with the sound produced by the heel strikes of the PLW. In order to assure synchrony between visual and auditory stimulus, the time delay between a sound stimulus and a visual flash was measured (Lamas et al., 2015). A delay of 15ms, $\pm 3ms$, was applied to the auditory signal for achieving the correct temporal alignment taking into account audiovisual signal propagation times.

Procedure

The experiment was conducted at the Laboratory of Visualization and Perception of the University of Minho and Centre for Computer Graphics. Prior to the experiment, the participants walked on a short walkway—a 13.50m x 0.92m wooden floor—in order to determine the comfortable walking velocity. Instructions were given: "walk comfortably but not too slow; walk as if you were walking with a friend". The participants walked as many times as needed until the velocity of three subsequent walks did not deviate more than 5% from the mean velocity of the three walks.

Subsequently, one PLW was chosen, according to gender, hip height (max. difference = 1.9cm) and the comfortable velocity (max. difference = 0.1m/s), to match relevant gait characteristics. Participants wore shorts and walked barefoot on the walkway, which was located next to the screen on which the PLW was displayed. See Figure 2 for a cartoon illustration of the setup. Participants started walking 2.4m before the screen began and stopped 3.2m after the screen end. Two reflective markers were attached to the malleolus of the ankle of the participant and four markers were attached to the head. Marker positions were captured at 240Hz by a Vicon motion capture system with 6 near-infrared cameras (MX F20 of 2 megapixels) and defined in a XYZ-Euclidian frame. The participant walked alongside the projected stimuli for 7.20m. Participants were instructed to: "Walk without interruption and do not reduce velocity until the walkway end. When the PLW is displayed, synchronize steps and maintain position at the side of the PLW. When not displayed, synchronize with the auditory footsteps. When PLW and footsteps are presented, synchronize with both and maintain the smallest distance as possible to the PLW". No instructions were given regarding gaze direction.

The availability of sensory information and the start phase of the stimulus were manipulated to create the following conditions: for the sensory information, 1) the PLW was displayed temporally aligned with the presentation of footstep sounds (audiovisual condition - AV), 2) only the PLW was displayed (visual condition - V), or 3) only the footstep sounds were presented (auditory condition - A); for the start phase, the PLW/footsteps started 1) in midstance, 2) in midswing, or 3) with a heel strike of the left foot. A within-subject design (Sensory condition [3] x Start phase [3] x Rep [3] x

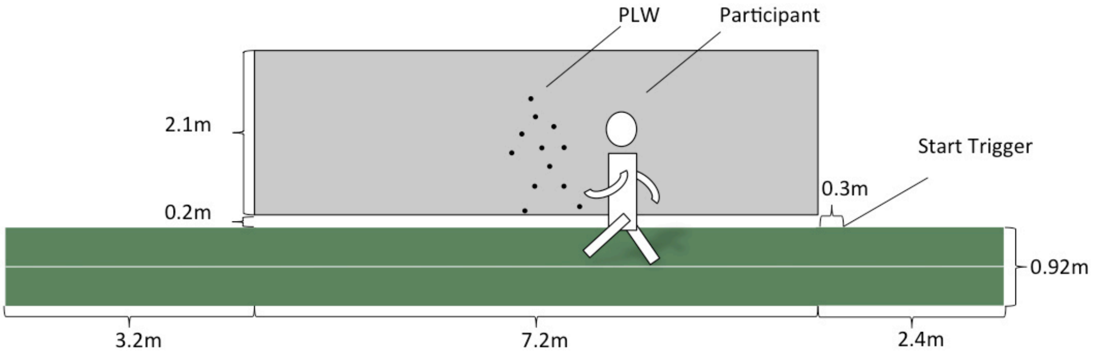


Figure 2: Setup of the Experiment 1. It is displayed the walkway (green) and the projection screen (gray).

Block [3]) was used.

0.2.2 Analysis

All analyses and statistical inference were conducted with R Studio version 0.98. In this work, we defined synchrony as a bounded temporal relationship. This can be captured by the variability of temporal differences of certain key events within the gait cycles of each pairing. Thus, the focus was the temporal aspect. However, it must be highlighted that spatiotemporal dimensions are intertwined. As illustrated in Section 0.1.3, time perception is affected by distance cues and spatial perception is affected by temporal cues. Consequently, for analyzing (temporal) synchronization, spatial factors have to be taken into account. The scope of the subsequent spatial analysis was mainly to serve as a control for the following temporal analysis. This is why its results and implications are discussed within the analysis section. We analyzed the distance between the walkers and their stride length differences.

Spatial analysis

Distance between head positions. Considering different audiovisual propagation delays, synchrony perception should be affected by the distance of the participant relative to the stimulus, which in turn should affect synchronization performance. Due to the particular setup (see Section 0.2.1), the distance was the same at the start but then could increase, should the walkers walk with different velocities. For this reason, we analyzed the distances between the head positions. The distances were analyzed at each step providing (Step[10] x Sensory condition [3] x Start phase [3] x Rep [3] x Block [3]) 810 distance records for each participant.

Due to missing information of the head position of 3 participants in Experiment 1 (in these cases we only reliably captured the foot positions), the head positions of 5 participants were analyzed. The distances between the head positions gradually increased as a function of step number (Figure 3). This increase differed across the participants. The mean absolute difference at step 10 was between 0.05m and 0.62m (Figure 4). For four of the five participants, A led to larger increases than AV, V, or both.

Difference between stride lengths. As a second spatial measure, we considered stride length differences. If the synchronization was achieved by adjusting spatial parameters, this could be reflected by an increasing similarity of the stride lengths as a function of the step/stride number. Eight step cycles were identified by the horizontal displacement of the ankle of the participant (see Figure 5). The distance between two consecutive heel strikes determined a step length. The distance between four consecutive heel strikes determined a stride length. We analyzed the differences between the stride lengths of the participant and the stimulus (see Figure 6). For each participant, there were a maximum of 324 stride length differences (Stride[4] x Sensory condition [3] x Start phase [3] x Rep [3] x Block [3]). Note that we considered stride and not step length for computational reasons but that for the scopes of this analysis both measures would lead to similar results (see Levine et al., 2012).

The strides did not change as a function of stride number (Figure 6). The last stride difference (stride 4) was different across the sensory conditions for four of seven participants (Figure 7). For these participants, A led to larger differences than AV, V, or both. The differences were between 0.04m and 0.23m.

We analyzed the differences between head positions and stride lengths as a function of step (stride) number. Comparing these differences across the sensory conditions revealed that the patterns of results were similar. A led to larger differences than AV or/and V. This is plausible, considering that in Experiment 1 the auditory cues did not provide relative distance cues of the stimulus. Overall, while the distances between the walkers seemed to increase slightly, we could not find such increases for the stride

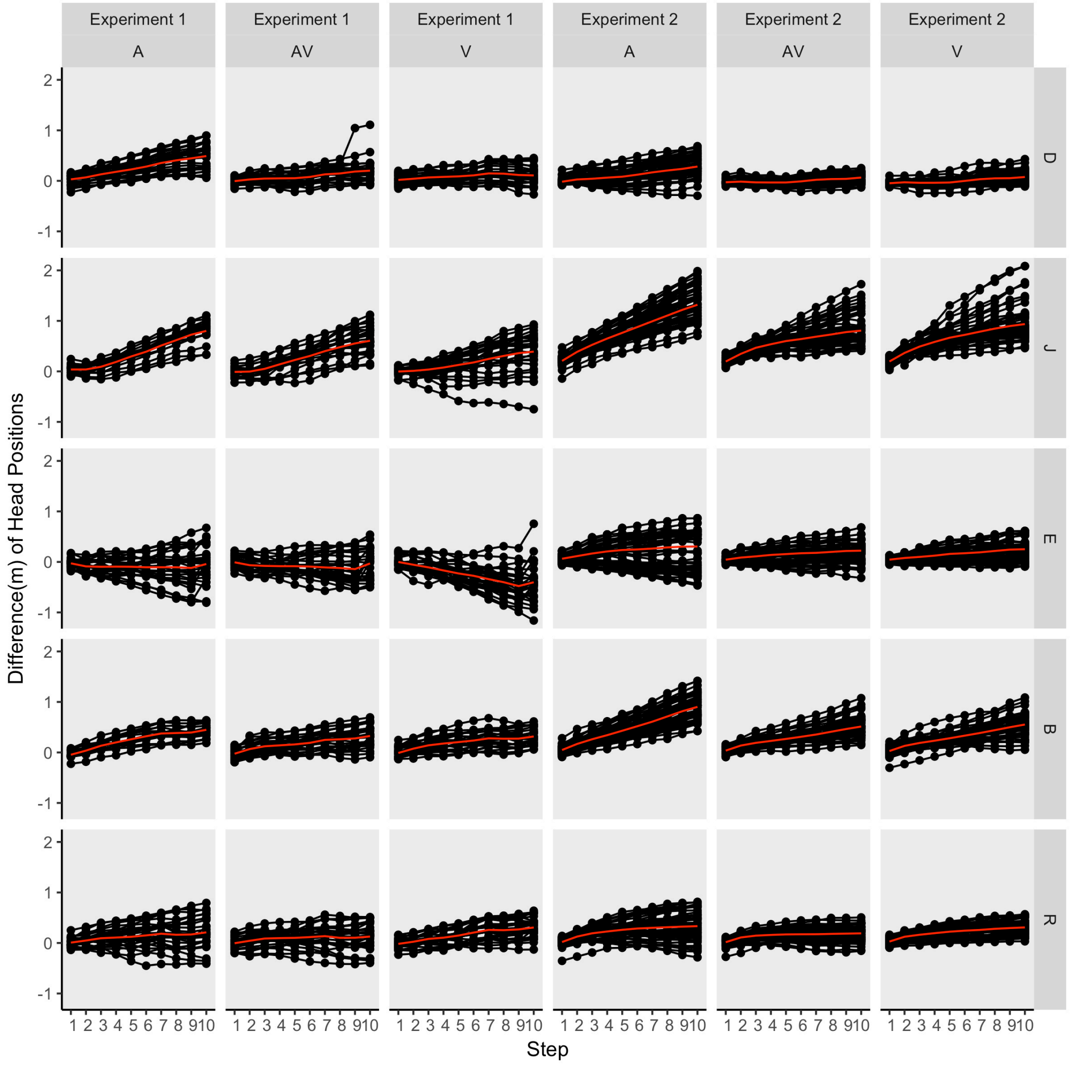


Figure 3: Scatterplots of the distances(m) between head positions as a function of step number of the 5 participants and the stimuli of Experiment 1 and Experiment 2. The black line-segments are the individual trials. The red line is the mean.

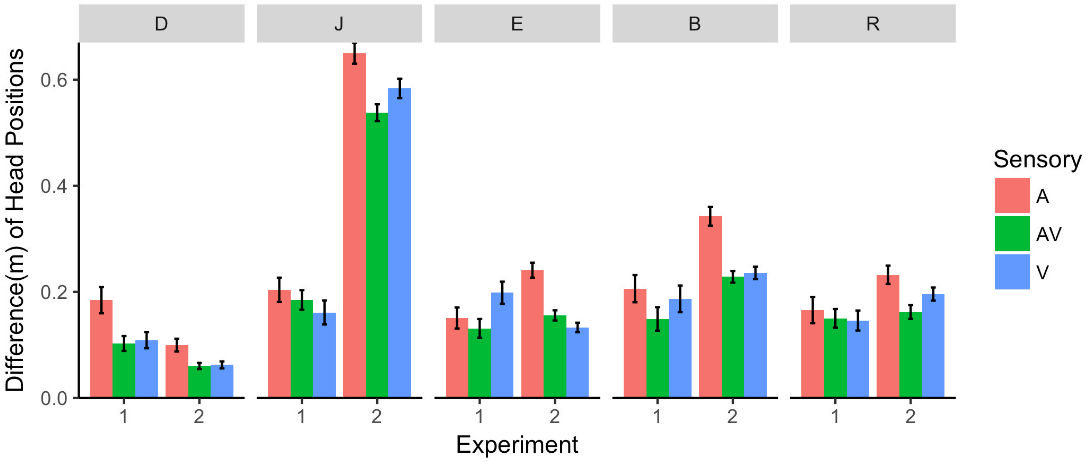


Figure 4: Barplots of the mean absolute distances between the head positions at step 10 of the 5 participants and the stimuli of Experiment 1 and Experiment 2. Error bars represent the 95% confidence intervals.

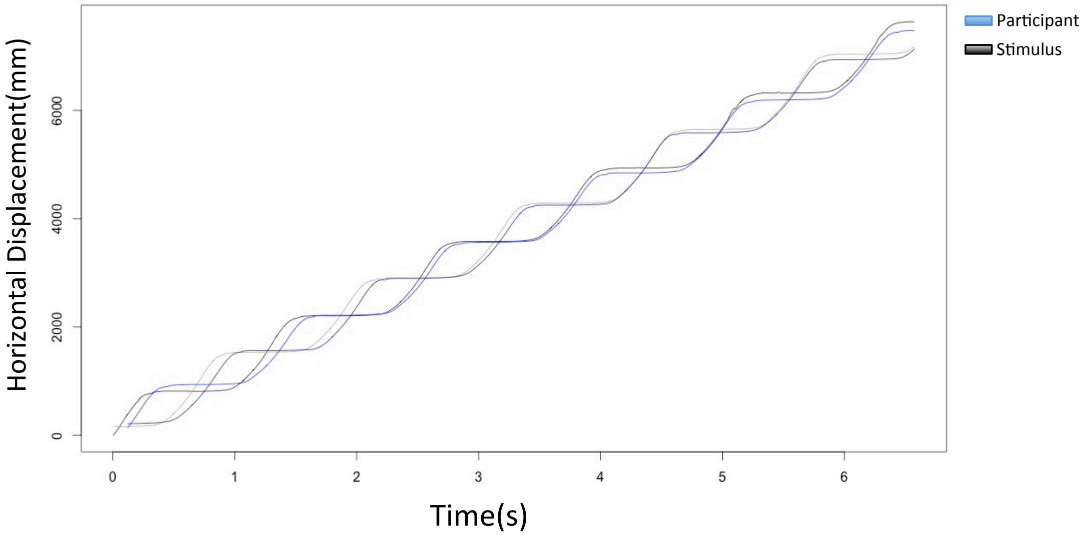


Figure 5: Horizontal displacement(mm) of the markers attached to the ankles as a function of time. Displayed are the position of left ankle (blue) and right ankle (dark blue) of the participant and the position of the left ankle (gray) and right ankle (black) of the PLW. Step cycles were identified at the inflection points—i.e., the first measurement after a movement; i.e., when the velocity reaches the first time zero.

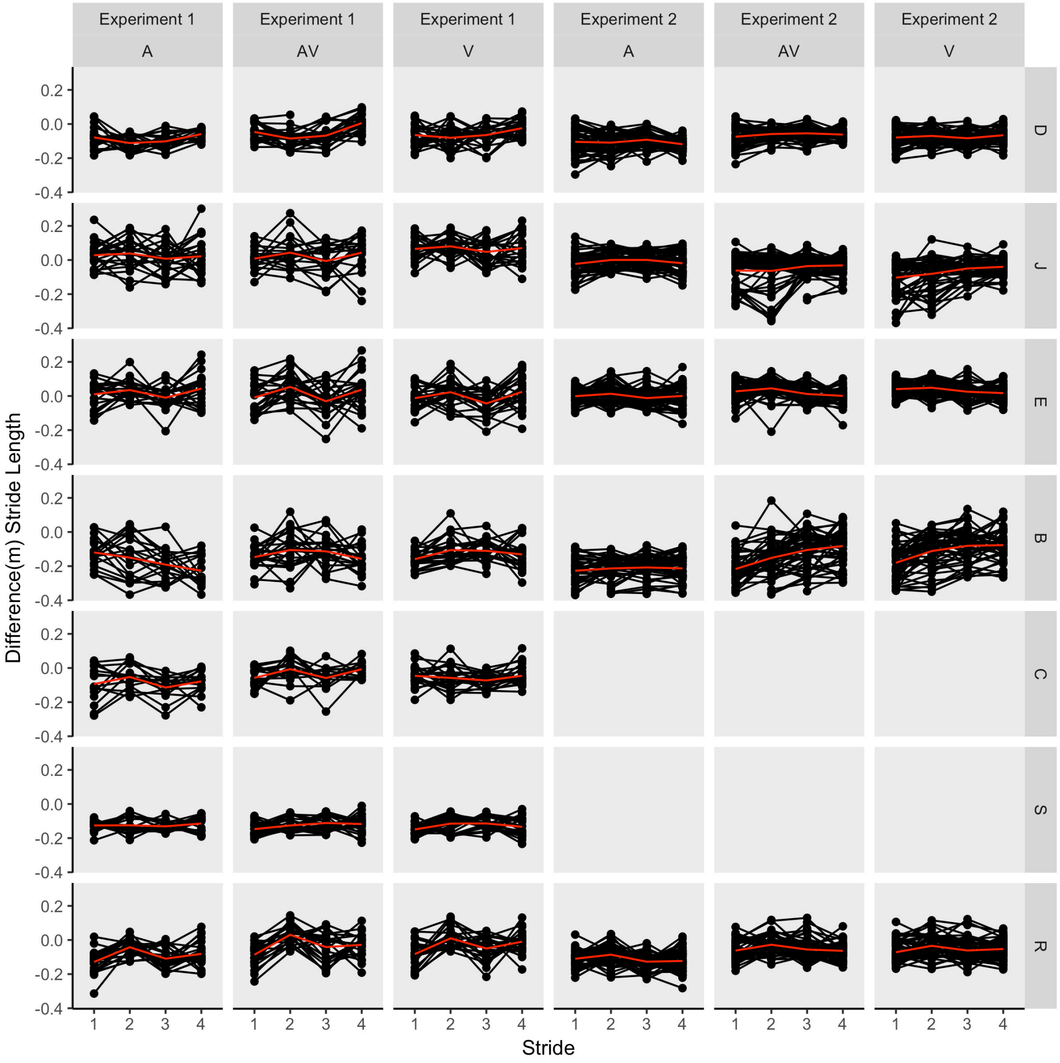


Figure 6: Scatterplots of the difference(m) between the stride lengths as a function of stride of the 7 participants of Experiment 1 and 5 participants of Experiment 2. The black line-segments are the individual trials. The red line is the mean.

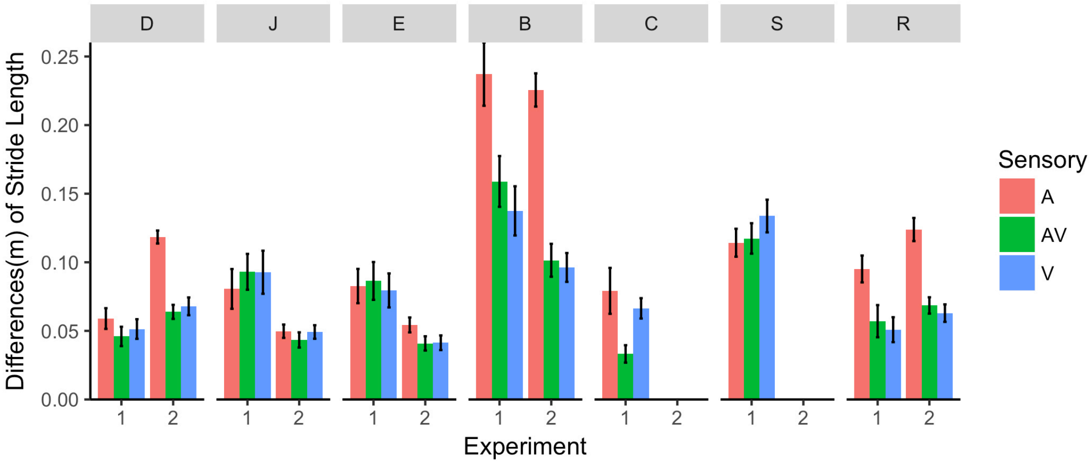


Figure 7: Bar plots of the absolute differences of the 4th stride lengths of 7 participants of Experiment 1 and 5 participants of Experiment 2 and the stimuli. Error bars represent the 95% confidence intervals.

lengths. However, distance differences of both measures were very small and most were still lying within the peri-personal space. We, therefore, suggest that spatial adjustments were marginal for the scopes of this study and that the subsequent temporal (synchrony) analysis should not be affected that much. (In Appendix A.4 we demonstrate the absence of any relation between both measures).

Temporal analysis

The temporal analysis was the main scope of this study. Ten heel strikes were identified by the vertical displacement of the ankle of the participant (see Figure 8) in each trial. The difference between the onsets of the heel strikes of the participant was subtracted from the onsets of the heel strikes of the PLW, to compute the temporal asynchrony. A within-subject design (Steps [10] x Sensory [3] x Start [3] x Rep [3] x Block [3]) was used. Therefore, maximal of 810 asynchrony records were obtained for each participant.

In order to capture synchrony between cyclic motions, two approaches were suggested: 1. Synchrony means walking with the same cadence (frequency locking). 2. Synchrony means walking with a stable phase difference (phase locking). Phase locking implies frequency locking but not vice versa. People can walk with the same frequency

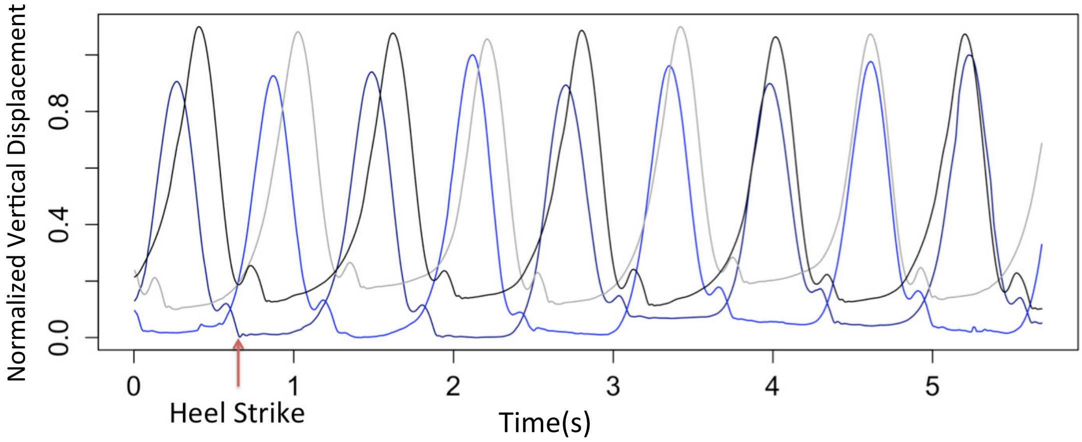


Figure 8: Normalized vertical displacement of the markers attached to the ankles as a function of time. Displayed are the position of left ankle (blue) and right ankle (dark blue) of the participant and the position of the left ankle (gray) and right ankle (black) of the PLW. For illustration, the time series of the PLW was vertically shifted by +0.1 on the y-axis. Heel strikes can be identified as the second local minima in the vertical displacement of the ankle within a stride cycle (red arrow). Due to signal noise and movement variability, this was sometimes not clearly identifiable and the position had then to be estimated manually.

but their phase difference can still vary and drift apart.

Frequency locking. Synchrony may be considered as frequency locking (see e.g., Pikovsky et al., 2001; Richardson et al., 2007). Frequency locking means that the step frequencies become matched. Considering the very short walking samples (10 steps), a trial was considered as frequency locked when the phase difference from start to end did not exceed 180 degrees, which is half a step cycle. This happened in 91% (D), 88% (E), 71% (B), 83% (C), 91% (J), 89% (S), 47% (R) of all trials. This suggests that participants were frequency locked since the start for a high proportion of trials. We attributed this to the fact that it was always presented the same stimulus. This manner, participants could easily anticipate the correct frequency and were making cadence adjustments before the stimulus appeared and the individuals' movements were captured. Therefore, we considered frequency locking as marginally relevant for the analysis of synchrony. Subsequently, phase locking was considered for these frequency locked episodes (e.g., Nessler et al., 2013; van Ulzen et al., 2008).

Phase locking. Phase locking is a more sensitive measure of synchrony. It means that phase differences obtained from two cyclic movements become stabilized. In general, the phase is the temporal position of a point within a cyclic movement. The phase difference is the difference between two such points from two different oscillating objects with cyclic motion (see e.g., Pikovsky et al., 2001). The phase difference can be represented as the absolute temporal difference (i.e., asynchronies in ms), but it is more conventional to represent it in degrees obtained from the angles of two movement cycles (we discuss this aspect later in this section).

The phase difference is also called relative phase. The concept relative phase comes from the Dynamical System approach. It was already introduced in Section 0.1.2 when describing the pendulum synchronization of grandfather clocks. The relative phase is the difference between the phases of two oscillators. When being simultaneously at the same position within the cycle, this is in-phase synchrony; relative phase = 0 degrees. When being 180 degrees apart, this is antiphase synchrony; relative phase = 180 degrees.

The relative phase can be represented continuously (Continuous Relative Phase = CRP) and discretely (Discrete Relative Phase = DRP). The CRP reflects the phase relation at each time instance throughout the cycle and can be computed by subtracting the instantaneous phase angles of participant and stimulus (for a detailed description and computation, see Appendix A.6). However, when a) the signal has a nonlinear phase increase, thus it is non-sinusoidal and b) particular events within the cycle are of interest, the DRP is more appropriate (Peters, Haddad, Heiderscheit, Van Emmerik, & Hamill, 2003). In our studies, sound information was mostly provided by the heel strikes. The continuous motion trajectory of the heel is non-sinusoidal. In addition, the heel strikes seem functionally very relevant during walking because they are short lasting events that signal a phase transition from single to double support (see Levine et al., 2012). For this reason, we decided to use the discrete heel strike moments as events of interest and considered the DRP rather than the CRP.

In general, the DRP is a point estimate that captures the timing of discrete events relative to the same (or comparable) events of another oscillating entity with cyclic motion. As pointed out above, these were in the present study the heel strikes of the participant and the stimulus. For computing the DRP, first, the temporal onsets of the steps of the participant had to be subtracted from the onsets of the steps of the PLW (PLW-participant) providing an asynchrony estimate (in ms). Then, the asynchrony was divided by the step cycle duration of the PLW (in ms) and multiplied by 360 degrees, revealing a DRP:

$$DRP_j = \frac{t_j^{PLW} - t_j^{participant}}{t_{j+1}^{PLW} - t_j^{PLW}} 360,$$

where t is the onset of the j heel strike. The DRPs could range from -360 degrees to 360 degrees. They were subsequently transformed to a different scale ranging from 0 degrees to 180 degrees, with 0 degrees being in-phase and 180 degrees being out-of-phase. Here, 180 degrees is not antiphase because we did not distinguish between left and right foot. The time between a heel strike of the right foot and the left foot was considered one cycle interval (see Figure 9). We did not make the conventional in-phase-antiphase distinction because it is actually not that clear what should be considered in-phase and antiphase during walking (van Ulzen et al., 2008) and because in pilot tests we did not find any differences between them.

Absolute asynchrony versus DRP. Depending on the context, asynchronies can be represented as a) the DRP expressed in degrees (see e.g., Schmidt & Richardson,

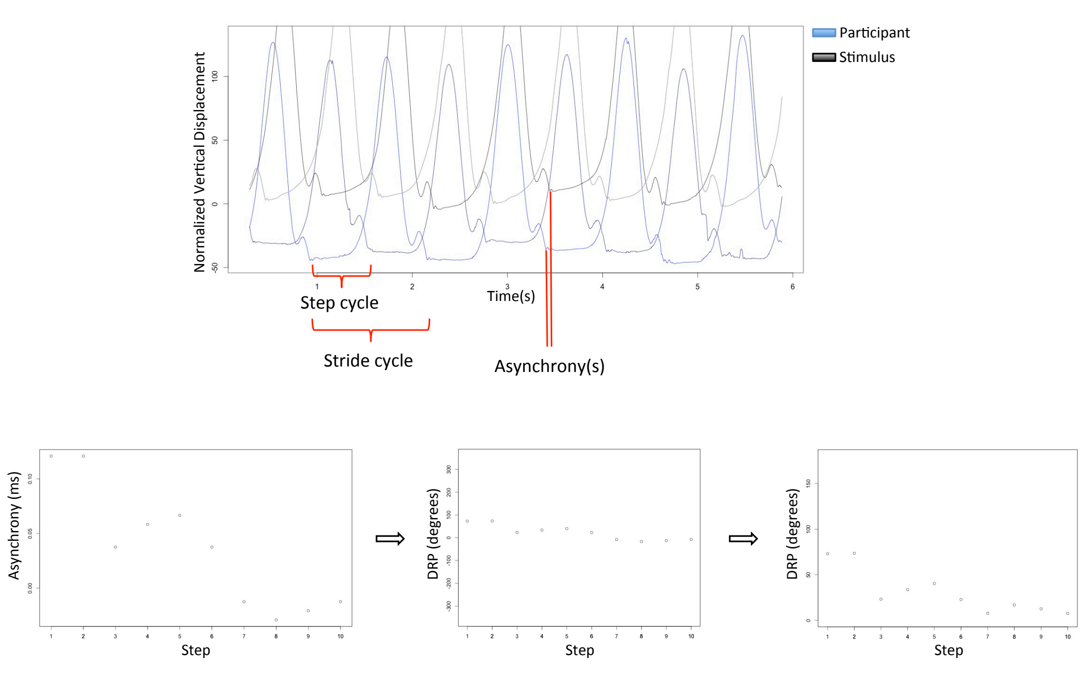


Figure 9: Transformation from original signal to final DRP.

2008) or b) its temporal separation expressed in milliseconds (see e.g., Repp, 2005). While the former is normalized by the cycle interval, the interpretation of the latter is more intuitive. Difficulties in using the absolute asynchrony measure arise when the walking frequencies of the walkers differ over time so that several synchronized and non-synchronized periods may be observed. Then, it is convenient to capture phase differences by the DRP because it is a circular measure of deviation. When DRP passes 180 degrees, it shrinks again while the asynchronies continue increasing/decreasing without upper/lower bound. However, in our study, we only considered frequency locked trials so that both reflect similar quantities. Nevertheless, due to a normalization by the cycle intervals, DRPs provide a relative measure of deviation. It, therefore, is a slightly different quantity than the asynchrony. One quantity may reveal patterns in the data that the other does not capture.

Moreover, because the DRP was often used in previous studies, we had a comparison standard (see e.g., Zivotofsky et al., 2012; Nessler & Gilliland, 2009; van Ulzen et al., 2008). For instances, slight variations of DRPs were usually tolerated. This means that DRPs that were falling within a particular interval were still considered as the same. DRP intervals from previous studies can serve as orientation when implementing such intervals in our study, as discussed later. Yet, in order to interpret the results within the framework of information processing theories, asynchronies (in ms) were still relevant. Consequently, both measures were used to analyze the results depending on the particular question on hand. Note that a similar argumentation justifies why we displayed alternately these measures as a function of step number or time. The step number naturally classifies the asynchronies into intervals while time is an absolute and more intuitive measure.

Quantification of synchrony. The extent of phase locking is usually quantified by the variability of the DRP (or asynchrony) over large periods of time (or steps) within a trial. Here, this is problematic because the implemented setup constraint the walking samples to very short periods ($\sim$5s).

Frequently, we observed that participants started “non-synchronized” and then attempted to reduce the asynchrony. Therefore, measuring asynchrony variability within trials is highly biased by the time that had passed since the start. Moreover, when starting occasionally more asynchronous, the variability would be elevated although the participant could be actually very good in synchronizing. A less biased measure would be to compute the asynchrony variability for the same and smaller temporal windows (or steps) across trials.

However, it is important to consider that synchronization could be a transient phenomenon (see van Ulzen et al., 2008). This means that dyads might have alternately entered into synchronized and non-synchronized states. Then, short walking periods might fail to capture reliably the synchronization process. Figure 10 shows the asynchronies as a function of step number for each individual and sensory condition. The plot illustrates that most asynchrony series converge logarithmically to a particular value. At steps ≥ 6, the slopes of these converging curves approach zero. Figure 11 shows the asynchrony variability. Similarly, the asynchrony variability seems to stabilize logarithmically at steps ≥ 6. The observations of Figure 10 and Figure 11 suggest that a) the participants followed the instructions and minimized the relative asynchronies and that b) once being minimized, there were no later transitions to less synchronized states. Note that the above-described pattern is far more consistent in Experiment 2 compared to Experiment 1. This probably owes to artifacts of the setup, as discussed in Section 0.2.4.

The stabilization of asynchrony magnitude and variability indicates that participants did not attempt to further minimize the asynchrony. Assuming that participants were motivated to do so, this pattern suggests that at later steps a) the asynchronies were perceived as synchronous, or b) that the perception-action system was unable to further reduce the asynchronies. Nevertheless, it is plausible to assume that the participants attempted to get the asynchronies close to the converging point. This point was therefore interpreted as minimal achievable asynchrony (see also Semjen, Vorberg, & Schulze, 1998). Asynchronies close to the converging point were framed synchronous and getting to this point was considered as an attempt to maximize the synchrony. The mean DRP at Step 10 (i.e., the last step that was considered) served as the best estimate of the converging point of each sensory condition. As in former studies, we considered DRPs as synchronous when falling into a 20 degrees interval around this point (see Nessler & Gilliland, 2009). Note that the converging point could vary within

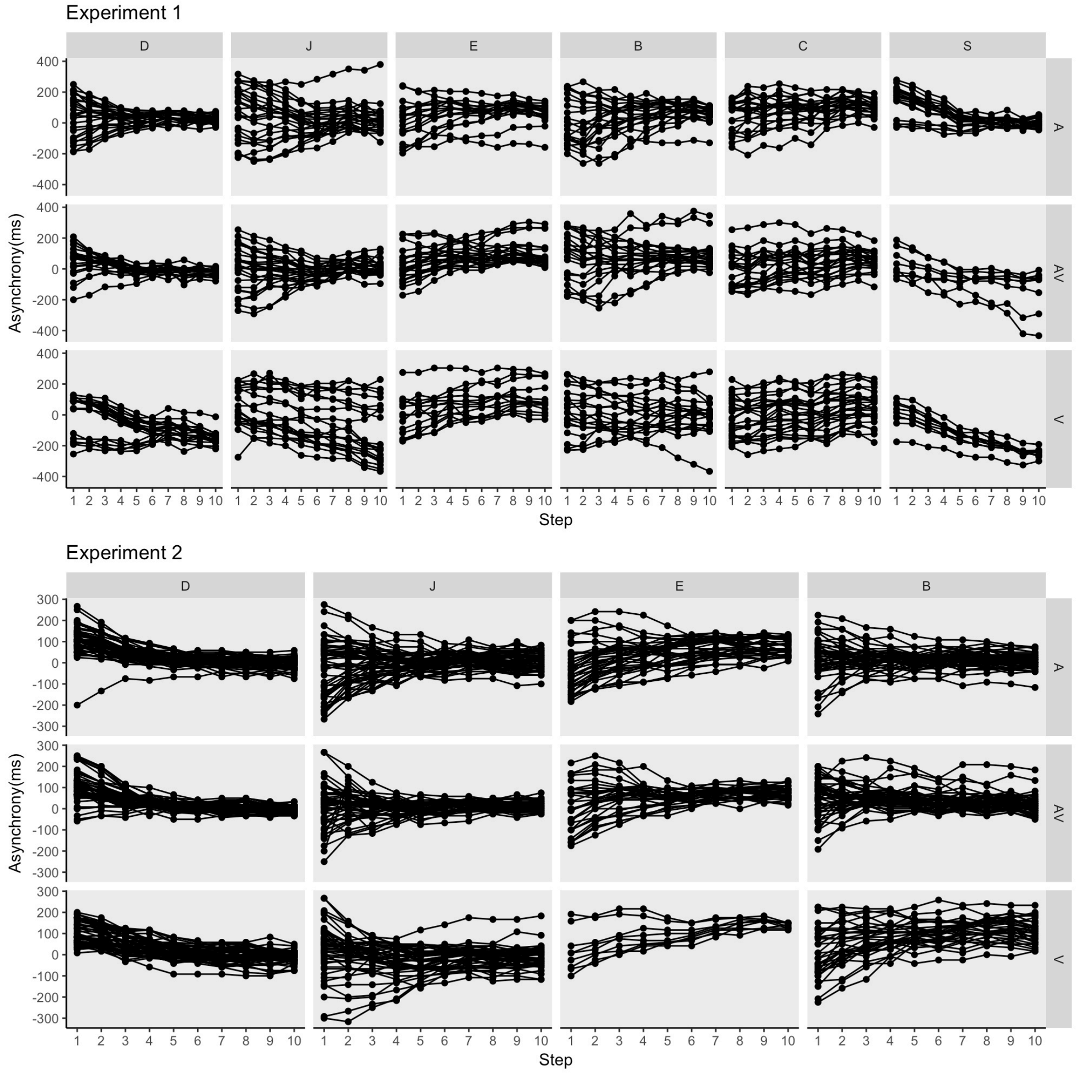


Figure 10: Asynchronies(ms) as a function of step number for visual (V), auditory (A), and audiovisual (AV) information of each participant of Experiment 1 and Experiment 2. Each line-segment represents one trial.

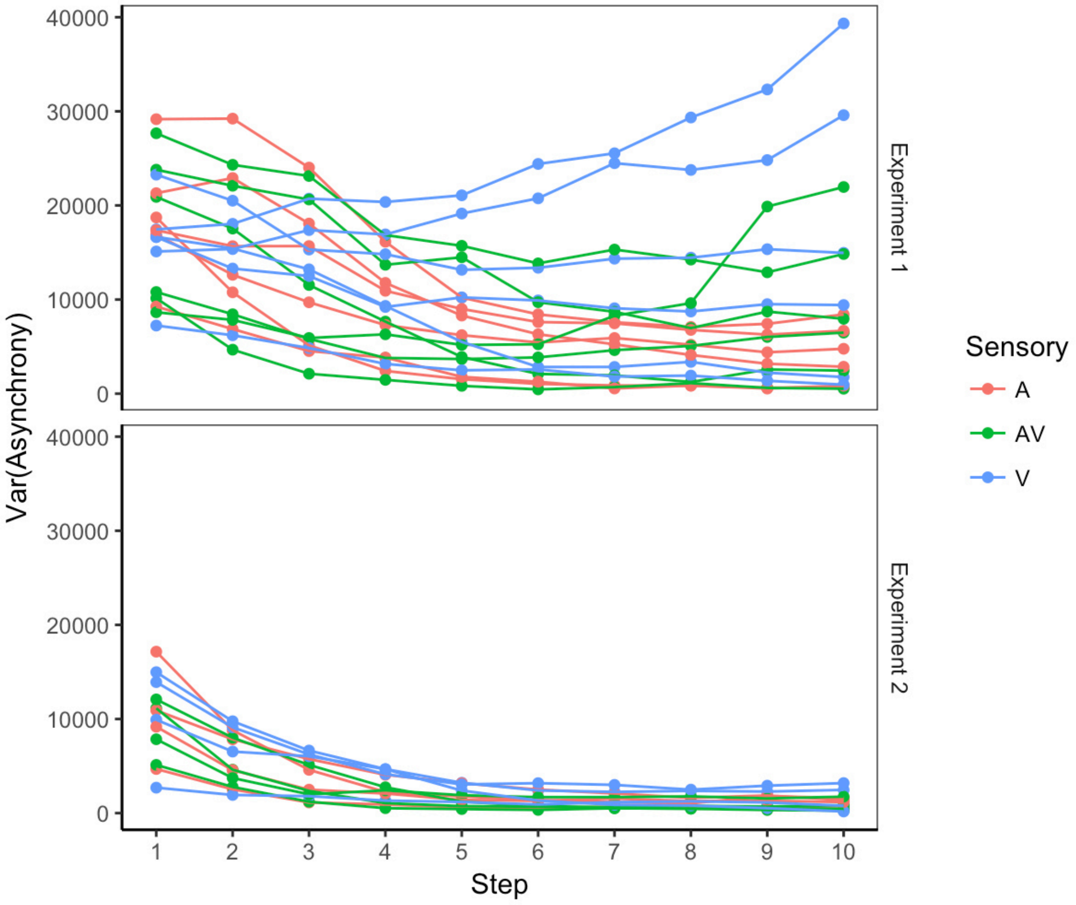


Figure 11: Asynchrony variance as a function of step number of Experiment 1 and Experiment 2. Each line-segment represents a participant. The sensory conditions are color-coded.

an individual between sensory conditions, probably due to perceptual processes (e.g., sensory system delays or focus of attention). We assumed that the asynchrony variability at the converging point computed across trials reflected an appropriate measure of synchronization variability.

Time to maximize synchronization. In addition to this variability measure, another measure of synchronization performance was, for our specific setup, how long it took until a participant maximized synchrony. This depended on the computation of the asynchrony and the selection and execution of adequate motor commands to reduce it. We expected that both processes are time-dependent and that synchrony should become maximal when a certain temporal threshold is reached. Considering that the underlying processes operate under noisy conditions (sensory noise, motor noise, etc.), synchrony can be formalized as a dichotomous random variable. The probability of observing more synchronized steps should increase with time. Then, the probability of synchronized steps can be represented by the proportion of synchronized steps across the trials. Proportions of synchronized steps were obtained by dividing the number of DRPs within the 20 degrees interval around the converging point by the total number of DRPs (see e.g., Nessler & Gilliland, 2009). A proportion was calculated for 5-time intervals ranging from 0s to 5s.

Trials that were synchronized since start were excluded because this way it was avoided that trials were considered in which maximal synchronization was achieved by chance. Moreover, trials were considered outliers and excluded when the asynchronies at the last step did exceed 3 s.d. from the converging point (see Appendix A.5 for a brief discussion of these outliers). Finally, two participants were excluded. One participant did not follow the instructions correctly. His data was not analyzed. A second participant was excluded because his performance was similar to that of a control condition in which the asynchronies were computed from heel strike onsets of randomly paired participants (see Figure 12 and Figure 13). Both show a constantly increasing asynchrony indicating that the individual (control) was not affected by the stimulus. After analyzing the results of Experiment 1 (which was conducted first), this participant actually reported that he forgot to synchronize. However, in Experiment 2, in which the same participant was explicitly called attention to synchronize, a similar pattern was found as in Experiment 1 (Figure 13). Therefore, we assumed that he was not able to execute the synchronization task correctly with any of the stimuli. After data exclusion, there were for each individual (Experiment 1, 6 individuals) at least 80 asynchronies in each sensory condition (8 for each step). Note that in most experimental conditions we registered many more asynchronies, that is up to 27.

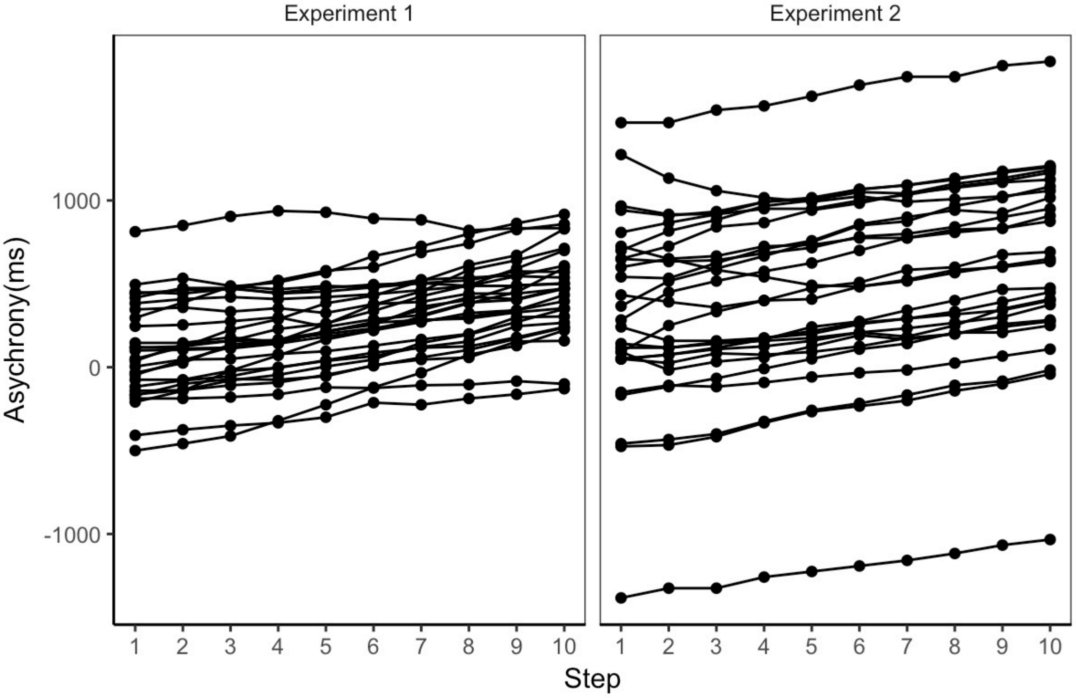


Figure 12: Asynchronies(ms) of Experiment 1 and Experiment 2 as a function of step number of 27 trials obtained from randomly paired participants.

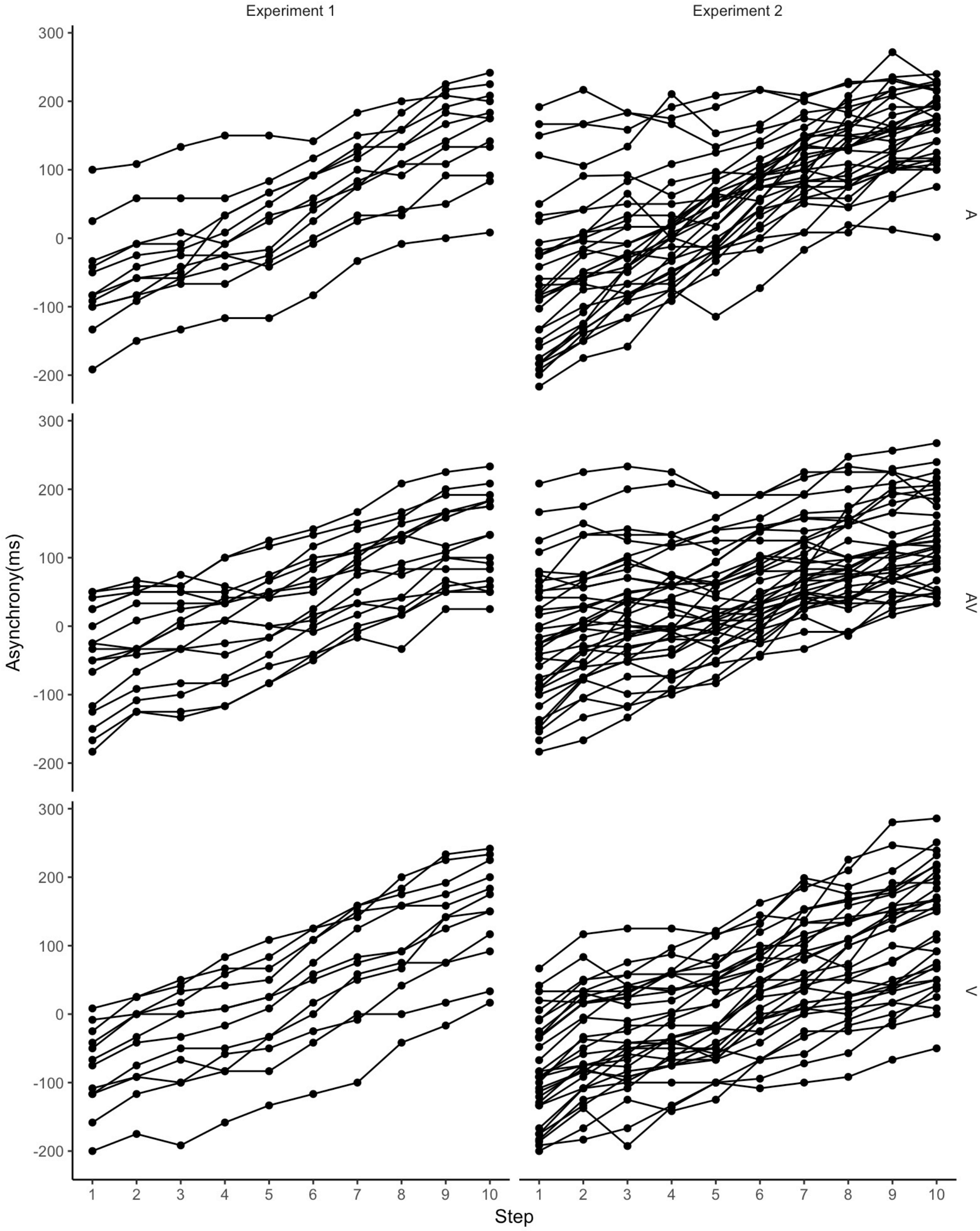


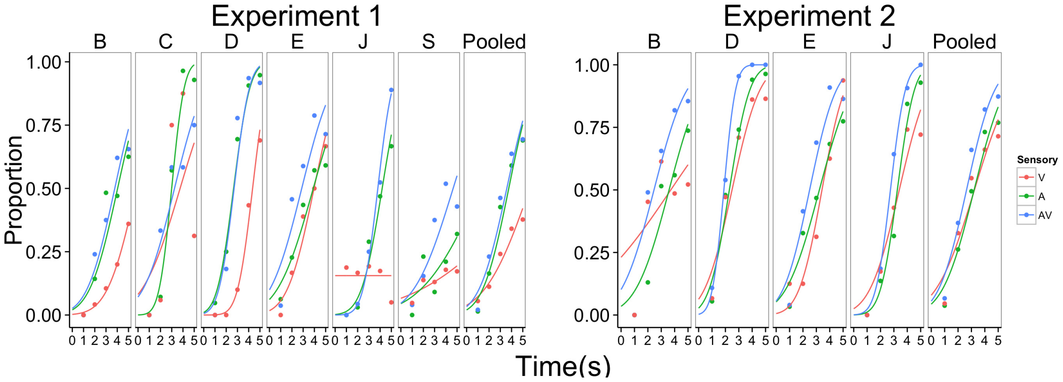
Figure 13: Asynchronies(ms) of Experiment 1 and Experiment 2 as a function of step number of the excluded participant.

0.2.3 Results

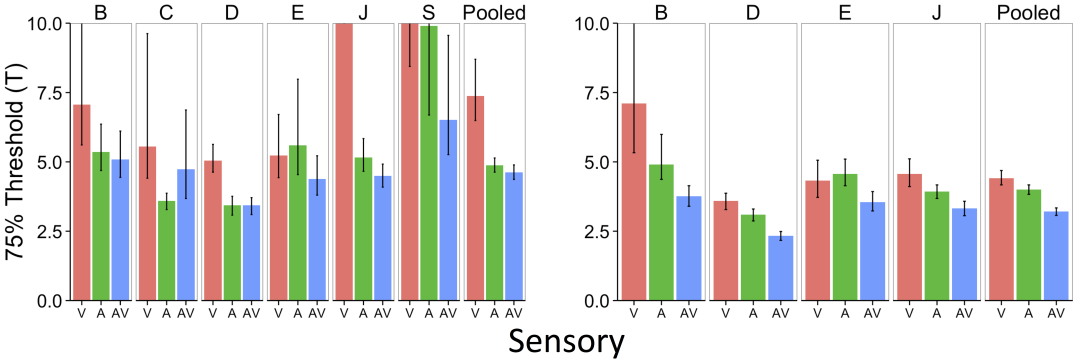
The aim of this experiment was to verify if synchronization improves over time and whether this process differs depending on the available sensory cues. In order to model the improvement, we fitted cumulative normal distribution functions to the proportions of steps considered as synchronous (cumulative Gaussians). We assumed that individuals attempt to minimize the perceived asynchrony as fast as possible. This moment is quantified by the point at which the cumulative Gaussian reaches 75%. This "threshold" ($P[X >= T] = 0.75$) represents the time at which 75% of the steps are synchronized (T). Here, T is interpreted as the time that is required to achieve synchronization.

In addition, once reached the synchronization threshold, the performance can be further assessed through asynchrony variability. Assuming that audiovisual integration should improve the precision of estimates of the heel strike onsets, integration processes might be manifested by lower variability. Thus, we captured two different aspects of the synchronization process. The first is quantified by T and is interpreted as the time that is required to synchronize. The second quantified by the s.d. of asynchronies reflects synchronization precision.

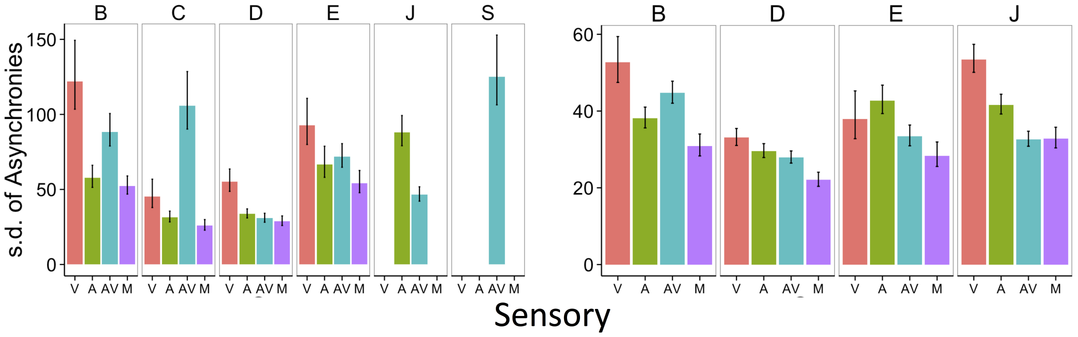
Qualitative assessment of the fitted Gaussians indicated an advantage of AV and A over V (Figure 14a, left). There might have also been a small advantage of AV over A for 5 out of 6 participants. The curves could be fully described by two parameters: A) the slope of the cumulative Gaussian (= the SD of the Gaussian distribution) indicates the form. It is frequently used in psychophysical studies to quantify perceptual sensitivity (see e.g., Mendonça et al., 2011). Here, such variable is not relevant. It reflects how quickly individuals switched from non-synchronized to synchronized states while we were interested in how quickly individuals synchronized overall. B) Any quartile of the curve (P[X >= T] = q) reflects the location on the abscissa at which q% of synchronized steps were reached. We calculated (P[X >= T] = 0.75) and interpreted it as synchronization threshold. T suggests that 5 of 6 individuals were faster synchronized in A than in V (Figure 14b, left). Three of 6 individuals were faster in AV than in A. Two individuals synchronized similarly in A and AV and for 1 individual performance in A was superior. Thus, considering the proportions obtained from DRPs, there may be a slight advantage in AV compared to A (Figure 14b, left). Overall, there were large variations within and



(a) The fitted proportions of synchronized steps against the time interval for each sensory condition (AV, A, V) for the individual and the proportions from the pooled observations. The curves are the cumulative normal distribution functions (cumulative Gaussians). For each condition, a dot represents one proportion obtained from at least 20 observations.



(b) Fitted 75% thresholds (T) obtained from the cumulative Gaussian. Errors bars are the 95% Bootstrap confidence intervals.



(c) S.d. of asynchronies obtained from the empirical observations. S.d. was calculated for asynchronies (in ms) of all steps occurring after time (T). Note that some conditions are lacking because in these cases T was estimated as being out of time range. M is the prediction of MLE model. Error bars represent the 95% confidence intervals (CI).

Figure 14: Synchronization performance with auditory, visual, & audiovisual cues for Experiment 1 (left) and Experiment 2 (right).

among all participants.

Next, s.d. of asynchronies was calculated from all steps occurring after time T (Figure14c, left).

To test the MLE Model, s.d. of asynchronies in AV was predicted by

$$\hat{\sigma}_{MLE} = \sqrt{\frac{\hat{\sigma}_A^2 \hat{\sigma}_V^2}{\hat{\sigma}_A^2 + \hat{\sigma}_V^2}},$$

where $\hat{\sigma}_{MLE}$ was estimated by the s.d. of asynchronies in the audiovisual conditions. S.d. of asynchronies obtained from the observations did not show a consistent pattern. MLE predictions (M in Figure 14c, left) failed for all but 1 participant, as indicated by 95% confidence intervals.

0.2.4 Discussion

In this first experiment, the focus was on performance differences during side-by-side walking when the available cues from the walking partners were auditory, visual, or audiovisual. We treated synchronization as a random variable and assumed that the probability to be synchronized increases with time. Cumulative Gaussians were used in order to describe this synchronization effect. From the models, we estimated the time required to synchronize. As the second indicator of synchronization, we measured s.d. of asynchronies directly from the observations. We expected that audiovisual cues should lead both to faster synchronization and to reduce variability.

Considering the time to synchronization, results revealed that participants were minimally faster with audiovisual cues compared to auditory cues. The absence of a clearer bimodal advantage contradicts the prediction of the MLE theory (see Bayesian Optimal Integration Hypothesis, Alais & Burr, 2004; Ernst & Bülthoff, 2004; Hove et al., 2013). These outcomes might be interpreted as a result of auditory dominance caused by a superior temporal processing. It would result in a higher reliability of the estimation of temporal cue onsets (see Modality Appropriateness Hypothesis, Welch et al., 1986).

According to MLE, the variability in estimating the onsets by multiple sensory cues is always lower than when using individual cues. Thus, even highly unreliable cues should positively contribute to the final estimate. Yet, in this first experiment, one participant was slower when presented with audiovisual cues compared with auditory cues alone. This fact indicates that audiovisual cues could be distracting or more demanding. Such assumption is strongly supported by the synchronization variability (s.d. of asynchronies) showing extensive deviations from our predictions for all but two participants (D & J).

However, artifacts of the auditory and visual cues in this experiment might have promoted the above-mentioned biases. PLWs preserve biomechanical and spatiotemporal properties of a walking person. They are appropriate for the study of e.g. recognition, velocity, and simultaneity judgments in focal vision (Mendonça et al., 2011; Silva et al., 2013). Yet, a crucial ability for PLW recognition is the extraction of structure from motion (Troje, 2008). Such ability was shown to be affected by eccentricity. Eccentricity is clearly related to the decreasing capability to resolve stimulus details when the distance from the fovea increases (Gurnsey, Poirier, Bluett, & Leibov, 2006).

Consistently, studies revealed perceptual deficits of PLWs when the eccentricity of PLWs was increased (Ikeda, Blake, & Watanabe, 2005), at least when the PLW was not magnified appropriately (Gurnsey, Roddy, Ouhnana, & Troje, 2008). Note that we did not instruct gaze direction in order to maintain the paradigm as natural as possible. Nonetheless, in a control analysis, we did not find any relation between head rotation and synchronization performance.

Moreover, in this first experiment, the image did not change perspective and the sound source location was fixed to the lateral right side of the participant. This implies that a) perspective and distant cues did not change as a function of the relative position of the participant and b) sound distance cues did not change at all. In addition, when for instances the PLW walked faster than the participant, the auditory delay should increase more than the visual delay due to the slower propagation of sound. Such real-time alignment of spatiotemporal signals did not happen. Incongruent signals can lead to the perception of asynchronies (Silva et al., 2013). This, in turn, should impair the integration of audiovisual signals (Spence, 2011) and therefore compromise the precision and accuracy of synchronization performance.

In sum, PLWs have been the standard stimuli in a wide range of perceptual experiments but might not be suitable for tasks that involve spatiotemporal processing, including side-by-side walking studies. Also, audiovisual incongruences might have impaired the synchronization performance. In Experiment 2, audiovisual cues were spatiotemporally congruent and changed in real-time as a function of participant behavior. In addition, further visual input was provided by replacing the PLW with a virtual mannequin stimulus.

0.3 Experiment 2

0.3.1 Methods

Participants

Five individuals (all male, $M = 29$, $SD = 2.3$), from Experiment 1 participated in Experiment 2. The experiments were separated by 7 to 9 month. Pilot testing with both setups with one “control” participant (Figure 15) and analyses of sequence effects (Figure 16) did not indicate any significant synchronization improvement through previous training.

Material and Stimulus

The experiment was programmed in Python (*Python 2.0*, 2016) and Blender’s logic bricks. BlenderVR 2.73 (Katz, Flinto, Tourain, Poirier-Quinot, & Bourdot, 2015) was used to coordinate and distribute the execution of the virtual environment.

A virtual mannequin (see Figure 17) was created to be used as visual stimulus. The spatiotemporal coordinates from the PLWs were used to determine the joint positions of the mannequin. The joints were connected by skin-colored cylinders, with relative sizes approximately proportional to the morphological dimensions. Because in Experiment 1 the joint position and the head were represented by small black dots and the virtual mannequin was built upon these dots, the virtual mannequin was larger. Increasing the size is one mean to magnify stimulation so that visual discrimination performance becomes equal across the entire visual field (Gurnsey et al., 2008). Thus, the larger virtual mannequin increased the sensitivity for perceiving visual cues. Moreover, like in a real-world scenario, body segments that were closer to the participant occluded segments that were farther away, providing additional depth cues.

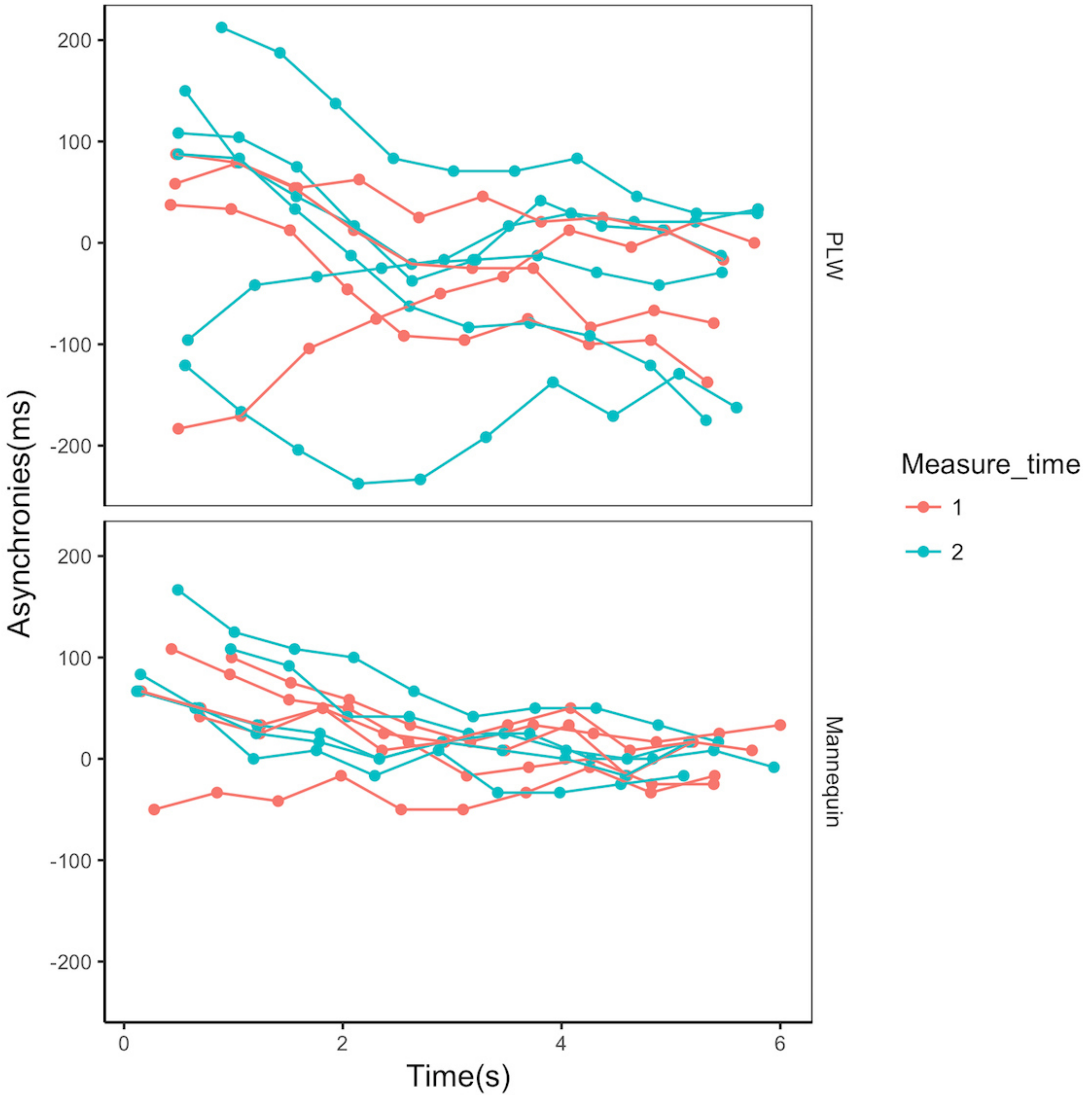


Figure 15: Asynchronies(ms) as a function of time(s) separated by the time of measurement and stimulus (PLW and Mannequin). The asynchronies were obtained from the performance of a participant who synchronized 4 to 6 times in the AV sensory condition with the two types of stimuli (PLW and Mannequin) at two measurement times (Measure time 1 and Measure time 2). Measure time 2 was 7 months later than Measure time 1. The plot illustrates that synchronization performance with the same stimulus does not seem to differ between the measurement times.

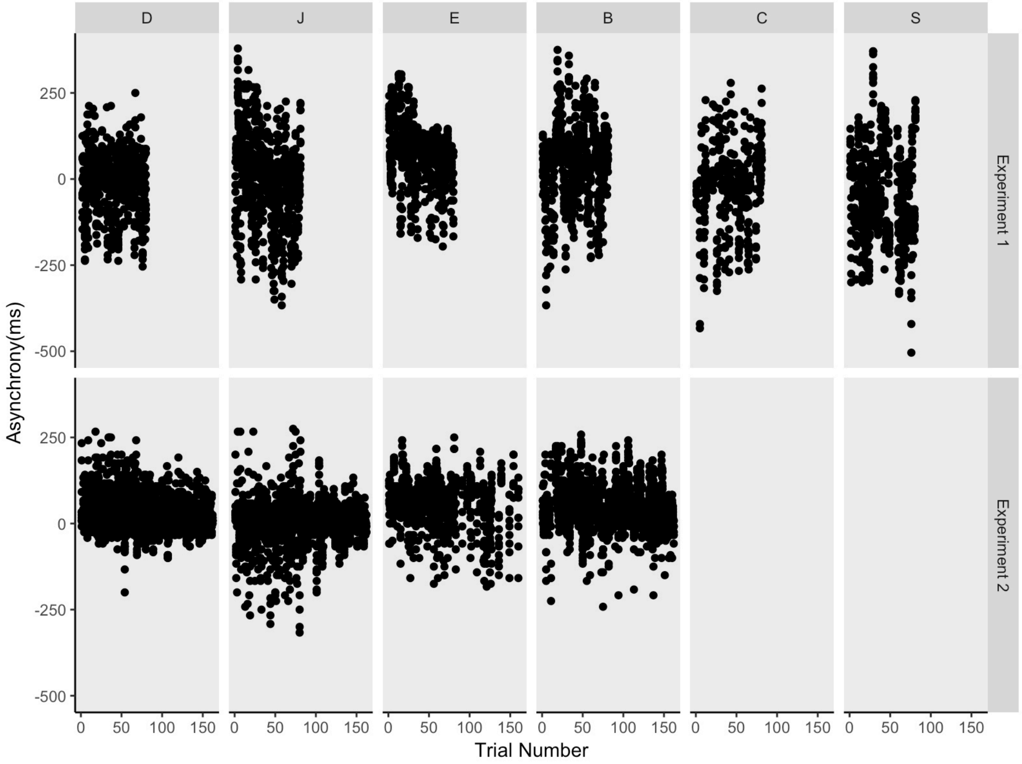


Figure 16: Asynchronies(ms) as a function of trial number, for each participant of Experiment 1 and Experiment 2. The plot illustrates that there is no consistent pattern of asynchrony change as the trial number increases. This suggests the absence of sequence effects.

The perspective for the projection of the mannequin and the sound source location were computed in real-time based on the relative position and the head rotation of the participant. To do so, the head coordinates of the participant were tracked by a Vicon motion capture system using Nexus 2.0 (*Nexus 2.0*, 2016). In order to synthesize sound properties, an auralization process using non-individualized HRTFs was used from (Oliveira et al., 2013) that included a simplified geometrical model of the experimental environment (e.g., reflections, distance, latencies). To prevent delays during online auralization, sound samples for 450 different positions (5 distances relative to participant [-100cm, -50cm, 0cm, +50cm, +100cm] * 90 head orientations of participants) were previously created and the appropriate ones played during the trial. In sum, both visual and auditory signals provided additional distance cues and an improved spatiotemporal

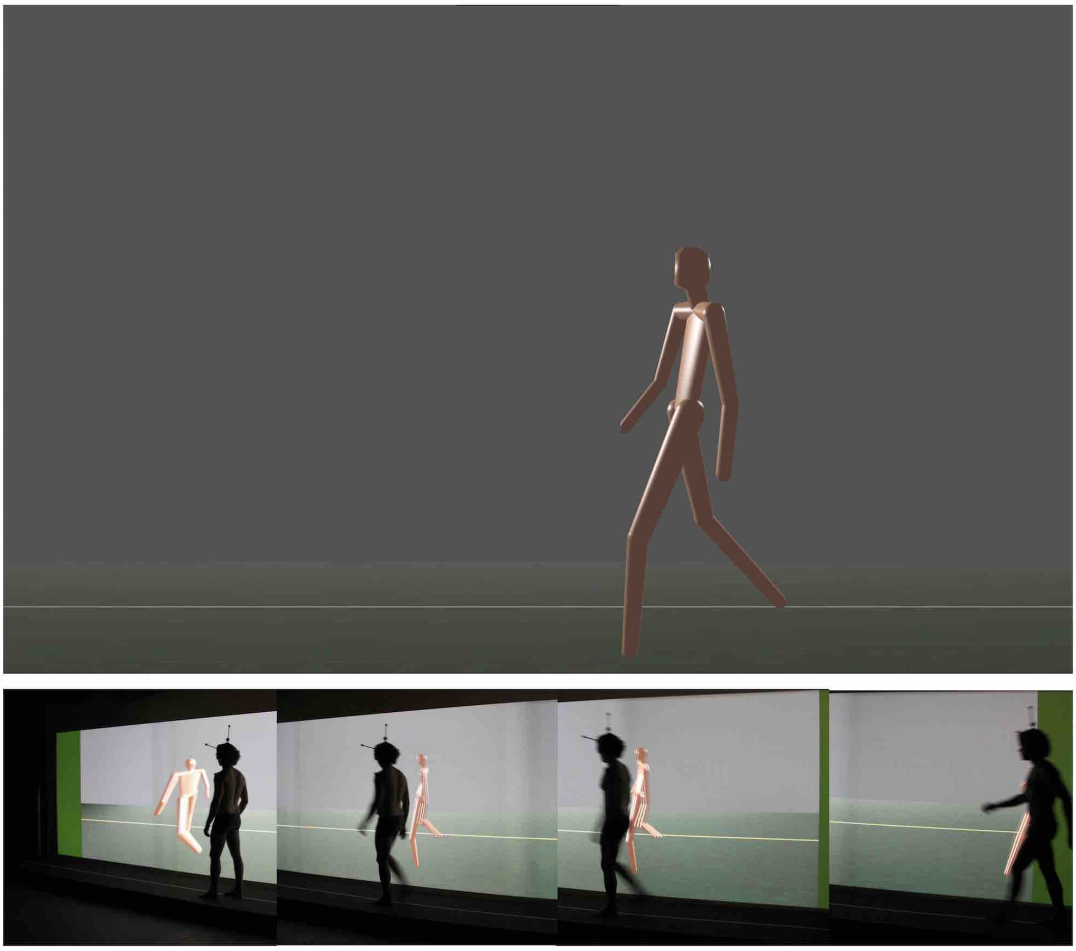


Figure 17: Stimulus and setup of Experiment 2. Top: Screenshot of the virtual mannequin from the computer screen. Bottom: Four snapshots of a participant walking next to the virtual mannequin projected on the screen. Note that the perspective changed as a function of position, which is well illustrated in the fourth snapshot (bottom left). The “antenna-like device” on the head of the participant delivered most reliable position and head rotation coordinates for the online render of image and sound.

congruency was achieved.

However, as expected in an immersive virtual environment, end-to-end system delays did occur from the motion capture to the update of stimulus presentation. The latencies for changing the perspective of the mannequin were of 93ms (4sd) and of 50ms (1sd) for the sound of footsteps. Therefore, to preserve the congruency in the audiovisual condition a further delay of 43 ms was applied to the sound signals in both A and AV condition.

Design and Procedure

Before the experiment, participants were trained in order to acquaint to location information of the auralized sound. Samples were presented at azimuth (30 degrees, 90 degrees, & 150 degrees) and elevation -72 degrees, relative to the position of the right ear.

The experimental design was the same as Experiment 1, i.e. three different start phases of the stimulus and three sensory conditions, i.e., footstep sounds (A), virtual mannequin (V), or combined (AV). Each condition was presented three times in three blocks, repeated in two sessions constituting 162 trials. The presentation was pseudo-randomized within a block. We expected an improvement in V due to the richer visual cues. Performance in A should increase due to spatiotemporal correspondence. Finally, AV should lead to higher performance than A or V alone due to the integration of both signals, as predicted by MLE.

Analysis

Spatial Analysis.

Similar to Experiment 1, we first assessed whether any spatial variables have to be controlled in order to conduct the temporal analysis. Therefore, the distances between the head position and the differences between stride lengths were analyzed. The distances were analyzed at each step providing (Step[10] x Sensory condition [3] x Start phase [3] x Rep [3] x Block [3]) 1620 distance records for each participant. The stride length differences were analyzed at each stride providing (Stride[4] x Sensory condition [3] x Start phase [3] x Rep [3] x Block [3]) 648 stride length differences.

The distances between the head positions gradually increased as a function of step number (Figure 3). The size of this increase differed among the participants. The mean absolute difference at step 10 was between 0.01m and 0.35m (Figure 4). For all five participants, A led to larger differences than AV and V. The strides did not change as a function of stride number (Figure 6). The last stride difference (stride 4) was different among the sensory conditions for four of five participants (Figure 7). For these participants, A led to larger differences than AV and V. The differences were between 0.035m and 0.22m.

Overall, the pattern was very similar to the pattern found in Experiment 1. Although there was a slight effect of the sensory condition, we concluded, as before, that the distance differences of both measures were small. In Appendix A.4 we demonstrate the absence of any relation between both measures. Therefore, those spatial adjustments seemed to be marginal for the scopes of this study and that the subsequent temporal (synchrony) analysis should not be affected that much.

Temporal Analysis.

A within-subject design (Steps [10] x Sensory [3] x Start [3] x Rep [3] x Block [3] x Sessions [2]) was used. It provided 1620 records of asynchronies for each participant. Exclusion criteria were the same as in Experiment 1 (see Appendix A.5). As mentioned in Section 0.2.2, one participant (R) was excluded because his performance was similar to that of a control condition in which the asynchronies were computed from heel strike onsets of randomly paired participants (see Figure 12 and Figure 13). The stepping frequency was matched since the start for 100% (D), 96% (B), 54% (E), and 90% (J) of the trials. Participant E demonstrated a considerably low proportion of frequency-matched trials compared to Experiment 1 because he applied a slightly different synchronization strategy in Experiment 2. Rather than starting with the same cadence since start, in Experiment 2 he often started walking very slowly and cautious and then got synchronized very quickly. At later steps, his behavior was similar to that of the others. However, to maintain consistency and have comparable results with Experiment 1, we subsequently conducted the phase locking analysis on the frequency locked trials.

0.3.2 Results

As previously, in order to capture performance differences in synchronization, we a) obtained the 75% Threshold (T) from the cumulative Gaussian functions and b) calculated the standard deviation (s.d.) of asynchronies.

Considering the threshold (T), 3 of 4 participants were faster in A than in V (Figure 14a, right & Figure 14b, right). All participants were faster in AV than in V and A (for pooled observations: V $[T = 4.4s]$ > A $[T = 4s]$ > AV $[T = 3.2s]$; A-AV: Bootstrap: $p < .001$). The pooled observations highlight the improvements in synchronization compared to Experiment 1 (Figure 14, left) for all sensory conditions, as indicated by 95% Bootstrap confidence intervals (Exp1: V $[T = 7.38]$ > A $[T = 4.87]$ > AV $[T = 4.62]$; Exp1 - Exp2, Bootstrap: $p < .001$).

For Experiment 2, the s.d. of asynchronies in AV was lower than in A and V, and in A it was lower than in V for 3 of 4 participants (Figure 14c, right). M is the prediction of the MLE. It correctly predicted s.d. reduction in AV for 2 (J & E) of 4 individuals. In addition, it pointed toward the correct direction for another individual (D) but here confidence intervals of AV and M did not overlap. All sensory conditions showed reduced s.d. of asynchronies compared to Experiment 1.

0.3.3 Discussion

In this Experiment 2, a virtual mannequin substituted the PLW and both visual and auditory stimuli locations and perspective were updated in real-time depending on the head coordinates of the participant. First, the modifications of the stimuli in Experiment 2 increased synchronization performance compared to Experiment 1. All four individuals improved in synchronization velocity and variability. Second, in Experiment 2 all participants synchronized faster with audiovisual cues. Although the audiovisual advantage is consistent with the optimal integration theory, estimates of the time required to achieve synchronization (T) by an MLE model is not meaningful here.

According to MLE, the effect of cue integration should be manifested in an optimal reduction of variability of the sensory representation (Ernst & Bülthoff, 2004). This, in turn, should lead to more precise timing (Elliott et al., 2010). Consistently, three of four participants synchronized more precisely when audiovisual cues were provided.

On the other hand, for only two of four individuals the MLE estimates matched the asynchrony variability of the audiovisual condition.

These inconsistencies might result from methodological shortcomings. In order to maintain sensory inputs as natural as possible, we investigated over-ground rather than treadmill walking. Because we conducted the experiments in a virtual environment, the walking distance was constrained to 7.2m (Experiment 1 and 2). It could be that measures of variability were affected by the reduced number of steps.†However, Figure 14c shows that the asynchrony variability consistently stabilized at minimal values after 5 steps. This suggests that 10 steps might be sufficient to maximize synchrony.

Since the PLW/mannequin was different across participants, another possible limitation is that some participants were trying to match signals that were more variable than that of other participants. Yet, the models, from which the stimuli were generated, were able to maintain an extremely constant pace (this was actually a model selection criteria). The step interval variances of all employed models were 1ms (D & E), 3.7ms (B), 3.4ms (C), 1.6ms (J) and, 2ms (S). In addition, we did not find any relation between these variabilities and the synchronization results. Although this does not rule out that the variability of other body segments may have produced some noise, it indicates that stimulus variability might have been a less relevant noise factor.

A more plausible explanation for the inconsistent results is that the MLE model implemented here does not allow the best fit. Asynchrony variability was computed from the steps. It specified the weights for each modality in the MLE model ($\hat{\sigma}_A$ & $\hat{\sigma}_V$). This model predicted then the variability of asynchronies with audiovisual cues ($\hat{\sigma}_{MLE}$). Thus, predictions were based on the asynchrony variability when the participants were trying to synchronize with visual cues and auditory cues alone.

However, for estimating the temporal onsets of the own heel strikes, audiovisual cues might be marginal. Other relevant cues are provided by the vestibular system, the proprioceptive system, and the somatosensory system. While each system might provide ambiguous spatiotemporal cues, their combination should allow much less ambiguous estimates of the heel strike onsets. Thus, an adequate MLE model should include parameters of the reliability of estimates with each and all of these cues within a cross-modal framework. In short, the MLE model used here did not account for all the perceptual variables involved in the estimation process.

In addition, the parameter estimates of the MLE were the variability of the observed asynchrony. Asynchronies were computed from the time difference between the motor responses (i.e., stepping pattern). Different subprocesses within the perception-action loop could cause the variability of the motor response. Variability may be inherent to the encoding of the events, which is modality dependent. However, it also can be caused by the time-keeping of temporal intervals or the motor responses implementation. The two latter processes are less dependent on the modality than the former process. Therefore, specifying parameters of the MLE model by the variability that is only caused by perceptual processes, might lead to an overestimation of variance reduction by the MLE model for perception-action loops (Elliott et al., 2010).

Nevertheless, variability reduction was overestimated only for two participants (B & D). In addition, for one participant (B), variability was lower for auditory cues when compared to audiovisual cues. The fact that an unimodal cue condition revealed lower variability indicates that also in Experiment 2 the visual stimulus may have caused some distraction or additional load in the audiovisual cue condition. The spatiotemporal congruency of information provided by cues on each individual modality cue and/ or combined was increased in Experiment 2 compared to Experiment 1. Yet, the visual perspective and the position of footsteps sound were updated with a delay of ∼90ms. Participants frequently rotated the head to the right side up to 70 degrees relative to the walking direction. This mostly happened in conditions in which the visual stimulus cues were available. For the visual condition, this implies that the perspective was updated with a delay. This might have produced some marginal noise. However, for the audiovisual condition, a fast head rotation with delayed updates might have promoted incongruence between auditory and visual cues. During head rotation, sound cues indicated the source as lateral to the ear, instead of being congruent with the visual stimulus by appearing slightly in front of the right ear. The rotation lasted only a few tenths of seconds but it might have been sufficient to create additional noise during heel strike.

In conclusion, the present experiment clearly demonstrated that higher synchronization precision is achieved by the combined presentation of congruent audiovisual cues, compared to auditory or visual cues alone. The MLE model suggests that this occurred due to cue integration. Our results confirm partly these assumptions but there were several sources of noise that prevent more robust conclusions.

As mentioned in 0.1.6, the integration of signals is of advantage when they are coming from the same event. Signals are not integrated when the cues indicate a temporal separation of the event (Berniker & Kording, 2011) in order to prevent the erroneous integration of cues from different sources (Elliott, Wing, & Welchman, 2014). The maximal temporal separation at which signals are integrated is called window of temporal integration (WTI) (Vroomen & Keetels, 2010). To examine whether the benefits of audiovisual cues were promoted by their integration, we conducted a third experiment. In Experiment 3, the visual and the auditory stimuli were presented with different levels of temporal onsets. Conditions in which the temporal asynchronies between auditory and visual signals were small should reveal lower synchronization variability, compared to conditions in which the temporal asynchronies were large.

0.4 Experiment 3

0.4.1 Methods

Participants

Three of the four individuals of Experiment 1 and 2 participated in Experiment 3 (all male, age: $M = 29$, $SD = 2$). The experiment was conducted three months after Experiment 2 and pilot tests showed no noticeable training effects.

Material, Stimuli, & Design

The available auditory and visual cues were the same as in Experiment 2 but the footstep sounds (A) and the virtual mannequin (V) were presented throughout all trials (AV). Here, A or V was temporally phase shifted. That is, the heel strike of the virtual mannequin and the footstep sound were displayed with disparate temporal onsets. Phase Shifts were a) in V or in A, b) positive or negative, and the amount of phase shift ranged from -250ms to +250ms in 50ms intervals (i.e., -250, -200, -150, -100, -50, 0, +50, +100, +150, +200, +250), constituting 23 conditions. Negative values signified that the shifted stimulus was presented earlier than the non-shifted stimulus; for positive values it was the opposite (see Figure 18). Each condition was repeated 10 times making up 230 trials presented in a pseudorandom order.

Procedure

As in Experiment 1 and 2, the participant started walking from 2.4m before the screen. Then, when passing a threshold of 0.3m before screen start, the AV stimulus was presented spatiotemporally synchronous for 3.3m. In this way, the participant had approximately five steps to get synchronized. The Figures 10 and 11, and the time to synchronization thresholds (T) (Figure 14a) illustrate that this should be sufficient. When the participant passed a 3.3m threshold, a phase shift was applied to V or A according to a predefined value. As a control, one synchronous condition (0ms phase shift) was included. In order to remove artifacts, the stimuli were occluded immediately before phase shift. That is, when the virtual trajectory of the stimulus passed 3.3m, the footstep sound disappeared for one step and the virtual mannequin disappeared behind a green square of 0.6m x 2.1m. During occlusion, the phase shift was applied and then the virtual mannequin re-appeared from behind the square and the footsteps sounds were presented again for 3.3m. Participants were instructed to: “synchronize steps with the mannequin and the footsteps sounds and maintain the smallest distance to the stimuli. If footstep sounds and mannequin are asynchronous, synchronize with which you feel more comfortable at that moment”.

0.4.2 Analysis

The walkway was 0.6m larger compared to the previous experiments. As in Experiment 1 and Experiment 2, in Experiment 3 were analyzed 10 steps (∼5s); 5 steps (∼2.5s) before phase shift and 5 steps (∼2.5s) after phase shift. We determined the effects of phase shift of auditory (A) and visual (V) cues on synchronization performance by comparing it with the performance before phase shift. For this reason, it first had to be determined whether synchronization was achieved before phase shift. The converging point was estimated by calculating the average of the asynchronies at the last step before phase shift. Figure 19 indicates the presence of a converging point at step 5.

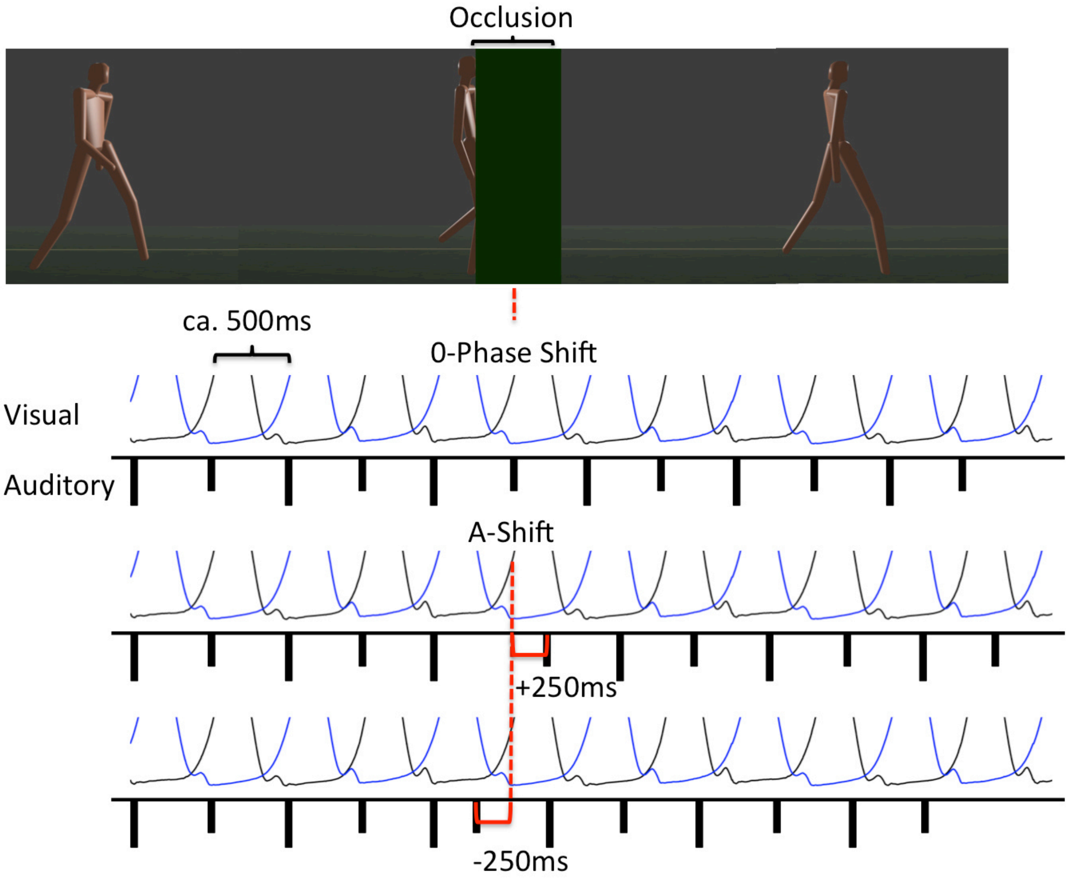


Figure 18: Stimulus manipulations of Experiment 3. Top: Presentation of virtual mannequin (V) and occlusion area (green). A Phase Shift (red) was applied during occlusion (green) of the stimuli. Bottom: Time series of the vertical ankle displacement (blue: left foot, black: right foot) (V) and footstep sounds representing the heel strike onsets (black stripes) (A). In red is exemplified the position of a positive (+) and negative (-) 250ms Phase Shift of A (A-Shift).

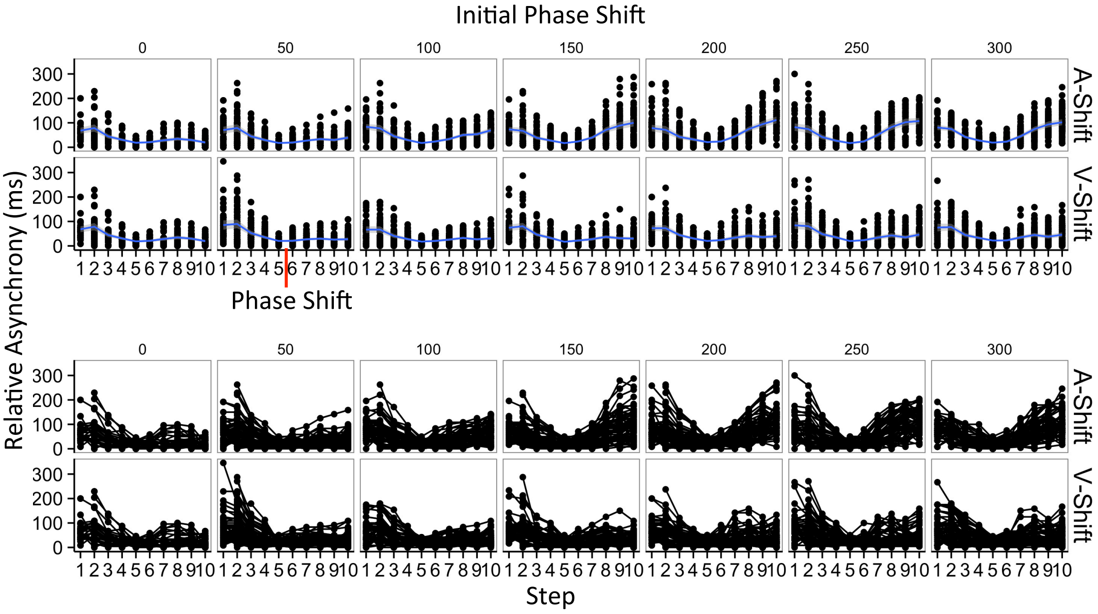


Figure 19: Absolute relative asynchronies of Experiment 3. These were obtained by subtracting the mean asynchrony, at the last step before phase shift, from each asynchrony. The pooled observations illustrate that most asynchronies approximate zero at step 5 and deviate from zero after phase shift.

Afterwards, the relative asynchronies were transformed into DRPs. As in Experiment 1 and 2, each trial in which the DRP at the last step deviated less than +-20 degrees from the converging point was considered as "synchronization succeeded" and was included in the further analysis. This was 100% (D), 97% (J), and 89% (E) of all trials. Therefore, for the subsequent analysis were considered 223 (D), 205 (J), and 163 (E) trials.

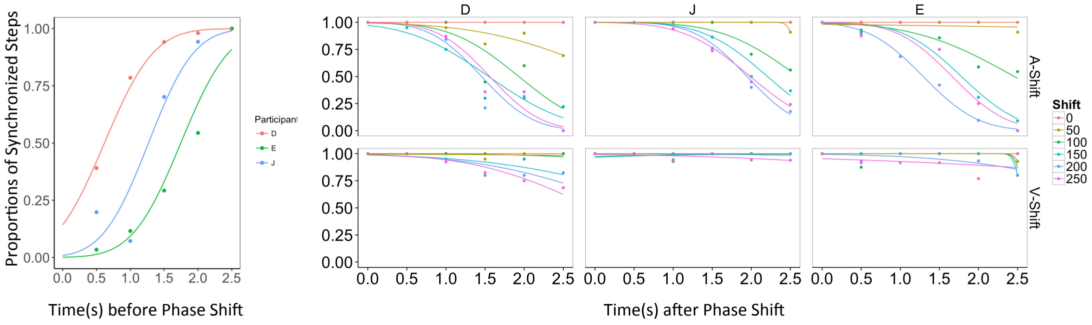
0.4.3 Results & Discussion

Experiment 3 was conducted in order to examine whether temporal coincidence of audiovisual cues improves synchronization through their integration. Temporally incongruent cues should lead to increased synchronization variability because their temporal separation prevents cue integration (Elliott et al., 2014; Ernst & Bülthoff, 2004). Note that the methodological shortcomings discussed in Experiment 2 also account for some variability in Experiment 3 since the same stimuli and a similar setup. Phase shifts were implemented to create the temporal asynchrony. Thus, we first estimated the time to achieve synchronization before phase shift from the proportions of synchronized steps. For each proportion, there were ∼100 trials. The 75% synchronization threshold was T = 1.01s (D), T = 1.63s (J), and T = 2.13s (E) (see Figure 20a, left).

The thresholds were smaller (faster synchronization) than in Experiment 2 mainly because of the implemented cut-off criterion through which we excluded all non-synchronized trials before phase shift. The criterion was such high because we were here interested in the effects of the phase shift and not on how much synchronization could be achieved overall. After phase shift, synchronization could be maintained or impaired. The impairment was determined against the non-shifted modality. The proportion of steps that were no longer synchronized quantified its probability of occurrence. We expected that it increased with the size and the time elapsed after phase shift. Cumulative Gaussians were fitted to the proportions of the asynchronous steps. The time at which 25% steps were not synchronized was estimated from the curves and termed 25% De-synchronization Threshold (DT).

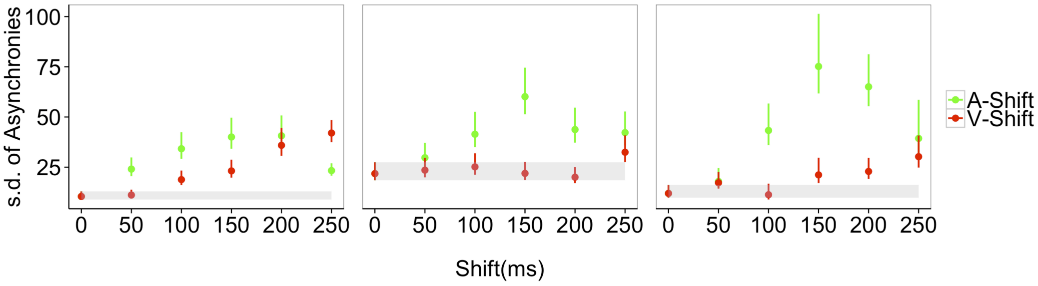
Overall, the cumulative Gaussians fit well the proportions of A-Shifts but not of V-Shifts (Figure 20a, right). A-Shifts mean that the auditory cue was shifted relative to the visual cue. For V-Shifts it was the opposite. Smaller A-shifts compared to V-Shifts led to DT. In addition, A-Shifts led to faster DT when being larger than 50ms. Thus, when cues were temporally separated, individuals' heel-strike onsets seem to shift into the direction of cues indicated by the auditory modality. Again, this suggests a stronger influence of auditory cues for the control of synchronization. In cases of DT, it is difficult to study cue integration because it cannot be clarified whether the time was sufficient to maximize synchronization. Nevertheless, it is unlikely that cue integration had happened within the available 2.5s after phase shift, considering that individuals required up to 2.13s to achieve synchronization with congruent audiovisual cues before phase shift. The asynchrony pattern in Figure 19 points towards that another converging point was not reached.

Next, we considered the upper limits of the 95% Bootstrap confidence interval of DT (Figure 20b, top). At most V-Shifts and smaller A-Shifts, the confidence limits exceeded the domain of the applied model, which was 20s. When DT was larger than 20s, we labeled it as infinite. An infinite DT indicates that synchronization was maintained



(a) Cumulative Gaussians fitted to the proportions of synchronized steps as a function of the time intervals before Phase Shift (left) and after Phase Shift (right). After Phase Shift, synchronization is defined relative to the non-shifted modality. In A-Shift, the auditory stimulus was phase shifted forward or backward. In V-Shift, the visual stimulus was phase shifted. Forward and backward shifts did not reveal differences and were averaged. The curves were obtained by fitting cumulative Gaussians to the proportions of de-synchronized steps. For illustration purposes, observations and predictions were inverted thus displaying the proportion of synchronized steps.

Upper CI Limit of 25% Threshold (DT)	D					J					E				
Shift	**50***	**100**	**150**	**200**	**250**	**50***	**100**	**150**	**200**	**250**	**50***	**100**	**150**	**200**	**250**
A	3.3	1.7	1.3	1.3	1.3	-	2.4	2	1.8	1.8	-	2.3	1.7	1.1	1.6
V	-	-	8.8	4.2	2.4	-	-	-	-	-	-	-	-	-	-



(b) Top: Upper confidence interval of the 25% De-synchronization Threshold (DT) as a function of Phase Shift. DT is the estimate of time (s) at which 25% of steps are not synchronized. The estimation was made for a time interval of maximal 20s. Dashes (-) indicate when DT exceeded this domain and hence did not occur. Asterisks (*) indicate when the 95% CI of A-Shifts and V-Shifts did overlap (for CI values, see Appendix A.7). Bottom: S.d. of asynchronies obtained from the observations at the last step for A-Shifts (green) and V-Shifts (red). The gray-shaded region is the 95% CI of the 0-Shift condition.

Figure 20: Effects of phase shifts on synchronization performance

with the non-shifted modality. Here, integration effects could be determined.

As previously, synchronization performance was quantified by the s.d. of asynchronies obtained from the observations. Results of Experiment 2 had revealed that s.d. with congruent AV was lower than with A. Here, we observed that also with small shifts, variability equaled AV. That is, below 100ms (D), 250ms (J), and 150ms (E) shifts, s.d. was within the 95% confidence interval of the 0-phase shift condition (i.e., the s.d. is within the gray-shaded area in Figure 20b, bottom). Note that the s.d. is not directly comparable between Experiment 2 and 3 because, in Experiment 3, all trials were excluded in which synchrony was not reached before phase shift. Nevertheless, these results indicate that cues became integrated because it maintained the variability low (Ernst & Bülthoff, 2004). When synchronization was maintained with the non-shifted modality, as indicated by an infinite DT, and the asynchronies between the cues became larger, s.d. increased. This indicates that cues with larger shifts functioned as distractors.

Finally, at A-Shifts between 200ms-250ms, s.d. decreased again. This may owe to a wear-off of the distractor effects. For V-Shifts, s.d. increased remarkably slower and did not decrease. These patterns are similar to findings of previous studies about finger-metronome synchronization using target-distractor paradigms (Bertelson & Aschersleben, 2003; Repp & Keller, 2004; Kato & Konishi, 2006; Hove et al., 2013). The asymmetric distractor effect could be attributed to a superior ability of the auditory system to extract temporal structure from isochronous stimulus sequences (Grahn, Henry, & McAuley, 2011; Su, 2014). Then, a temporal shift in a widely regular step sound leads to the perception of disruption of a sequence. This elicits stronger error correction processes than when the disruption is not being perceived (Repp, 2005). Temporal shifts of the visual mannequin may not have been perceived that fast.

In conclusion, both, the asymmetric effects on DT thresholds and s.d. of asynchronies clearly suggest a stronger reliance on the auditory modality. When DT did not happen, cues might have been integrated serving as an aid when being small and as distractors when being larger. Overall, the results of Experiment 3 offer further support to the claim that the temporal onset of audiovisual signals is crucial and that their misalignment reduces synchronization performance.

The stronger reliance on auditory information might be particular to the present experiments. Here and in most studies, synchronization during walking was quantified by considering the temporal asynchrony of discrete events of the gait cycle. These are usually the heel strike onsets (Murray-Smith et al., 2007; Nessler & Gilliland, 2009; Nessler et al., 2011; Nessler & Gilliland, 2010; Varlet & Richardson, 2015). To our knowledge, the only exceptions are van Ulzen et al. (2008) assessing the continuous displacement of the lower legs and Zivotofsky et al. (2012) assessing trunk acceleration.

When walking together, a principal objective seems to be, besides navigation, the maintenance of an interpersonal comfort space (see e.g., Hall, 1966 for seminal research in Quesque et al., 2016). For achieving it, the position of the trunk and the head should be more relevant than the feet. It was shown that the maintenance of an interpersonal distance during forward-backward movements is mediated by global motion information of the partner's body and that the head displacement might be controlled when facing the interaction partner during locomotion (Ducourant, Vieilledent, Kerlirzin, & Berthoz, 2005). Also, during conversations, game playing, and dancing, the shared activity of head, trunk, and center of body mass (COM) were usually the variables of interest (Shockley & Fowler, 2003; Shockley, Baker, Richardson, & Fowler, 2007; Schmidt, Nie, Franco, & Richardson, 2014). Furthermore, although heel strikes appear to be the most salient events within the gait cycle, the control of locomotion does not seem to originate from them (Ivanenko, Poppele, & Lacquaniti, 2004). It was also claimed that the active involvement of higher brain functions during walking is restricted to sending a reduced set of "global" commands that affect the CPG activity and reflexes (Lacquaniti, Ivanenko, & Zago, 2012; Kandel et al., 2000; Latash, 2012; Flash & Hochner, 2005). Thus, while synchronization during side-by-side walking is mostly studied by assessing the variability of asynchronies between the heel strike onsets, it is not assured that these are the control variables with the highest priority.

In the previous three experiments of this thesis participants were instructed to synchronize without that we gave further instructions which variables—i.e. the part of one's body and the walking partner's body—should be focused to achieve synchronization. When only visual cues were provided, it is less plausible that individuals synchronized heel strike onsets. The exact event—when the heel contacts the ground—is difficult to derive from the visual information of the foot motion. Furthermore, people usually do not focus their feet and the feet of the walking partner during walking (Lappe et al., 1999). As previously mentioned, most visual cues that the individual receives from one's movements are provided by optic flow. Optic flow is integrated with vestibular, somatosensory, and auditory cues to perceive self-motion (Cullen, 2012). It could be, for example, that a continuous self-motion signal was matched with other signals from the walking partner.

Previous perceptual judgment studies have shown that the gender of PLWs can be identified by considering a few motion signals that represent the co-variation of several joint kinematics (Troje, 2002). These signals can be captured by matrix factorization techniques like the Principle Component Analysis (PCA). The first principle component obtained by PCA is the dimension of most co-variation of elementary variables like the joint displacement. It was shown that the temporal variation of the principal component provides sufficient information for retrieving the position of a PLW within its gait cycle (i.e., phase) (Das, Lazarewicz, & Finkel, 2004).

These signals could be processed in global motion sensitive units within the brain. These are larger neuronal receptive fields that do not require complete spatiotemporal coherent information. They are located in the medial superior temporal area and the middle temporal area. These receptive fields integrate small spatiotemporal asynchronies and then code the percept as a coherent whole (Y. Chen, Nakayama, Levy, Matthysse, & Holzman, 2003). Also in a synchronization task with a random dot paradigm, it was demonstrated that only up to 50% spatiotemporal coherence (i.e., co-variation of the motion of dots) was required for individuals to adapt to tempo changes (Ceux, Wagemans, Rosas, Montagne, & Buekers, 2006).

Thus, many cues provide information about the global motion of the walking partner from which the instantaneous position of the gait cycle could be retrieved. We speculated that this perceived (sinusoidally accelerating and decelerating) signal of the global motion of the partner could be matched with the global motion of the individual to achieve synchronization. The investigation of this question was the primary goal of Experiment 4. Its mechanisms are further discussed in Section 0.6.

Moreover, we assessed whether synchronization differed depending on the type of sensory channel that is used. It is reasonable to expect that individuals match heel strikes when these are provided aurally and being the only available cues from the walking partner. However, when there are only visual cues, heel strike synchronization should be reduced, and global body synchronization could be elevated. Finally, when audiovisual cues are available, the auditory cues could disambiguate the visual perception of heel strikes onsets. This also would make the heel strike onsets more salient. Therefore, we expected an increased heel strike synchronization when both cues are available. Another reason for this might be the facilitation through an unimodal matching of auditory signals from one's movements and the movements of the partner, as discussed previously and indicated by the results of Experiment 1 to 3.

Another critical shortcoming of Experiment 1 to 3 is that the participants synchronized with non-adaptive stimuli. This is similar to SMS tasks in which movements have to be synchronized with metronome events. It allows deriving conclusions about mechanisms underlying human timing capabilities but it differs from interpersonal coordination processes in real life. Due to an uni-directional information exchange, the walking partner cannot compensate for movement errors. This could alter underlying synchronization mechanisms. Thus, to scrutinize the above-mentioned arguments, we conducted Experiment 4 and analyzed movement synchrony of different body segments and the global motion of human participants walking side-by-side.

0.5 Experiment 4

0.5.1 Methods

Participants

Sixteen participants (12 naïve, nine female, seven male, age: $M = 28$, $SD = 3$) without gait disabilities took part in the experiment. They were combined into eight pairs by matching hip height (the maximal difference was 2cm) and gender. There was one exception because a non-naïve male was paired with a female participant. All individuals gave informed consent for their participation.

Material

Six reflective markers were attached to one lateral side of the head and the position of the principal joints of the shoulder, elbow, wrist, knee, and ankle. Marker positions were captured at 240Hz by the Vicon motion capture system with 16 near-infrared cameras (MX F20 of 2 megapixels), defined in a XYZ-Euclidian frame. The stimuli were the visual cues and auditory cues provided by the partner walking at daylight. The sound amplitude was amplified by attaching a plastic film to the foot sole (48dB LAeq). Visual cues were constrained by side-blinders. This restricted the visual field to an area of approximately 60 degrees, 30 degrees to the left and 30 degrees to the right, relative to the heading direction. Auditory cues were hampered by in-earplugs and external hearing protections, leading to the prevention of sounds inferior to 55dB LAeq.

Procedure

The experiment was conducted in a large indoor field. The participants wore sports shorts and walked barefoot (with attached plastic film) on a 20m walkway side-by-side. In a bi-directional coupling condition, participants could a) see and hear (Audiovisual condition - AV), b) see (Visual condition - V), or c) listen to each other (Auditory condition - A). In a unidirectional coupling condition, for one participant, the same manipulations were carried out while the other participant should neither be able to hear nor to see the walking partner. Thus, the 3 (Sensory Information) x 2 (Coupling) conditions made up to six conditions. Because in the unidirectional coupling condition both participants acted as coupled and non-coupled individual, each pair walked overall nine times. The order was pseudo-randomized. The participants started walking when an auditory signal was presented at random starting onsets. The following instructions were given: “Walk constant, comfortable, not too slow and not too fast. Walk on the side of your walking partner. Synchronize as best as possible with your partner. You can turn your head if you want/need to”. When vision was prevented, it was added: “Focus heading direction. Use auditory cues to synchronize. If the partner appears in your field of view, accelerate slowly so that you cannot see your partner anymore”. When auditory information was prevented, it was added: “Use visual cues to synchronize”.

The participant who supposedly did not receive cues from the partner was instructed to “walk as constant as possible with preferred tempo”.

0.5.2 Analysis

Data were analyzed approximately for 10m after the participants had started walking for 5m. Due to poor capture quality at start and end, the interpreted walking distance differed. However, within participant pairs, it was similar throughout the conditions. The motion capture of the marker displacements provided the spatiotemporal variations of 6 segments (i.e., head, shoulder, elbow, hand, knee, & ankle).

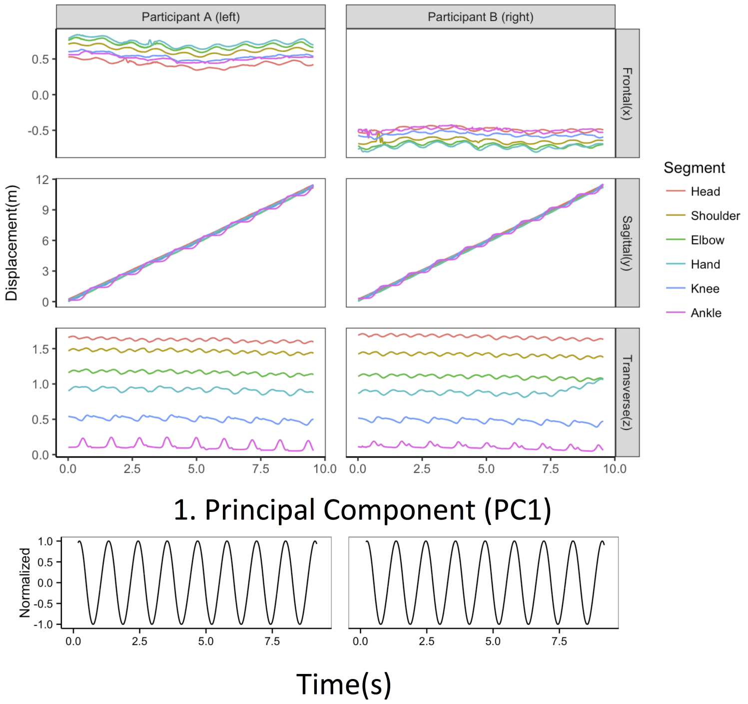
Principal Components

The principal scope was to determine whether individuals synchronize global body motion rather than individual segments. Comparing synchrony measures of the displacements of the joints and the global motion could answer this question. As previously, synchrony performance was quantified by the s.d. of asynchronies from discrete events of the signals. We assumed that synchrony would be highest for those variables that are considered as most relevant by the CNS. We hypothesized that during synchronization of side-by-side walkers, these are the global motions of the walkers. Next, it is briefly illustrated how the global motion signal was obtained.

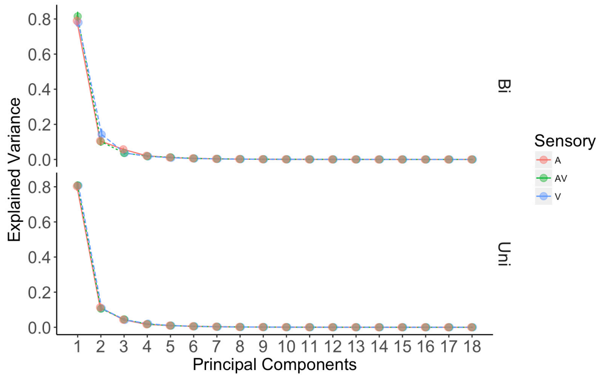
The motion trajectories of the several segments are correlated in sagittal, frontal, and transverse plane directions. Most segments move simultaneously, or with some temporal lag but consistently, into similar directions. Our premise was that the principal dimension of co-variation of the different segments represents the global body motion of each walker. The principal dimension of co-variation could be identified by PCA (see e.g., Troje, 2008) (see Figure 21a for an example of the obtained time series). The PCA transforms the set of correlated segment motions into uncorrelated dimensions, called components. When the motions are correlated, fewer components are necessary to capture the entire motion variability. As shown in former studies, during walking there is one single dimension, called the principal dimension (PC1), that captures most of the variability. We implemented a PCA and determined whether the PC1 could be interpreted as the global motion of the walker. For a description of the PCA algorithm, see A.8 and Troje (2008).

The PCA revealed a dimension of covariation that explained approximately 80% of the variation of the segment motion. This pattern was similar across the sensory conditions and the coupling conditions (see Figure 21b). It demonstrates that the individual body segment motions are highly correlated and can be captured by one single dimension. Therefore, the temporal displacements of the coordinates projected onto this first principal component (PC1) were considered the global motion components of the walkers (see e.g., Troje, 2008).

Next, we computed asynchronies between the different segment motions and between the global motions of the two individuals (see Figure 22). Because most signals were non-sinusoidal (i.e., the phase increased non-linearly), we considered the discrete events within the gait cycle (i.e., local minima in the time series of each signal) (see Peters et al., 2003). Previous studies indicated that individuals might spontaneously start synchronizing in in-phase or antiphase and that walkers can switch between both of these stable states (Haken et al., 1985; van Ulzen et al., 2008, 2010). However, we did not observe any flips between these two phases. Besides being variable, the apparent stable phase difference remained the same throughout a trial. Due to the absence of phase flips and considering that we did not apply any predetermined cut-off criterion—as it was done in Experiment 1, 2 and 3—, there was no need to transform the asynchronies into DRPs (see also Wheat & Glazier, 2005). As previously, we treated in-phase synchronization and antiphase synchronization as equal.



(a) Time series of individual body segments and principal motion component (PC1). Top: Times series of each marker of one pair in a trial. The displacement is in sagittal (y - walking direction), frontal (x - perpendicular to walking direction), and transverse (z - vertical) motion direction. Bottom: Displacement of all coordinates projected onto the PC1. Afterwards, the signal was normalized. Note that the translational component (sagittal) was removed before applying PCA.



(b) The mean proportions and 95% confidence intervals of explained variance of the joint displacement by the first principal component (PC1), separated by sensory and coupling condition. On the ordinate are displayed the principle components (6 segments x 3 dimensions = 18), ordered by how much variance they explain in the data. The sensory conditions are color-coded (red=V, green=A, blue=AV). The variability that is explained by the principle components is very similar for all sensory conditions and coupling conditions. Due to a superposition (i.e., similar results), the different sensory conditions are not easily distinguishable. The plots indicate that the PC1 captured consistently most variability of the segment motion in any of these conditions.

Figure 21: Output of Principal Component Analysis (PCA) of the motions of the segments of both participants.

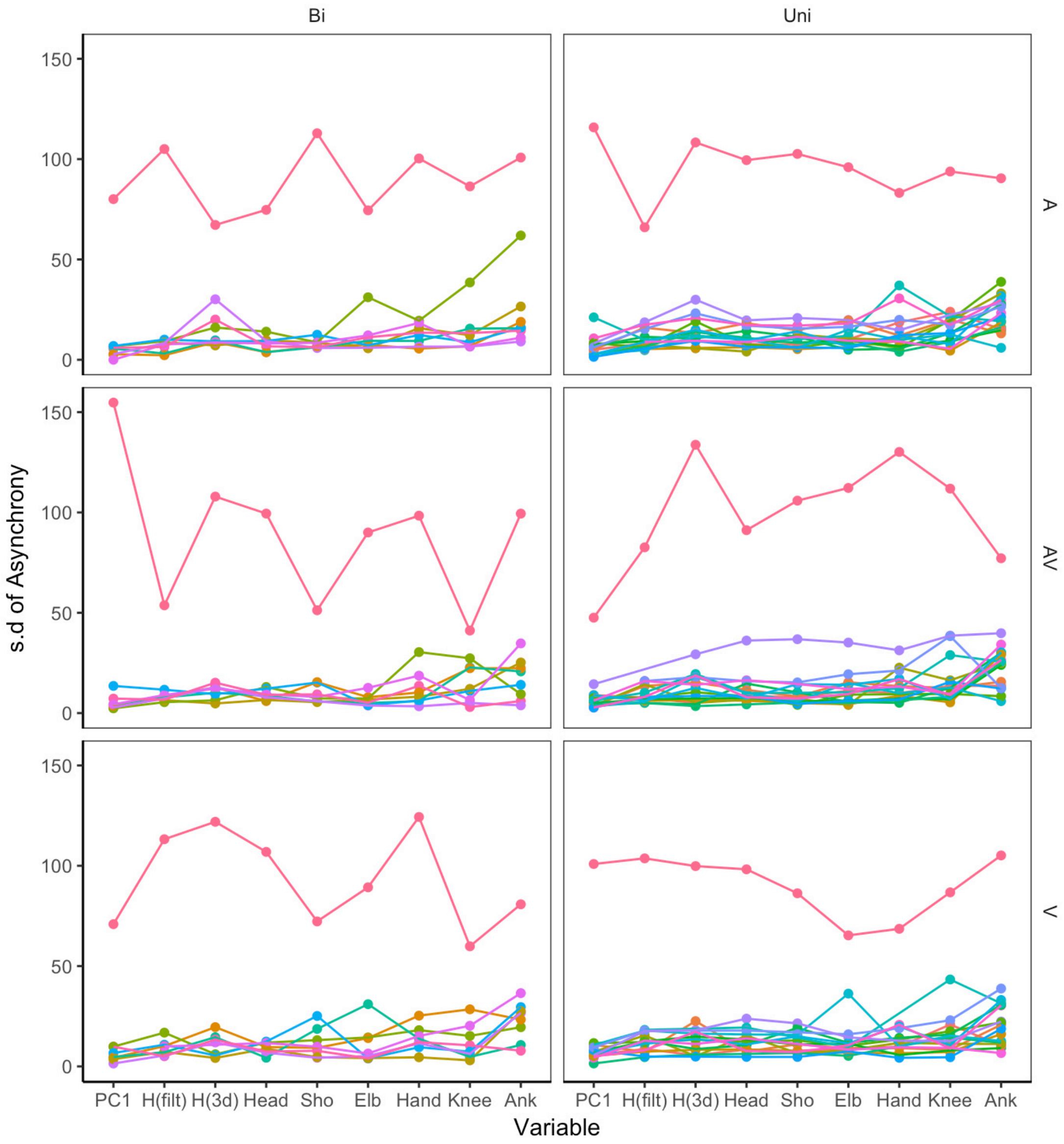


Figure 22: S.d. of asynchronies (synchronization performance) as a function of the pair of comparison: segment-segment and the first Principal Component (PC1-PC1), separated by sensory condition and coupling direction condition. Dyads (left n=8, right n=16) are color-coded. The s.d. of asynchronies was computed from the vertical (1-dimensional) displacement of each variable. In red are displayed the median s.d. of asynchronies computed from the displacement of 20 randomly paired individuals, serving as control condition. In H(filtered), the signal of the head was filtered by a moving average of order 5. In H(3d), all three dimensions (i.e., X, Y, & Z) were combined for estimating the asynchronies. Note that the largest s.d. of asynchronies can be attributed to some parts to an increased uncertainty of estimating the exact minima of the ankle displacement. We did not meet further difficulties of estimating the minima of other variables.

0.5.3 Results

The main goal of Experiment 4 was to provide hints which motion variable individuals might control when they are instructed to synchronize walking. A higher control should result in improved synchronization performance. Synchronization performance was quantified by the s.d. of asynchronies. For computing asynchronies, the movement time series of each trial were divided into several intervals in which local minima could be identified.

We assumed that the more variable asynchronies (higher s.d. of asynchronies) are the dimensions of coordination that are less central to CNS control (Schöner, 2016). Therefore, we expected the lowest s.d. of asynchronies when computed from the global motion signal. We expected that the s.d. is lower in AV and A compared to V when computed from the heel strikes. Also, we expected the s.d. to be lower in the bidirectional coupling condition compared to the unidirectional coupling condition.

The asynchronies were computed from the minima of the time series of equal segments (asynchrony PC1 = PC1 of participant 1 - PC1 of participant 2, asynchrony head = head of participant 1 - head of participant 2, asynchrony ankle = ankle of participant 1 - ankle of participant 2, etc.). See Table 2 for all relevant variable pairs. We determined whether s.d. of asynchronies differed across the experimental conditions (i.e., segment of interest, available sensory information, coupling direction). The observations did not meet requirements for parametric testing (i.e., non-normal distributed s.d. of asynchronies, heteroscedasticity, skewed distributions, & outliers). A robust technique for hypothesis testing is here the non-parametric Bootstrap because it does not make any assumptions about the underlying probability distribution that might have produced the data. CIs of 95% were constructed for each experimental condition. Because the present study used a repeated measurement design, the Bootstrap samples were drawn from the differences of the s.d. of asynchronies across experimental conditions. A CI including zero would then mean that there is no statistical evidence for a difference between the compared conditions. See Table 2, Table 3, Table 4 and Table 5 for explanations about the relevant combinations that were compared.

Table 2: Asynchronies computed from different segment pairings. Asynchronies can be computed from the minima of all variable pairings of the participants (i.e., all combinations of the segments and the PC1). The first variable is from the participant walking on the right side and the second variable is from the participant walking on the left side. We presented here statistical analysis of the s.d of asynchronies that were computed from equal variables (PC1-PC1, Ank-Ank, etc.).

	PC1	H(filt)	H(3d)	Head	Sho	Elb	Hand	Knee	Ank
PC1	PC1-PC1	.	.	.	.	.	.	.	.
H(filt)	.	H(filt)-H(filt)	.	.	.	.	.	.	.
H(3d)	.	.	H(3d)-H(3d)	.	.	.	.	.	.
Head	.	.	.	Head-Head	.	.	.	.	.
Sho	.	.	.	.	Sho-Sho	.	.	.	.
Elb	.	.	.	.	.	Elb-Elb	.	.	.
Hand	.	.	.	.	.	.	Hand-Hand	.	.
Knee	.	.	.	.	.	.	.	Knee-Knee	.
Ank	.	.	.	.	.	.	.	.	Ank-Ank

Table 3: Bootstrapped differences between variables (segments and PC1). We were interested if s.d. of asynchronies differed between the segments and the PC1. Therefore, the s.d. of asynchronies of any variable pair could be subtracted from the s.d. of asynchronies of any other variable pair. Of main interest was the difference between the s.d. of asynchronies of the pair PC1-PC1 with any other equal body segment pair. In parentheses are presented the combinations of the variable-pairings of interest. From these were computed the s.d. of asynchronies. In order to test if the s.d. of asynchronies was less for the pair (PC1-PC1) compared to the other pairs (Head-Head, Sho-Sho, etc.), the differences of the s.d. of asynchronies was computed and a Bootstrapped 95% nonparametric confidence interval was determined. The pairings for which the differences were not Bootstrapped are indicated by dots. The 95% confidence intervals were Bootstrapped within each combination of the sensory conditions and the coupling direction conditions (i.e., AV-Bi, AV-Uni, A-Bi, A-Uni, V-Bi, V-Uni).

	sd(PC1-PC1)	sd(H(filt) -H(filt))	sd(H(3d) -H(3d))	sd(Head -Head)	sd(Sho -Sho)	sd(Elb -Elb)	sd(Hand -Hand)	sd(Knee -Knee)	sd(Ank -Ank)
sd(PC1-PC1)	.	.	.	.	.	.	.	.	.
sd(H(filt)-H(filt))	sd(PC1-PC1) -sd(H(filt)-H(filt))	.	.	.	.	.	.	.	.
sd(H(3d)-H(3d))	sd(PC1-PC1) -sd(H(3d)-H(3d))	.	.	.	.	.	.	.	.
sd(Head-Head)	sd(PC1-PC1) -sd(Head-Head)	.	.	.	.	.	.	.	.
sd(Sho-Sho)	sd(PC1-PC1) -sd(Sho-Sho)	.	.	.	.	.	.	.	.
sd(Elb-Elb)	sd(PC1-PC1) -sd(Elb-Elb)	.	.	.	.	.	.	.	.
sd(Hand-Hand)	sd(PC1-PC1) -sd(Hand-Hand)	.	.	.	.	.	.	.	.
sd(Knee-Knee)	sd(PC1-PC1) -sd(Knee-Knee)	.	.	.	.	.	.	.	.
sd(Ank-Ank)	sd(PC1-PC1) -sd(Ank-Ank)	.	.	.	.	.	.	.	.

Table 4: Bootstrapped differences between the sensory conditions. Here we were interested whether the s.d. of asynchronies differed across sensory conditions. Therefore, the s.d. of asynchronies was computed for equal variable pairs. The differences of s.d. of asynchronies were computed for the same variable pairs (sd(PC1-PC1) in AV - sd(PC1-PC1) in A; sd(Head-Head) in AV - sd(Head-Head) in A, etc.). From these differences, it were Bootstrapped 95% nonparametric confidence intervals. This was done within each combination of the equal variable pairs and the coupling condition (i.e., Head-Head-Bi, Head-Head-Uni, Sho-Sho-Bi, Sho-Sho-Uni, etc.)

	A	AV	V
A	.	sd(PC1-PC1) - sd(PC1-PC1)	sd(PC1-PC1) - sd(PC1-PC1)
AV	sd(PC1-PC1) - sd(PC1-PC1)	.	sd(PC1-PC1) - sd(PC1-PC1)
V	.	.	.

Table 5: Bootstrapped differences between the coupling direction conditions. Here we were interested whether the s.d. of asynchronies differed between coupling direction conditions. Therefore, the s.d. of asynchronies was computed for equal variable pairs. The differences of s.d. of asynchronies were computed for the same variable pairs (sd(PC1-PC1) in Bi - sd(PC1-PC1) in Uni; sd(Head-Head) in Bi - sd(Head-Head) in Uni, etc.). From these differences, it were Bootstrapped 95% non-parametric confidence intervals. This was done within each combination of the equal variable pairs and the coupling condition (i.e., Head-Head-AV, Head-Head-A, Head-Head-V, Ank-Ank-AV, etc.)

	Uni	Bi
Uni	.	sd(PC1-PC1) - sd(PC1-PC1)
Bi	.	.

Summarizing Table 2, Table 3, Table 4 and Table 5, the bootstrap functions received the difference between s.d. of asynchronies of each dyad 1) of the segment combinations (e.g., $\sigma_{PC1_1-PC1_2} - \sigma_{Elbow_1-Elbow_2}, \sigma_{PC1_1-PC1_2} - \sigma_{Ankle_1-Ankle_2}$, etc.), within each sensory and coupling condition (Figure 23), 2) of the sensory combinations ($\sigma_{PC1_1-PC1_2}$ in AV - $\sigma_{PC1_1-PC1_2}$ in A, $\sigma_{PC1_1-PC1_2}$ in AV - $\sigma_{PC1_1-PC1_2}$ in V, $\sigma_{Head_1-Head_2}$ in A - $\sigma_{Head_1-Head_2}$ in V), within each segment and coupling condition (Figure 24a), and 3) of the coupling combinations ($\sigma_{PC1_1-PC1_2}$ in Uni - $\sigma_{PC1_1-PC1_2}$ in Bi, $\sigma_{Head_1-Head_2}$ in Uni - $\sigma_{Head_1-Head_2}$ in Bi, etc.), within each segment and sensory condition (Figure 24). The σ represents the s.d. of asynchronies and the subscripts $_1$ and $_2$ represent participant 1 and participant 2. The 95% CI of each difference value was obtained from 1000 Bootstrap samples.

First, we expected that s.d. of asynchronies computed from the vertical displacement of the segments is larger than when computed from the PC1s. Results revealed that the CIs did not include zero (see Figure 23). This means that s.d. of asynchronies was lowest for PC1. An exception was the elbow in Bi-AV, where the s.d. of asynchronies did not significantly differ from PC1. The s.d. of asynchronies between other segments was consistently greater than the PC1.

A possible limitation is that PCA filters out the noise from the signals. This filtering function reduces variability. To control for this, we applied a moving average filter of order 5 to each segment time series and then estimated the local minima from the filtered signal (see H[filt] for the filtered head signal). Moreover, because the PC1 was computed from 3 coordinates, we also estimated the asynchronies from local minima of the 3-dimensional signals (see H[3d]) (see Figure 23 and Figure 24a). S.d. of asynchronies was similar when computed from the filtered, the 3-dimensional, and the non-filtered 1-dimensional signals.

Second, we expected that the s.d. of asynchronies may be higher in V followed by A and then by AV. The factor Sensory (with 3 levels, AV, A, V) was compared within each condition (e.g., AV-A for [PC1, Bi], [PC1, Uni], [Head, Bi], etc.). Sensory differed in Bi. S.d. of asynchronies was smaller in AV than in A for Head(3d), Elbow, and Ankle. It was smaller in AV than in V for Shoulder, Elbow, and Ankle. It was larger in A than in V for Head(3d) and Knee, and smaller for Shoulder (see Figure 24a). For

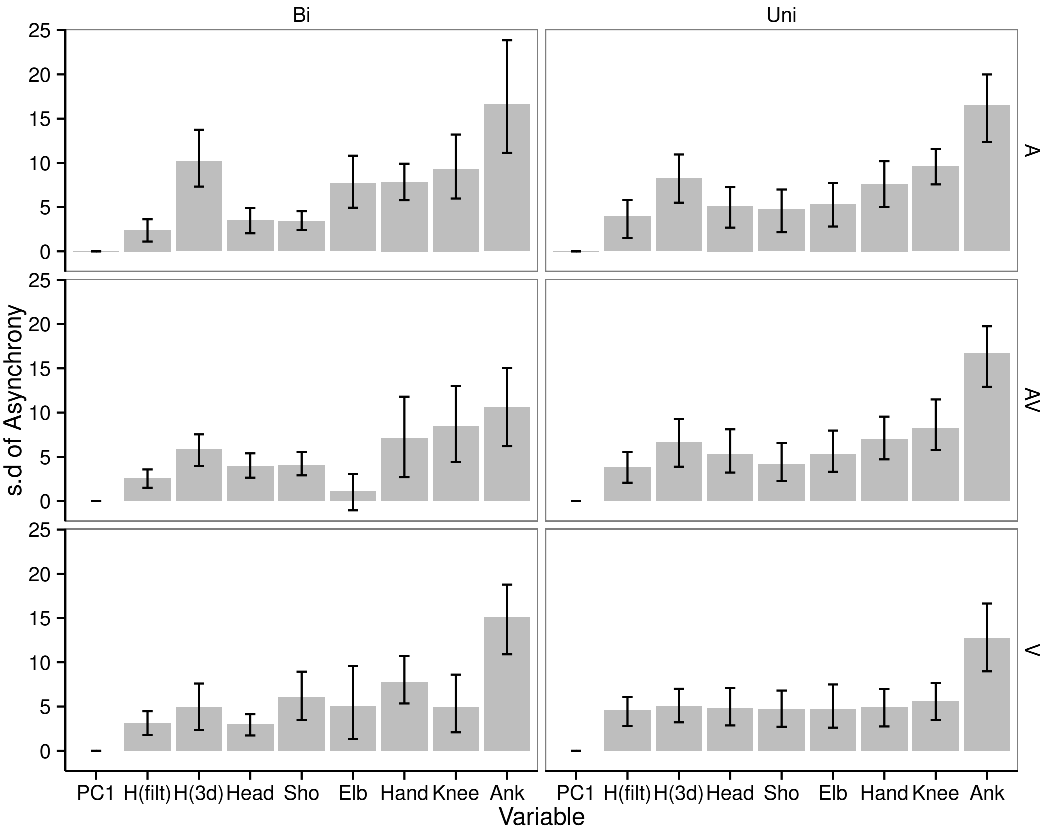


Figure 23: Synchronization performance differences between PC1 and the segments. Positive values indicate that the s.d. of asynchronies is lower for PC1 compared to the other segments. Note that at the position labeled “PC1”, the barplot indicates zero mean and then absence of confidence intervals since at this point, the s.d. of asynchronies of PC1 was subtracted from the s.d. of asynchronies of PC1, thus, the signal was subtracted from itself.

Uni, s.d. of asynchronies was smaller in V than in AV and A for the Ankle.

Third, we expected that the s.d. of asynchronies may be higher in Uni than in Bi. The differences between Uni and Bi were compared within each condition (e.g., Bi-Uni for [PC1, AV], [PC1, A], [PC1, V], [Head, AV], etc.). The CIs included zero for all conditions (see Figure 24).

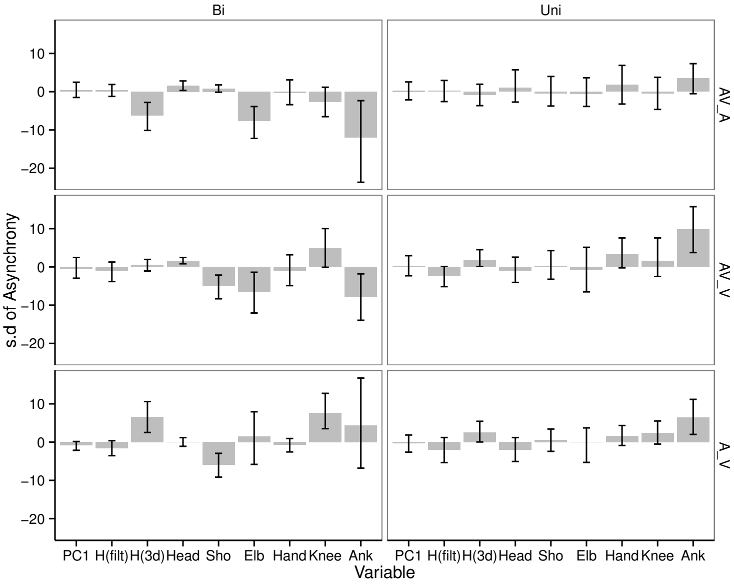
In summary, consistent with our expectations, results revealed that the s.d. of asynchronies between segment pairs was greater than between the PC1 pair. Inconsistent with our expectations, there was no clear pattern of differences across the sensory conditions and there were no differences between the coupling conditions.

0.5.4 Discussion

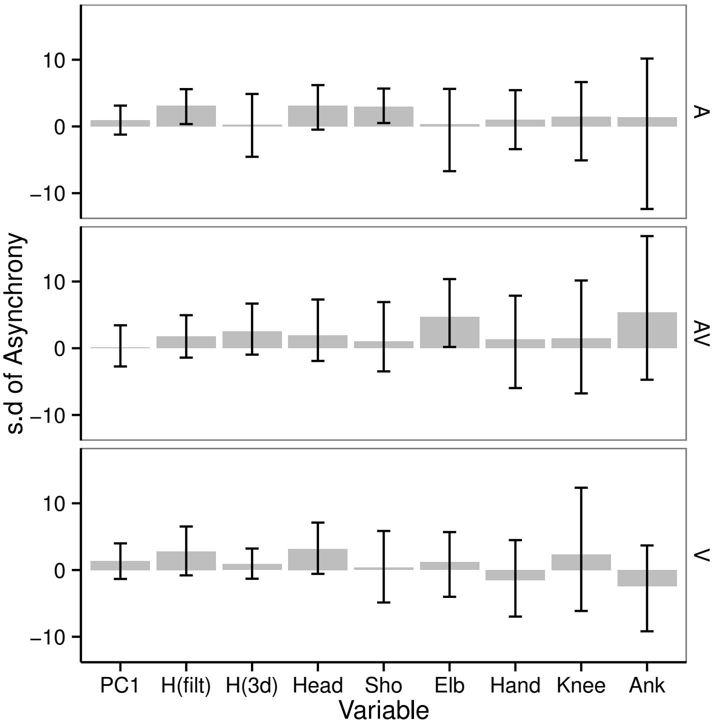
In Experiment 4, dyads synchronized movements during side-by-side walking. Synchrony was quantified by the s.d. of asynchronies between the onsets of discrete time points within the movement cycles of each individual. The s.d. of asynchronies was remarkably lower compared to the s.d. of asynchronies in Experiment 1 and 2 (compare Figure 22 with Figure 14). We attribute this to three factors. First, in Experiment 4, the s.d. of asynchronies was calculated from one single walk. In the previous experiments, an asynchrony was obtained from the last step in each trial and the s.d. was then computed from asynchronies of multiple trials. Second, in half of the trials in Experiment 4, participants were bi-directionally coupled. This probably reduced the s.d. of asynchronies. Third, the virtual environment of Experiment 1 to 3 had delays between motion capture and stimulus presentation.This probably increased the s.d. of asynchronies in Experiment 1 to 3.

In Experiment 1 to 3, we focused on step synchronization. The main goal of Experiment 4 was to provide hints which motion variable individuals might control when they are instructed to synchronize walking. We presumed that the variability of asynchronies of the segments—computed between the local minima of the time series signals—reveals these relevant variables.

We speculated that individuals achieve synchronization by temporally matching (controlling) the global motion of the bodies rather than focusing on individual segments. The global motion can be represented by the principal co-variation of all body



(a) Difference across the sensory conditions. Positive values indicate that the s.d. of asynchronies in the first mentioned sensory condition is larger.



(b) Difference between the coupling direction conditions. Positive values indicate that the s.d. of asynchronies in the Uni-directional condition is larger.

Figure 24: Mean and 95% non-parametric Bootstrap confidence intervals of the difference between conditions obtained from 1000 Bootstrapped replicates of the sample.

segments. This was here captured by PCA. Consistent with our speculations, synchronization variability was lowest when the asynchronies were computed from the first Principal Component (PC1). This indicates that individuals could synchronize global motion during side-by-side walking. An exception was the elbow segment, which did not significantly differ from the PC1 in the bidirectional audiovisual condition. This might simply be a type I error that was likely to occur due to the multiple comparisons.

As a second goal, we were interested whether movement synchronization is different when individuals are unidirectionally and bidirectionally coupled. Against our expectations, synchronization in the unidirectional coupling condition was not more variable than in the bidirectional coupling condition. However, the absence of clear differences seems contra-intuitive taking into account that in the bidirectional coupling condition both individuals could compensate for the variability produced by the other individual. We attribute this lack to the fact that in the unidirectional coupling condition information exchange was not entirely restricted to one direction. That is, in the first three experiments, we created the stimuli in a virtual environment. For this reason, in A, the only available cues from the walking partner were actually auditory and in V they were actually visual.

In contrast, in Experiment 4, although we attempted to prevent auditory or visual information in the respective sensory condition, this might have been insufficient and the manipulations did not prevent other sensory cues from being perceived by the walking partner entirely. On the one hand, it was controlled for head rotation and the distance between the individuals. This revealed that the participants apparently did not rotate the head and did not increase the distances between each other. Consistent with these results, participants reported that the partner did not appear in the visual field of view in conditions in which this should have been prevented. It indicates that at least the visual cues were manipulated appropriately.

On the other hand, in the visual condition, sounds produced by the heel strikes might still have been heard. The heel strike sounds were occluded by in-earplugs and external hearing protections. In-earplugs prevent that external sounds are perceived up to 30 dB LAeq. Hearing protections prevent a perception of sounds up to 25 dBLAeq. Therefore, all external sounds that were inferior to 55 dBLAeq should be occluded.

However, this occlusion technique was not very reliable and depended highly on the individual use. Moreover, while we made a quick sound perception test before the start of the experiment in order to validate whether the occlusion device was applied appropriately, such technique can loose effectiveness with the time of use. Furthermore, vibrations from heel strikes on the surface, air pressure due to body displacement, and other variables might have exist that were not controlled here. Particularly, the surface vibrations can also be sensed through, haptic, proprioceptive, auditory and the vestibular system, providing relevant cues for synchronization (see Section 0.1.5).

An insufficient sensory manipulation would also explain the absence of a clear difference between the sensory conditions. It is not clarified how global body synchronization emerges when only the heel strike sounds are available. But, there were probably further auditory cues like the friction of clothes, breathing, etc. Alternatively, this could suggest that when the sensory channels are constraint, a global model might still be used and estimated from the reduced set of cues. For a more detailed discussion, see Section 0.6.3.

Nonetheless, there was a complex interaction between the coupling direction and the type of sensory manipulation. In contrast to the unidirectional coupling condition, in the bidirectional coupling condition, there were differences in the variability between several segments depending on the sensory condition. It was lower for the elbow, head, and ankle when additional visual cues were provided and for the shoulder, elbow, and ankle when additional auditory cues were provided. For the unidirectional coupling condition, the variability was greater for the ankle in the audiovisual and the auditory condition than in the visual condition. Although this pattern is complex and challenging to explain and might, to some extent, be a result of such insufficient sensory manipulations, it overall indicates that bidirectionally coupled individuals benefit more from the presence of audiovisual cues than unidirectionally coupled individuals.

In short, the above-mentioned shortcomings of the sensory manipulations may have resulted in more variability and could partly explain the intricate pattern of differences. Nevertheless, assuming that the here constructed lower-dimensional variable (PC1) captured the global motion of the walking partners, the results point towards that the global motion of the individuals became synchronized and that this was achieved by using multiple (combined) available sensory cues.

0.6 General Discussion

In this dissertation, we investigated the contribution of sensory cues for the synchronization of the movements of side-by-side walkers. Because testing such assumption is complicated, in this first attempt, we experimentally approached the audiovisual cues from the walking partner and studied intentional synchronization. In order to achieve high control over the stimuli while maintaining the situation as natural as possible, three experiments were conducted using virtually simulated stimuli that represent the partner. With a fourth experiment, we subsequently scrutinized the possible shortcomings of the first three experiments and also investigated the relation between local versus global body synchronization.

0.6.1 Audiovisual Cues Increase Step Synchronization

Experiment 1 revealed highest synchronization performance when auditory cues were present, regardless of the visual ones. This auditory dominance effect was attributed to experimental artifacts. Subsequently, through Experiment 2 (with an improved setup), we provided evidence that the presence of audiovisual cues increased synchronization performance compared to auditory or visual cues alone.

In order to synchronize steps, an individual has to estimate the time of the upcoming heel strike onsets. This can be based on multiple cues. But here, we focused on the integration of heel strike sounds and visual cues of the motion of the virtual walking partner. An optimal way of estimation is to integrate signals from both modalities. This shall lead to estimates with increased reliability (Ernst & Bülthoff, 2004). Increasing the reliability of the estimate should then lead to more stable synchronization (reduced s.d. of asynchronies). In Experiment 2, the MLE model predicted partly the synchronization variability for audiovisual cues. However, the task required that the participant cross-modally matched audiovisual cues from the partner with audiovisual, somatosensory, and kinesthetic cues from one's own movements. Up to now, neither the MLE model nor any other currently available model seems appropriate to describe the behavior in such a cross-modal sensorimotor matching paradigm. A possible experimental approach would be to manipulate all of these cues within a cross-modal framework (Elliott et al., 2010).

Multimodal signals should be integrated when the signalized unimodal event onsets fall into a window of temporal integration (WTI) (Vroomen & Keetels, 2010). Hence, in Experiment 3, it was tested for bimodal integration by manipulating the asynchrony between the stimulus onsets. At small asynchronies, the variability was as small as in the synchronous condition. This is consistent with the MLE model because the integration of slightly asynchronous signals should still reduce the variability of the final estimate of the onsets (Ernst & Bülthoff, 2004). At larger asynchronies, the synchronization variability increased. This could be attributed to distractor effects of the temporally displaced stimuli. These arise when signals from supposedly separate (or independent) events are integrated to code the same event (e.g., Repp & Penel, 2004). When the asynchrony is very large, this does not happen. Then, asynchronous cross-model events are coded and perceived as separate (Keetels & Jean, 2012). Therefore, the point where distraction is maximal might indicate the size of the WTI. However, results of Experiment 3 weaken this assumption because the points of maximal distraction were much larger than expected from the WTIs determined by previous studies (see e.g., Mendonça et al., 2011; Vatakis & Spence, 2006). Moreover, a decreasing variability at larger shifts was only observed when the auditory stimulus was shifted but not when the visual stimulus was shifted.

Yet, one should be cautious when drawing conclusions about the exact sensory integration mechanism. The size of the WTI differed largely between previous studies being much larger and different for complex biological stimuli (see e.g., Repp & Penel, 2002, 2004; Vatakis & Spence, 2006; Arrighi, Alais, & Burr, 2006). Furthermore, in the present experiments, individuals synchronized with a non-adaptive stimulus. This is similar to SMS studies in which movements had to be synchronized with metronome events (e.g., Repp, 2005), but it differs from interpersonal coordination studies where both individuals were mutually adaptive (e.g., Schmidt & Richardson, 2008). The underlying mechanism of both tasks might be different and for complex stimuli and tasks cues that are incongruent might still be combined and are beneficial in other ways.

It must be highlighted that Experiment 2 and Experiment 3 used individuals that had also participated in Experiment 1. The pilot tests do not entirely rule out that participants might have improved their performance in subsequent experiments due to training. However, since this was a very novel and complex paradigm, we used the same individuals in order to have some comparison standard across the experiments. Because the sample sizes were very small, using different participants across the experiments would probably have hampered the interpretations to a greater extent. Figure 15 provides some hints for the absence of large training effects. Nevertheless, previous walking studies investigating spontaneous synchronization have shown that there is a large amount of variability among pairings (Zivotofsky et al., 2012). Although we assume that variability across pairings may be reduced for intentional synchronization, the small sample size limits the generalizability of our results.

It is also unclear if these findings can be generalized to spontaneous synchronization. Specifically, we do not know if the same mechanisms are shared by both the spontaneous and the intentional synchronization. Nevertheless, our results are generally consistent with those from Nessler and Gilliland (2009), where spontaneous synchronization was greatest with audiovisual cues followed by auditory cues and then visual cues. Furthermore, several former studies on spontaneous synchronization reported an auditory dominance effect (see e.g., Repp, 2005), which was also found in our study. Besides one participant in Experiment 2, synchronization was consistently faster and less variable with auditory cues and the distraction effect was stronger with auditory shifts than with visual shifts. Therefore, the underlying sensory integration mechanisms in intentional and spontaneous synchronization might be comparable.

0.6.2 Unimodal Matching Increases Step Synchronization

Rather than assuming a superior temporal processing for the above mentioned auditory biases (see e.g., Repp, 2005; Van Wassenhove, Grant, & Poeppel, 2007), we here suggest that the auditory bias might be explained by an increased ease of matching unimodal stimuli. For asynchrony estimation, all available information can be used. We assume that during walking without obstacles, people do usually not observe their feet. Then, kinesthetic and somatosensory signals of one's body and sounds produced through ground contact signalize the heel strike onsets. It was shown that temporal estimates are much more accurate when evaluating stimuli with a sharper rise time of energy, which were here both heel strike sounds (Van der Burg, Cass, Olivers, Theeuwes, & Alais, 2009). It is widely accepted that when the reliability to estimate events by individual cues increases, the integration of multiple cues decreases. This is called Principle of Inverse Effectiveness, as mentioned in the introduction (Holmes, 2009) (see Section 0.1.6). This principle is partly captured by the MLE model (see Section 0.1.4). The higher is the reliability of estimating the event by an individual cue—i.e., the lower is the SD of the assumed underlying normal distribution—the less other cues with lower reliability contribute to the final estimate. Therefore, when an unambiguous perception is established based on information from a single cue, there is no need to integrate further less reliable cues. Moreover, fusion limits at which two cues are perceived as one are much lower for multimodal stimuli (∼4Hz for audiovisual) (Fujisaki & Nishida, 2009) compared to unimodal stimuli (∼25Hz for visual) (Fujisaki & Nishida, 2005). In addition, in the present setup, the online generation of auditory stimulus position and visual perspective was delayed leading to spatiotemporal incongruencies. This could have further promoted the reliance on single modalities. Thus, assuming that with the auditory cues alone, considerably high synchronization performance can be achieved, it seems plausible that the individuals of Experiment 1 to 3 matched externally the auditory signals rather than cross-modally matching auditory, somatosensory, and kinesthetic signals with external visual signals. Support for this is provided by a study showing that tempo matching during stepping on place was facilitated through unimodal matching sounds with auditory feedback of the heel strikes compared to cross-modally matching sounds with haptic feedback (Maculewicz et al., 2016).

In conclusion, first, when instructing individuals to synchronize movements during side-by-side walking, they seem to benefit from the integration of audiovisual cues. There is currently no appropriate model but the MLE model captures by parts these observations. Second, auditory cues seem to have a stronger impact on synchronization performance because they seem to reliably represent one's own and the walking partner's heel strike onsets, which could then be matched unimodally.

Future studies could systematically manipulate the reliability of the involved sensory systems that inform about one's body movements. This would provide a more comprehensive picture of the sensory integration mechanisms. Visual cues like optic flow could be manipulated by virtual environment techniques and by comparing treadmill walking with overground walking (see e.g., Durgin, Reed, & Tigue, 2007). Somatosensory cues could be blurred by the use of vibration techniques. For haptic perception, vibrations could be applied to the foot sole at heel strike (Giordano et al., 2012). For proprioception, vibrations could be directly applied to the muscles (Bove, Courtine, & Schieppati, 2002). There have also been attempts to manipulate vestibular cues by caloric and galvanic stimulation. It was shown that these techniques evoke self-motion perception mediated by vestibular signals (St George & Fitzpatrick, 2011; Fasold et al., 2002; Preuss, Hasler, & Mast, 2014; Fitzpatrick et al., 1999). Yet, it also partly affected visual and somatosensory processes and posture control (Lopez, 2016).

Overall, it is challenging to manipulate single sensory systems in isolation because most function in connection with multiple systems and therefore have to be studied in a multidisciplinary way. Nevertheless, a possibility would be to use patients with deficits or brain lesions within the related sensory system. For instances, labyrinthectomy removes the end organs so that signals from the inner ear do not reach the brain. Other techniques could be the manipulation of functions through medication. Due to recalibration (Durgin et al., 2005), the manipulation of a single sensory system might be possible without disturbing the gait pattern if sufficient training is provided. So, although it is hard, some manipulations might provide insight into the multimodal contribution of each sensory system that should be approached in future studies.

0.6.3 Synchronization is Achieved by Matching Global Body Motion

For technical reasons, in Experiment 1 to 3, we quantified synchronization in walking by the variability of asynchronies computed from the minima of the heel strike onsets. By focusing on lower limb movements, we followed the convention of most studies investigating synchronization during walking (Murray-Smith et al., 2007; Nessler & Gilliland, 2009; Nessler et al., 2011; Nessler & Gilliland, 2010; Varlet & Richardson, 2015; Zivotofsky et al., 2012; Zivotofsky & Hausdorff, 2007). Importantly, however, it is neither obvious nor necessary that two individuals, who are instructed to synchronize walking, focus on the lower limbs. Apparently, we could have given such specific instructions, but this was avoided to capture the synchronization process as naturally as possible.

We assumed that rather than synchronizing with individual segments (like the heel strikes), individuals could perceive the global motion of the walking partner and match this signal with the global motion of one's body. In Experiment 1 to 3, we assessed the discrete heel strike onsets so that we could not test for such assumptions. Moreover, these are not easy to test because we do not know exactly which local and global information is provided by each sensory cue and there do not exist appropriate models that capture the gait cycle estimation process, not to mention the entire synchronization process.

Experiment 4 was, therefore, a first attempt to provide some insights which cues (local or global) of the walking partner might be used. Synchrony variability was quantified by discrete events within the cycles of the six body segments (i.e., local cues) and the principal component of covariation of these segments (i.e., global cues). This technique was adapted from an earlier study, which had indicated that the perception of PLWs might be based on signals that represent the covariation of the principal joints of the PLW (Troje, 2002). Experiment 4 revealed highest synchronization performance when considering these principal components of covariation compared to individual segments. Consistent with our assumptions, it suggests that the global motions of side-by-side walkers become synchronized.

These conclusions are in line with results of Experiment 1. Previously, it was demonstrated that peripheral vision impairs the extraction of global motion components from PLWs (B. Thompson, Hansen, Hess, & Troje, 2007; Troje & Westhoff, 2006). In Experiment 1, PLWs were implemented which provided mostly peripheral visual cues during the walking. If synchronization is based on global rather than local motion, then the poor performance with PLWs in Experiment 1 is plausible.

Next, drawing on the results of Experiment 1 to 3, we expected that auditory cues should increase the heel strike synchronization. Contrary to our expectations, this was not observed. The absence of such auditory dominance could owe to less salient heel strike sounds in Experiment 4. The participants walked on a rubber-like surface. To amplify the low-frequency sounds, we attached a plastic film to the foot sole resulting in heel strike sounds of 48dB LAeq. This amplitude was much lower and the sound had a wider rise time of energy than the simulated "sharp" wooden-floor sounds of 63dB LAeq used in Experiment 1 to 3. Also, in Experiment 4, the visual quality was better because participants were walking with real partners. Although the use of the virtual mannequin in Experiment 2 and 3 reduced some visual deficits of Experiment 1, it is evident that there might still be artifacts that we did not control. For these reasons, we assume that the relative reliability of the auditory and the visual cues were proximate in Experiment 4 and therefore no auditory dominance was observed, and consequently the unimodal matching was not promoted in a similar manner as it was in Experiment 1 to 3.

Furthermore, the combined presentation of audiovisual cues should also increase the global body motion synchronization, as it increased heel strike synchronization in Experiment 2 and 3. Contrary to these expectations, Experiment 4 revealed no superior performance with audiovisual cues. Moreover, it is surprising that global body synchronization was similarly high in the auditory condition, in which supposedly only the discrete footsteps sounds of the walking partner were available. We attribute both results to an insufficient sensory manipulation. In this experiment, it was impossible to entirely eliminate all visual cues in the auditory conditions, and all auditory cues in the visual condition. Moreover, there might have been other haptic, proprioceptive, and vestibular cues, which could be used for synchronization. Due to the implementation of a virtual environment, we are confident that the sensory channels were manipulated sufficiently in Experiment 1 to 3. In Experiment 4, this cannot be assured. In all sensory manipulation conditions, there were still other sensory cues that were not controlled (see Section 0.5.4). Consequently, it might be that individuals always retrieve the global motion of the walking partner and match this signal with the global motion of one's movements based on as many cues as possible. We speculate that heel strike synchronization (assessed in Experiment 1 to 3) might be then a mere by-product of this more global synchronization processes. In the following, we discuss this idea and present a more "speculative" theory of how synchronization can be achieved.

In light of the Information Processing Approach

Individuals perceive global rather than local motion of the walking partner. To synchronize with a single segment of the walking partner like the foot, an individual has to estimate foot motion. This can be done based on continuous information or one uses some repeating and discrete events of the gait cycle. These could be, for example, the moments when the feet contact the ground (as considered in Experiment 1 to 3). The individual can estimate when and where a foot reaches the ground by relying on audiovisual cues from the partner. Single sensory cues provide this information ambiguously, as discussed in many previous sections.

Considering that all sensory cues are processed to some extent ambiguously and noisy, a strategy that might lead to the best synchronization could be to base one's estimations on a combined signal. In Experiment 1 to 3 we discussed this and concluded that synchronization was increased because participants were synchronizing with an audiovisual signal. In this context, we referred, however, to the effects of sensory integration on the estimate of an event that can also be predicted by individual sensory cues alone. Auditory and visual cues were here mostly redundant. Although it is more noisy and ambiguous, heel strike onsets can still be estimated based on single cues. It was demonstrated that the auditory cues were most reliable and that visual cues seem to further disambiguate the perception of the "where and the when" of the heel strike onsets. This was manifested in increased synchronization.

Nonetheless, another aspect is that cues from several features of the environmental scene can create a perceptual experience that does not exist without their combination. Multiple sensory stimuli from different features (bottom-up) and top-down processes could contribute to this perceptual experience. It might be, for example, that the multitude of sounds and visual images from the motion of several "local" segments of the walking partner contribute to the perception of a "global whole" body motion (e.g., B. Thompson et al., 2007; Bruce, Green, & Georgeson, 2003). The estimation of the partner's gait cycle could be based on such global signal. Individual cues could be irrelevant for the exact estimation of the segment motion, and all cues could contribute to some extent to the process of estimating the globally moving body. From this, the gait cycle position could then be retrieved (directly or indirectly).

Such a mechanism would increase the accuracy and precision of the estimation process. Estimations based on individual sensory signals are biased and noisy. Different environmental cues (i.e., auditory monaural and binaural and visual monocular and binocular cues) inform to some degree about individual variables (e.g., heel strike sounds, visually oscillating head motion, etc.), and to some degree, they report about the same variable (e.g., globally oscillating body motion). Suppose, that noise and bias of estimating individual variables are independent. A way to separate independent noise components from joined information is to consider the spatiotemporal covariation of all individual variables. By relying on this principal covariation—which lies in a lower dimensional subspace of the space spanned by all considered variables—, noise of each individual estimation is canceled out and bias is reduced. In this way, the estimation process would be most efficient. Consequently, we could interpret the observed matching of discrete events of global motion as a strategy of the CNS to optimize synchronization performance.

Individuals match self-motion with the global motion of the walking partner. The results point toward that the individual's global motion signal was synchronized with the global motion of the partner. One's global motion could be interpreted as self-motion. When considering each sensory system individually, they ambiguously provide temporal and/or spatial cues for estimating the motion of individual segments. This was extensively discussed in earlier sections and we illustrate this here with some examples for one's heel strike onsets: A) Watching one's feet may provide relative reliable information about the position of the foot in the marching direction and orthogonal to the marching direction. Yet, from the individual's point of view, it is tricky to estimate the vertical distance between the heel and the surface. B) Footsteps sounds provide more reliable temporal information about heel strike onsets. But in order to estimate the exact onsets, the temporal auditory cues have to be combined with spatial information provided by, for instances, the Interaural Time Differences (ITD) and the Interaural Level Differences (ILD). These cues provide information about the distance to the source so that the temporal heel strike onsets may be estimated less ambiguously by integrating distance information. Yet, these cues provide spatial information still very ambiguously (Brungart et al., 1999; Kolarik et al., 2016;

Goldstein, 2008). C) The vestibular system might mainly provide information about the walking pattern by signals from the otolith organ (De Winkel, 2013). The walking pattern can be described by its linear acceleration with maximum acceleration at heel strike. Yet, acceleration cannot be distinguished from gravity based on otolith stimulation alone (De Winkel, 2013). D) Signals from the somatosensory system are affected by the surface quality and the force with which the foot contacts the ground. Such examples make us believe that the "when and where" of the exact heel strike onset are provided ambiguously when relying on signals from single cues.

It was shown that for perceiving heading direction during walking, sensory ambiguities could be resolved by integrating signals of multiple sensory systems into a single representation of self-motion (Green & Angelaki, 2010). Perceiving self-motion is the result of the integration of signals from the body (vestibular, proprioceptive, motor) and signals that arise externally when displacing the body (optic flow, auditory flow, haptic flow) (Durgin et al., 2007). Signals of multiple sensory systems (visual, vestibular, proprioceptive) are integrated relatively early while the more complex signal is integrated later with existing action-effect representations. Therefore, the final percept of self-motion can be understood as a higher-level cognitive process that does not exist without the integration of multiple signals (Angelaki & Cullen, 2008). It was suggested that the global self-motion signal is the principal variable that is used by internal models of action control for displacing the whole body during walking (see e.g., Durgin et al., 2005). Further support comes from studies claiming that self-motion resides from coordinate transformations of local-centered to a body-centered reference frames (Angelaki & Cullen, 2008) and that global reference frames are employed for controlling multiple body segments in complex motor tasks (Mergner & Rosemeier, 1998). Accordingly, we suggest that self-motion could be the variable that is used by the individual to estimate the phase of one's gait cycle for achieving synchronization during side-by-side walking. This could, as above, eliminate the noise and bias of estimating spatiotemporal motion components ("the when and the where") based on individual events.

Concluding, we suggest that the perceived self-motion could be matched with the global motion signals of the walking partner to achieve movement synchronization during side-by-side walking because this might be the most robust strategy of the CNS to prevent biases and to be efficient. Note that this does not rule out our previous conclusions that an elevated synchronization can be achieved due to unimodal matching and that this is increased by the integration of audiovisual cues. Nevertheless, we suggest that the improved heel strike synchronization in Experiment 1 to 3 in the audiovisual condition could be a by-product of an elevated global motion matching. This was not observed in Experiment 4 because in this experiment the multiple sensory channels were insufficiently manipulated so that it is impossible to attribute the observed effects to single cues (and modalities).

In light of the Dynamical System Approach

Up to now, the results of Experiment 4 were interpreted in light of the Information Processing Approach. Synchronization could be based on Inverse, Feedback, and Feed-forward Models. The concept of matching may be decomposed into a) the perception of asynchronies (Feedback Loop), the computation of asynchrony reducing commands (Inverse Model), and the prediction of upcoming asynchronies (Feed-Forward Model) (Ghez & Krakauer, 2000; Wolpert, Ghahramani, & Jordan, 1995).

Nevertheless, the results may also be interpreted within the framework of the Dynamical System Theory. It was suggested that a fundamental problem of the Information Processing approach is the reliance on the active computation of motor commands based on Inverse Models (see e.g., Grush, 2004). The coordination of segments during walking requires a fast recruitment of many degrees of freedom. Degrees of freedom are parameters of the neuromuscular system that are widely independent of each other. They can be defined on all levels of the system like as neural firing, muscle force, or joint angle changes, etc. The problem is that there are more possible combinations of parameter values than necessary to control the movement. How the subset of degrees of freedom is selected from an almost infinite number of possible degrees of freedom is known as the Degrees of Freedom Problem (see Bernstein, 1967 in Latash, 2012).

Mapping of spatiotemporal to motor commands (Inverse Model) can take place if there is a finite number of parameters (degrees of freedom) to estimate that equals or is less than the number of dimensions (see Latash, 2012). Consider the following simplified example:

$$\begin{aligned} X &= ax1 + bx2 + cx3 \\ Y &= ay1 + by2 + cy3 \\ Z &= az1 + bz2 + cz3. \end{aligned} \tag{6}$$

$$\begin{aligned} X &= ax1 + bx2 + cx3 + dx4 \\ Y &= ay1 + by2 + cy3 + dy4 \\ Z &= az1 + bz2 + cz3 + dz4, \end{aligned} \tag{7}$$

where X, Y, and Z are the spatial coordinates registered via the sensory system; x, y, and z are variables of the neuromuscular system; a, b, c, and d are parameters that have to be specified. In Equation 6, the system can be solved because it has as many variables as spatial coordinates. In Equation 7, there are many ways how spatial coordinates can be mapped onto the variables (i.e., the motor command coordinates of the neuromuscular system). This is just a simple example of a system that has one parameter more than there are spatial coordinates. However, it illustrates the inverse problem. That is, it is difficult to find a solution when there are more ways to specify the parameters than there are available equations (Latash, 2012).

To move to a particular position in the gait cycle, there is a near-infinite number of ways neurons can communicate, joints can be configured, muscles can be activated, and limbs are coordinated. The costs associated with the computation of each possible configuration constitute an insurmountable obstacle for the CNS.

According to Bernstein, the Degree of Freedom Problem can be alleviated by assuming synergies on different levels (i.e., neuronal, joint, muscle, motor command “level”). A synergy is defined as a grouping of degrees of freedom that co-vary to achieve particular movement goals (Latash, 2012). It was proposed that a reduced set of commands controls the muscles by the activation of muscle and joint synergies during walking (Flash & Hochner, 2005). Its final output is reflected in motor neuron activity.

This is consistent with the CPG mechanism introduced in section 0.1.5. The rhythmic activity of motor neurons, muscles, and segments is suggested to originate from the CPG (Lacquaniti et al., 2012). Supraspinal input specifies CPG and reflex functioning by modulation of parameters from which the exact timing emerges. For example, the modulation of a single parameter should be sufficient to increase the angular frequency of oscillations within the CPG. Consistently, it was shown that simple signals from the brain stem initialize walking and control walking speed (Pearson & Gordon, 2000). This then implies that no explicit computation of temporal aspects of each segment is required (Lacquaniti et al., 2012) and that supraspinal control of walking is based on a reduced set of control commands.

Based on the estimation of asynchrony between the individual and the walking partner's global motion, the individual might modulate particular parameters that are not necessarily related to time. The threshold of the tonic stretch reflex (see the Equilibrium Point Hypothesis, Latash, 2012) could be changed which alters the stiffness of the limb or the manipulation of properties of the CPG architecture could alter the angular frequency of the rhythmic activity.

As a consequence, the temporal discrepancies between event and response are not directly corrected between the cycles (K. Torre & Balasubramaniam, 2009). Therefore, in order to synchronize, a reduced set of commands might change the internal oscillations (CPG activity) so that synchrony emerges (Latash, 2012; Ivanenko et al., 2004). Then, synchronization could be understood as matching global motion commands, also in circumstances in which single and discrete cues from, for example, the heel strikes are available. The heel strike onsets would inform about the global motion but are not directly synchronized.

This is consistent with the dynamical system approach. It does not rely on the internal computation of desired motor commands. Some configurations of the neuromuscular system, coupled to environmental variables, lead to different stable states to which the entire system converges. A particular configuration constraints most and give rises to a few possible behavioral solutions. Thus, the control does not lie inside the individual but emerges through interactions within the person-environment system. In this way, the degree of freedom problem seems managed (Warren, 2006; Riley, Richardson, Shockley, & Ramenzoni, 2011; Kelso, 1997).

Often, the choice of approaching synchronization processes by Dynamical System or Information Processing theory is based on the discrete nature of the stimulus, the movement, or whether intentional processes are involved. It was suggested that (intentional) synchronization with discrete events involves event-based timing (captured by the Linear Phase Correction model (LPC) (Schulze & Vorberg, 2002)) while (spontaneous) synchronization with continuous events involves emergent timing (captured by coupled oscillator models) (K. Torre & Balasubramaniam, 2009). We illustrated here that for the intentional synchronization of complex discrete movements and stimuli, both approaches could at least theoretically account for the findings. Nonetheless, in order to test for the two approaches, future studies have to be conducted. For instances, it could be assessed the correlation structure of succeeding steps because the LPC predicts that these are correlated (see e.g., Jacoby, Tishby, Repp, Ahissar, & Keller, 2015).

0.6.4 Practical Implications

The outcomes of this thesis have implications for many areas involving motor coordination:

Rehabilitation. People with Parkinson Disease show elevated variability in several gait characteristics such as cadence and stride time (Hausdorff, 2005). A common treatment is the presentation of external cues. They are supposed to substitute the lacking internal cues that are usually provided by the basal ganglia (Keus, Bloem, Hendriks, Bredero-Cohen, & Munneke, 2007). Auditory metronomes may provide an external rhythm and this way function as time-keeper. Cadence and walking speed can be improved (see e.g., Freedland et al., 2002; Lim et al., 2005; Willems et al., 2007; Hausdorff et al., 2007). Visual cues—like stepping stones, walking sticks, floor markers, and to visually focus environmental objects—provide additional information about walking speed (i.e., optical flow) and give more orientation in space. Furthermore, these cues were implemented as a means to improve spatial characteristics like stride length (Keus et al., 2007). Rhythmic visual cueing was also implemented (Rochester et al., 2010, 2005, 2007), but most studies showed that auditory cueing results in better treatment outcomes (Rochester et al., 2005, 2007). A limitation of auditory cueing might be that it more strongly alters the natural stride dynamics. This increases falling risks (Sejdić et al., 2012).

There seems to be no clear evidence that the combined presentation of audiovisual cues improves gait characteristics in such circumstances. It was shown that the audiovisual cue combination reduces the beneficial effects of the visual cues on stride length (Suteerawattananon, Morris, Etnyre, Jankovic, & Protas, 2004). This might be due to interference effects. Interference is the reduction of cognitive processing speed when multiple information is processed simultaneously. This is consistent with the finding that executing a secondary task impairs movement performance of patients with Parkinson disease (Morris, Iansek, Matyas, & Summers, 1996). In the study of Suteerawattananon et al. the auditory cues were metronome events, and the visual cues were colored parallel lines on the floor (Suteerawattananon et al., 2004). There might always be a tradeoff when attempting to improve spatial and temporal characteristics simultaneously by different stimuli. We suggest that a more efficient way to combine audiovisual cues could be to exploit audiovisual cues that co-occur in nature.

A walking partner provides naturally co-occurring visual and auditory cues of cadence, walking speed, stride length, etc. Taking into account that both represent a different aspect of the environment with higher precision, the integration of audiovisual cues can complement each other. This thesis was a first attempt to study the effects of audiovisual cue integration for complex movements. This should be extended and similar studies can be conducted in the framework of gait rehabilitation. A possible rehabilitation technique might be to instruct patients to synchronize movements with a careful training partner. Considering the great ease of synchronizing movements with a social partner, this might facilitate the motor control processes by providing relevant spatiotemporal cues.

Team sports. Another area for which the outcomes of this thesis can have implications is team sports. Most team sports require a highly precise and accurate coordination of movements with the interaction partner. Success and failure depend on the time differences of tens of seconds. For instances, success in crew rowing requires the synchronization of rowing movements of the crew members to most efficiently dislocate the shell. More recently, there have been findings that antiphase rowing is mechanically more efficient than in-phase rowing because it minimizes power loss due to fluctuations in velocity. However, it is an uncommon practice because individuals find it easier to synchronize in-phase.

From a dynamical system perspective, this may reside from a greater in-phase stability (de Brouwer et al., 2013). From an Information Processing Approach, this may reside from the priming of actions through action observation (Morsella et al., 2009). Results of the present thesis suggest that the additional presentation of auditory cues at discrete endpoints of the rowing movement cycle might decrease the variability (and therefore increase stability) of the movements. Consistently, it was shown that auditory cues increased movement synchronization and boat speed (see Schaffert & Mattes, 2016). Considering that in our experiments auditory cues had stronger effects than visual cues when they were temporally displaced, antiphase movement synchronization may be improved through the presentation of discrete auditory cues in antiphase for each team member.

Skill acquisition in virtual environments. Another example is the learning or relearning of particular movement sequences based on the observation of other individuals performing these intended movements. The imitation of others is a common practice for skill acquisition (see Ashford, Bennett, & Davids, 2006). This imitation learning implies that the learner translates the observed action into action execution that is similar to the observed one (Rizzolatti & Craighero, 2004). In this context, it were usually studied the effects of visual cues (see Schaffert & Mattes, 2016). The present thesis demonstrated that a focus on visual cues comes too short and that a multimodal approach should be implemented.

For explaining imitation-learning effects, much attention received the Common Coding Approach (Hommel et al., 2001). According to this approach, movement execution and learning are facilitated when the individual observes a human model performing the intended actions. This is supposed to occur because the perceived movements prime the intended movements via shared cognitive representations. As underlying neural correlate was identified the MNs system. It has been found that MNs also respond to stimuli showing PLWs or other point light displays that performed complex actions such as climbing (Gallese & Goldman, 1998; Rizzolatti, Fogassi, & Gallese, 2001). Taking into account that perceiving complex biological stimuli can activate the MN system, the spatiotemporal alignment of movements to the movements of a perceived partner should be facilitated.

However, observational learning may have shortcomings. According to Prinz, the perception-action priming effect weakens if the observed actions are less familiar to actions of one's motor repertoire (see Rizzolatti & Craighero, 2004). Thus, the presentation of a model with dissimilar spatiotemporal characteristics might be inefficient for skill acquisition. A pilot study conducted in our laboratory did not reveal spontaneous gait synchronization when body characteristics were different. Similarly, according to the Dynamical System Theory, coupled oscillators do not synchronize when their angular frequencies differ (Pikovsky et al., 2001). Consistently, the adjustment to an external rhythm takes more time, the more the temporal intervals between events (frequency) deviates from the preferred cadence of the individual (Castiello, Ansuini, Bulgheroni, Scaravilli, & Nicoletti, 2009). Because morphology and dynamics of individuals can differ, this highlights the challenges of using models for imitation learning. Finding appropriate models for the particular individual might be difficult and cost intensive.

This might be overcome by using a virtual environment. The spatiotemporal properties of the virtually generated stimuli can be systematically manipulated to construct an optimal human model. A possible caveat is that humans perceptual-action system is less sensitive to virtually generated movements. Kilner et al. (2003) demonstrated that observing motion trajectories of another person leads to interferences when the observed and the to-be-executed movements were perpendicular to each other (i.e., Congruency Effect) (Kilner et al., 2003). The congruency effect was not replicated when observing a robot executing the movements. In agreement, it was shown that participants imitated faster human motion (see Iacoboni, 2009) and had greater cortical activation when observing human motion compared to robotic motion (Kessler et al., 2006). Moreover, when coordinating with a robot, no spontaneous synchronization was witnessed (Mörtl et al., 2012). This suggests that only the perception of the actions of other humans facilitate action execution. One study proposed that this is mediated by the attribution of agency to the stimulus (Stanley, Gowen, & Miall, 2007) while another study argued for the particular kinematics as mediating factor (Oztop, Franklin, Chaminade, & Cheng, 2005). Although it is yet not clear and there are also studies claiming that differences disappear when having experience with the artificial stimulus (Press, Gillmeister, & Heyes, 2007), a solution for this possible limitation could be the construction of appropriate stimuli using motion capture. Previous studies have demonstrated its successful implementation for the acquisition of movement routines (Bardy, 2011; Varlet et al., 2013). We here demonstrated its implementation for creating virtual interaction partner and for updating these stimuli online during movement execution.

The efficient learning of skills is also a crucial aspect in many professions such as medicine. Nowadays, virtual environment techniques are used to increase surgical skills by observational learning and feedback. Although there is evidence for skill acquisition and transfer, its effectiveness is still an issue to debate (Seymour, 2008). Another area is manufacturing (see e.g., Mujber, Szecsi, & Hashmi, 2004). Virtual environments can model structure, status, and behavior of the particular system. Their function and control can be trained cheaper and safer. Many tasks are repetitive in nature. In both cases, a multimodal approach complementing visual cues with auditory and haptic cues may improve skill acquisition and increases safety. This thesis demonstrated that temporally displaced cues and the quality of the visual stimuli decrement timing performance. This has implications when using virtual environments for the learning or/and relearning of movements in contexts where there is a high demand for temporally precise motor actions such as in rehabilitation, sports, or surgery. It demands spatiotemporal alignment when displaying the stimuli. Considering computational complexity for achieving real-time rendering of audiovisual stimuli, this is non-trivial. Besides, we have shown that synchronization behavior might differ depending on the coupling direction. This imposes high demands on the virtual environment. It requires audiovisual spatiotemporal alignment in real-time so that the stimulus can adapt to the movements of the individual.

0.6.5 Conclusions

This dissertation approached the issue of how complex movements (here walking) become synchronized in a complex movement coordination task. This was here the movement synchronization during side-by-side walking. It is quite difficult to study such phenomena in a psychophysical context. In an attempt to bridge the gap between highly controlled sensorimotor synchronization paradigms and more ecologically valid studies of interpersonal coordination, we employed elaborate virtual environment techniques. The model used in Experiment 3 is a prime example for this. Here, we employed the classical cross-modal perception-action paradigm by systematically manipulating the temporal asynchronies between naturally occurring synchronous audiovisual cues of an interaction partner. By virtually simulating the stimuli, we could create sensory conflicts that seem impossible to implement when using natural stimuli. Of course, this placed very high demands on the team and equipment of the virtual environment. For instances, it is challenging to synchronize on a millisecond scale a sensitive virtual environment system consisting of motion capture, wireless sound projection, and 3-channel video projection systems. Then, individuals walked for hours next to a 9m screen while their position was captured to update in real-time the audiovisual information they received. Here, it was required a high capture quality and all system delays must have been known and held constant. I (the candidate) always tried to maintain a balance between the technical proficiency for solving problems related to such systems and my ambitions for understanding the complex theoretical concepts of perception-action mechanisms. Although we have met many obstacles throughout the years, we pronounce that our goals seem to be met. This thesis provides the following theoretical main conclusions: A) Audiovisual cues are integrated to improve movement synchronization during side-by-side walking. B) Unimodal (here auditory) matching seems to facilitate step synchronization when the cues are salient. C) To achieve synchronization, self-motion might be matched with the global body motion of the walking partner.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current biology*, *14*(3), 257–262.

Alink, A., Euler, F., Galeano, E., Krugliak, A., Singer, W., & Kohler, A. (2011). Auditory motion capturing ambiguous visual motion. *Frontiers in psychology*, *2*.

Allison, A., & Draghici, M. (2013). *Gonyaulax bioluminescence.* (`http://www.mbari.org/staff/conn/botany/dinos/alimon/biolumin.htm`.)

Andersen, T. S., Tiippana, K., & Sams, M. (2005). Maximum likelihood integration of rapid flashes and beeps. *Neuroscience letters*, *380*(1), 155–160.

Angelaki, D. E., & Cullen, K. E. (2008). Vestibular system: the many facets of a multimodal sense. *Annu. Rev. Neurosci.*, *31*, 125–150.

Arezes, P. M., Barroso, M. P., Cordeiro, P., Costa, L. G. d., & Miguel, A. S. (2006). Estudo antropométrico da população portuguesa. *ISHST*.

Arrighi, R., Alais, D., & Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of vision*, *6*(3), 6–6.

Arrighi, R., Marini, F., & Burr, D. (2009). Meaningful auditory information enhances perception of visual biological motion. *Journal of Vision*, *9*(4), 25–25.

Arshavsky, V. Y., Lamb, T. D., & Pugh Jr, E. N. (2002). G proteins and phototransduction. *Annual review of physiology*, *64*(1), 153–187.

Ashford, D., Bennett, S. J., & Davids, K. (2006). Observational modeling effects for movement dynamics and movement outcome measures across differing task constraints: a meta-analysis. *Journal of motor behavior*, *38*(3), 185–205.

Atmanspacher, H., & Ruhnau, E. (2012). *Time, temporality, now: Experiencing time and concepts of time in an interdisciplinary perspective.* Springer Science & Business Media.

Bardy, B. G. (2011). Learning new skills in multimodal enactive environments. In *Bio web of conferences* (Vol. 1, p. 00008).

Barraclough, N. E., Xiao, D., Baker, C. I., Oram, M. W., & Perrett, D. I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience*, *17*(3), 377–391.

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*(1), 149–159.

Beek, P., Peper, C., & Stegeman, D. (1995). Dynamical models of movement coordination. *Human Movement Science*, *14*(4), 573–608.

Bentvelzen, A., Leung, J., & Alais, D. (2009). Discriminating audiovisual speed: optimal integration of speed defaults to probability summation when component reliabilities diverge. *Perception*, *38*(7), 966–987.

Berniker, M., & Kording, K. (2011). Bayesian approaches to sensory integration for motor control. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(4), 419–428.

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension: 1. evidence from auditory–visual temporal order judgment. *International Journal of Psychophysiology*, *50*(1), 147–155.

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.

Blauert, J. (1997). *The psychophysics of human sound localization.* MIT Press.

Bohannon, J. (2017). *Synchronised marching boosts confidence.* (`http://www.sciencemag.org/news/2014/08/synchronized-marching-boosts-confidence?rss=1`)

Bove, M., Courtine, G., & Schieppati, M. (2002). Neck muscle vibration and spatial orientation during stepping in place in humans. *Journal of neurophysiology*, *88*(5), 2232–2241.

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.

Bruce, V., Green, P. R., & Georgeson, M. A. (2003). *Visual perception: Physiology, psychology, & ecology.* Psychology Press.

Brungart, D. S., Durlach, N. I., & Rabinowitz, W. M. (1999). Auditory localization of nearby sources. localization of a broadband source. *The Journal of the Acoustical Society of America*, *106*(4), 1956–1968.

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S., & Toiviainen, P. (2013). Influences of rhythm-and timbre-related musical features on characteristics of music-induced movement. *Frontiers in psychology*, *4*, 183.

Burr, D. C., & Santoro, L. (2001). Temporal integration of optic flow, measured by contrast and coherence thresholds. *Vision research*, *41*(15), 1891–1899.

Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral cortex*, *11*(12), 1110–1123.

Campos, J. l., & Bülthoff, H. H. (2012). Multimodal integration during self-motion in virtual reality. In M. Murray & M. Wallace (Eds.), *The neural bases of multisen-*

sory processes. Taylor and Francis.

Carpinella, I., Crenna, P., Rabuffetti, M., & Ferrarin, M. (2010). Coordination between upper-and lower-limb movements is different during overground and treadmill walking. *European journal of applied physiology*, *108*(1), 71–82.

Castiello, U., Ansuini, C., Bulgheroni, M., Scaravilli, T., & Nicoletti, R. (2009). Visuomotor priming effects in parkinson's disease patients depend on the match between the observed and the executed action. *Neuropsychologia*, *47*(3), 835–842.

Cattaneo, L., & Rizzolatti, G. (2009). The mirror neuron system. *Archives of neurology*, *66*(5), 557–560.

Ceux, T., Wagemans, J., Rosas, P., Montagne, G., & Buekers, M. (2006). Perceptual-motor adaptations in a synchronization task: the joint effects of frequency and motion coherence manipulations. *Behavioural brain research*, *168*(2), 226–235.

Chen, K.-M., & Yeh, S.-L. (2009). Asymmetric cross-modal effects in time perception. *Acta psychologica*, *130*(3), 225–234.

Chen, Y., Nakayama, K., Levy, D., Matthysse, S., & Holzman, P. (2003). Processing of global, but not local, motion direction is deficient in schizophrenia. *Schizophrenia research*, *61*(2), 215–227.

Coelho, C. M., & Balaban, C. D. (2015). Visuo-vestibular contributions to anxiety and fear. *Neuroscience & Biobehavioral Reviews*, *48*, 148–159.

Coey, C., Varlet, M., Schmidt, R. C., & Richardson, M. J. (2011). Effects of movement stability and congruency on the emergence of spontaneous interpersonal coordination. *Experimental brain research*, *211*(3-4), 483–493.

Colling, L. J., & Williamson, K. (2014). Entrainment and motor emulation approaches to joint action: Alternatives or complementary approaches? *Frontiers in human neuroscience*, *8*.

Crowell, J. A., Banks, M. S., Shenoy, K. V., & Andersen, R. A. (1998). Visual self-motion perception during head turns. *Nature neuroscience*, *1*(8), 732–737.

Cruse, H., Dean, J., Heuer, H., & Schmidt, R. (1990). Utilization of sensory information for motor control. In *Relationships between perception and action* (pp. 43–79). Springer.

Cullen, K. E. (2012). The vestibular system: multimodal integration and encoding of self-motion for motor control. *Trends in neurosciences*, *35*(3), 185–196.

Cullen, K. E. (2016). The neural encoding of self-generated and externally applied movement: implications for the perception of self-motion and spatial memory. In S. Bernard, C. Lopez, T. Brandt, P. Denise, & P. Smith (Eds.), *The vestibular system in cognitive and memory processes in mammals.* Frontiers.

Cutting, J. E., Moore, C., & Morrison, R. (1988). Masking the motions of human gait. *Attention, Perception, & Psychophysics*, *44*(4), 339–347.

Das, S., Lazarewicz, M., & Finkel, L. H. (2004). Principal component analysis of temporal and spatial information for human gait recognition. In *Engineering in medicine and biology society* (Vol. 2, pp. 4568–4571).

de Brouwer, A. J., de Poel, H. J., & Hofmijster, M. J. (2013). Don't rock the boat: how antiphase crew coordination affects rowing. *PloS one*, *8*(1), e54996.

Demos, A. P., Chaffin, R., Begosh, K. T., Daniels, J. R., & Marsh, K. L. (2012). Rocking to the beat: Effects of music and partner's movements on spontaneous interpersonal coordination. *Journal of Experimental Psychology: General*, *141*(1), 49.

De Winkel, K. (2013). *Multisensory perception of spatial orientation and self-motion* (Unpublished doctoral dissertation). Utrecht University.

de Winkel, K. N., Weesie, J., Werkhoven, P. J., & Groen, E. L. (2010). Integration of visual and inertial cues in perceived heading of self-motion. *Journal of vision*, *10*(12), 1–1.

Dittrich, W. H. (1993). Action categories and the perception of biological motion. *Perception*, *22*(1), 15–22.

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition*, *77*(3), 251–288.

Ducourant, T., Vieilledent, S., Kerlirzin, Y., & Berthoz, A. (2005). Timing and distance characteristics of interpersonal coordination during locomotion. *Neuroscience letters*, *389*(1), 6–11.

Durgin, F. H., Pelah, A., Fox, L. F., Lewis, J., Kane, R., & Walley, K. A. (2005). Self-motion perception during locomotor recalibration: more than meets the eye.

Journal of Experimental Psychology: Human Perception and Performance, *31*(3), 398.

Durgin, F. H., Reed, C., & Tigue, C. (2007). Step frequency and perceived self-motion. *ACM Transactions on Applied Perception (TAP)*, *4*(1), 5.

Duysens, J., & Van de Crommert, H. W. (1998). Neural control of locomotion; part 1: The central pattern generator from cats to humans. *Gait & posture*, *7*(2), 131–141.

Eisenhammer, T., Hübler, A., Packard, N., & Kelso, J. A. (1991). Modeling experimental time series with ordinary differential equations. *Biological cybernetics*, *65*(2), 107–112.

Elliott, M. T., Wing, A. M., & Welchman, A. E. (2010). Multisensory cues improve sensorimotor synchronisation. *European Journal of Neuroscience*, *31*(10), 1828–1835.

Elliott, M. T., Wing, A. M., & Welchman, A. E. (2014). Moving in time: Bayesian causal inference explains movement coordination to auditory beats. *Proceedings of the Royal Society of London B: Biological Sciences*, *281*(1786), 20140751.

Ernst, M. O. (2006). A bayesian view on multimodal cue integration. *Human body perception from the inside out*, *131*, 105–131.

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

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

Fasold, O., von Brevern, M., Kuhberg, M., Ploner, C. J., Villringer, A., Lempert, T., & Wenzel, R. (2002). Human vestibular cortex as identified with caloric stimulation in functional magnetic resonance imaging. *Neuroimage*, *17*(3), 1384–1393.

Fitzpatrick, R. C., Wardman, D. L., & Taylor, J. L. (1999). Effects of galvanic vestibular stimulation during human walking. *The Journal of Physiology*, *517*(3), 931–939.

Flash, T., & Hochner, B. (2005). Motor primitives in vertebrates and invertebrates. *Current opinion in neurobiology*, *15*(6), 660–666.

Freedland, R. L., Festa, C., Sealy, M., McBean, A., Elghazaly, P., Capan, A., … Rothman, J. (2002). The effects of pulsed auditory stimulation on various gait measurements in persons with parkinson's disease. *NeuroRehabilitation*, *17*(1), 81–87.

Frenkel-Toledo, S., Giladi, N., Peretz, C., Herman, T., Gruendlinger, L., & Hausdorff, J. M. (2005). Treadmill walking as an external pacemaker to improve gait rhythm and stability in parkinson's disease. *Movement Disorders*, *20*(9), 1109–1114.

Fujisaki, W., & Nishida, S. (2005). Temporal frequency characteristics of synchrony–asynchrony discrimination of audio-visual signals. *Experimental Brain Research*, *166*(3-4), 455–464.

Fujisaki, W., & Nishida, S. (2009). Audio–tactile superiority over visuo–tactile and audio–visual combinations in the temporal resolution of synchrony perception. *Experimental brain research*, *198*(2-3), 245–259.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, *2*(12), 493–501.

Garcia, M., Chatterjee, A., Ruina, A., Coleman, M., et al. (1998). The simplest walking model: stability, complexity, and scaling. *J Biomech Eng Trans ASME*, *120*(2), 281–288.

Gazzola, V., & Keysers, C. (2009). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fmri data. *Cerebral Cortex*, *19*(6), 1239–1255.

Gelfan, S., & Carter, S. (1967). Muscle sense in man. *Experimental Neurology*, *18*(4), 469–473.

Ghez, C., & Krakauer, J. (2000). The organization of movement. *Principles of neural science*, *656*, 668.

Gibson, J. J. (1958). Visually controlled locomotion and visual orientation in animals. *British journal of psychology*, *49*(3), 182–194.

Giordano, B. L., Visell, Y., Yao, H.-Y., Hayward, V., Cooperstock, J. R., & McAdams, S. (2012). Identification of walked-upon materials in auditory, kinesthetic, haptic, and audio-haptic conditions. *The Journal of the Acoustical Society of America*, *131*(5), 4002–4012.

Goldstein, E. B. (2008). *The blackwell handbook of sensation and perception.* John Wiley & Sons.

Grahn, J. A. (2012). Neural mechanisms of rhythm perception: current findings and future perspectives. *Topics in Cognitive Science*, *4*(4), 585–606.

Grahn, J. A., Henry, M. J., & McAuley, J. D. (2011). Fmri investigation of cross-modal interactions in beat perception: audition primes vision, but not vice versa. *Neuroimage*, *54*(2), 1231–1243.

Grasso, R., Prévost, P., Ivanenko, Y. P., & Berthoz, A. (1998). Eye-head coordination for the steering of locomotion in humans: an anticipatory synergy. *Neuroscience Letters*, *253*(2), 115–118.

Green, A. M., & Angelaki, D. E. (2010). Multisensory integration: resolving sensory ambiguities to build novel representations. *Current opinion in neurobiology*, *20*(3), 353–360.

Grondin, S. (2010). Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics*, *72*(3), 561–582.

Grossman, E. D., Blake, R., & Kim, C.-Y. (2004). Learning to see biological motion: brain activity parallels behavior. *Journal of Cognitive Neuroscience*, *16*(9), 1669–1679.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and brain sciences*, *27*(03), 377–396.

Gurnsey, R., Poirier, F. J., Bluett, P., & Leibov, L. (2006). Identification of 3d shape from texture and motion across the visual field. *Journal of Vision*, *6*(5), 1–1.

Gurnsey, R., Roddy, G., Ouhnana, M., & Troje, N. F. (2008). Stimulus magnification equates identification and discrimination of biological motion across the visual field. *Vision Research*, *48*(28), 2827–2834.

Haken, H., Kelso, J. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological cybernetics*, *51*(5), 347–356.

Harrison, S. J., & Richardson, M. J. (2009). Horsing around: spontaneous four-legged coordination. *Journal of Motor Behavior*, *41*(6), 519–524.

Hartmann, W. M. (1983). Localization of sound in rooms. *The Journal of the Acoustical Society of America*, *74*(5), 1380–1391.

Hausdorff, J. M. (2005). Gait variability: methods, modeling and meaning. *Journal of* *neuroengineering and rehabilitation*, *2*(1), 1.

Hausdorff, J. M., Lowenthal, J., Herman, T., Gruendlinger, L., Peretz, C., & Giladi, N. (2007). Rhythmic auditory stimulation modulates gait variability in parkinson's disease. *European Journal of Neuroscience*, *26*(8), 2369–2375.

Hollands, M. A., Patla, A., & Vickers, J. (2002). "look where you're going!": gaze behaviour associated with maintaining and changing the direction of locomotion. *Experimental Brain Research*, *143*(2), 221–230.

Holmes, N. P. (2009). The principle of inverse effectiveness in multisensory integration: some statistical considerations. *Brain topography*, *21*(3-4), 168–176.

Hommel, B. (2009). Action control according to tec (theory of event coding). *Psychological Research PRPF*, *73*(4), 512–526.

Hommel, B., Elsner, B., Morsella, E., Bargh, J., & Gollwitzer, P. (2009). Acquisition, representation, and control of action. *Oxford handbook of human action*, 371–398.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). Codes and their vicissitudes. *Behavioral and brain sciences*, *24*(05), 910–926.

Honisch, J. J., Elliott, M. T., Jacoby, N., & Wing, A. M. (2016). Cue properties change timing strategies in group movement synchronisation. *Scientific reports*, *6*.

Hove, M. J., Iversen, J. R., Zhang, A., & Repp, B. H. (2013). Synchronization with competing visual and auditory rhythms: bouncing ball meets metronome. *Psychological research*, *77*(4), 388–398.

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual review of psychology*, *60*, 653–670.

Ikeda, H., Blake, R., & Watanabe, K. (2005). Eccentric perception of biological motion is unscalably poor. *Vision research*, *45*(15), 1935–1943.

Issartel, J., Marin, L., & Cadopi, M. (2007). Unintended interpersonal co-ordination: "can we march to the beat of our own drum?". *Neuroscience Letters*, *411*(3), 174–179.

Ivanenko, Y. P., Poppele, R. E., & Lacquaniti, F. (2004). Five basic muscle activation patterns account for muscle activity during human locomotion. *The Journal of physiology*, *556*(1), 267–282.

Iversen, J. R., Patel, A. D., Nicodemus, B., & Emmorey, K. (2015). Synchronization to auditory and visual rhythms in hearing and deaf individuals. *Cognition*, *134*, 232–244.

Jacoby, N., Tishby, N., Repp, B. H., Ahissar, M., & Keller, P. E. (2015). Parameter estimation of linear sensorimotor synchronization models: phase correction, period correction, and ensemble synchronization. *Timing & Time Perception*, *3*(1-2), 52–87.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of neural science* (Vol. 4). McGraw-Hill New York.

Kato, M., & Konishi, Y. (2006). Auditory dominance in the error correction process: a synchronized tapping study. *Brain research*, *1084*(1), 115–122.

Katz, B. F., Flinto, D. Q., Tourain, D., Poirier-Quinot, D., & Bourdot, P. (2015). Blendervr: Open-source framework for interactive and immersive vr. In *Ieee virtual reality conference (ieeevr)* (p. 203-204).

Kay, B., Kelso, J., Saltzman, E., & Schöner, G. (1987). Space–time behavior of single and bimanual rhythmical movements: Data and limit cycle model. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(2), 178.

Kaya, D. (2014). Proprioception: The forgotten sixth sense. *Proprioception and Gender. Foster City, USA: OMICS Group eBooks*.

Keetels, M., & Jean, V. (2012). Perception of synchrony between the senses. In M. Murray & M. Wallace (Eds.), *The neural bases of multisensory processes* (p. 147-177). London: CRC Press/Taylor and Francis.

Keetels, M., & Vroomen, J. (2012). Exposure to delayed visual feedback of the hand changes motor-sensory synchrony perception. *Experimental brain research*, *219*(4), 431–440.

Kelso, J. S. (1997). *Dynamic patterns: The self-organization of brain and behavior.* MIT press.

Kessler, K., Biermann-Ruben, K., Jonas, M., Siebner, H. R., Bäumer, T., Münchau, A., & Schnitzler, A. (2006). Investigating the human mirror neuron system by means of cortical synchronization during the imitation of biological movements. *Neuroimage*, *33*(1), 227–238.

Keus, S. H., Bloem, B. R., Hendriks, E. J., Bredero-Cohen, A. B., & Munneke, M.

(2007). Evidence-based analysis of physical therapy in parkinson's disease with recommendations for practice and research. *Movement disorders*, *22*(4), 451–460.

Kilner, J., Paulignan, Y., & Blakemore, S. (2003). An interference effect of observed biological movement on action. *Current Biology*, *13*(6), 522–525.

Kolarik, A. J., Moore, B. C., Zahorik, P., Cirstea, S., & Pardhan, S. (2016). Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. *Attention, Perception, & Psychophysics*, *78*(2), 373–395.

Kopčo, N., & Shinn-Cunningham, B. G. (2011). Effect of stimulus spectrum on distance perception for nearby sources a. *The Journal of the Acoustical Society of America*, *130*(3), 1530–1541.

Kurtzer, I. L. (2014). Long-latency reflexes account for limb biomechanics through several supraspinal pathways. *Frontiers in integrative neuroscience*, *8*.

Lacquaniti, F., Grasso, R., & Zago, M. (1999). Motor patterns in walking. *Physiology*, *14*(4), 168–174.

Lacquaniti, F., Ivanenko, Y. P., & Zago, M. (2012). Patterned control of human locomotion. *The Journal of physiology*, *590*(10), 2189–2199.

Lamas, J., Silva, C. C. L., Silva, R. M. M. T. d., Mouta, S., Campos, J. C., & Santos, J. A. (2015). Measuring end-to-end delay in real-time auralisation systems. In *10th european congress and exposition on noise control engineering, maastrich, netherderlands* (pp. 791–796).

Lappe, M., Bremmer, F., & Van den Berg, A. (1999). Perception of self-motion from visual flow. *Trends in cognitive sciences*, *3*(9), 329–336.

Larsson, P., Västfjäll, D., & Kleiner, M. (2004). Perception of self-motion and presence in auditory virtual environments. In *Proceedings of seventh annual workshop presence* (pp. 252–258).

Latash, M. L. (2012). *Fundamentals of motor control.* Academic Press.

Lécuyer, A., Vidal, M., Joly, O., Mégard, C., & Berthoz, A. (2004). Can haptic feedback improve the perception of self-motion in virtual reality? In *Haptic interfaces for virtual environment and teleoperator systems, 2004. haptics'04. proceedings. 12th international symposium on* (pp. 208–215).

Levine, D., Richards, J., & Whittle, M. W. (2012). *Whittle's gait analysis.* Elsevier

Health Sciences.

Lim, I., van Wegen, E., de Goede, C., Deutekom, M., Nieuwboer, A., Willems, A., … Kwakkel, G. (2005). Effects of external rhythmical cueing on gait in patients with parkinson's disease: a systematic review. *Clinical rehabilitation*, *19*(7), 695–713.

Loeb, G. E. (1989). Neural control of locomotion. *BioScience*, *39*(11), 800–804.

Lopez, C. (2016). The vestibular system: balancing more than just the body. *Current opinion in neurology*, *29*(1), 74–83.

Lysakowski, A., & Goldberg, J. M. (2004). Morphophysiology of the vestibular periphery. In *The vestibular system* (pp. 57–152). Springer.

Maculewicz, J., Erkut, C., & Serafin, S. (2016). An investigation on the impact of auditory and haptic feedback on rhythmic walking interactions. *International Journal of Human-Computer Studies*, *85*, 40–46.

Marmelat, V., Delignières, D., Torre, K., Beek, P. J., & Daffertshofer, A. (2014). Human paced walking: Followers adopt stride time dynamics of leaders. *Neuroscience letters*, *564*, 67–71.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle* (Vol. 1). Mit Press.

Mates, J., Müller, U., Radil, T., & Pöppel, E. (1994). Temporal integration in sensorimotor synchronization. *Journal of cognitive neuroscience*, *6*(4), 332–340.

Mendonça, A. C. (2012). *Audiovisual perception of biological motion* (Unpublished doctoral dissertation). University of Minho, Braga, Portugal.

Mendonça, A. C., Santos, J. A., & López-Moliner, J. (2011). The benefit of multisensory integration with biological motion signals. *Experimental brain research*, *213*(2-3), 185–192.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. *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.

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

Mergner, T., & Rosemeier, T. (1998). Interaction of vestibular, somatosensory and visual signals for postural control and motion perception under terrestrial and microgravity conditions-a conceptual model. *Brain Research Reviews*, *28*(1), 118–135.

Merker, B. H., Madison, G. S., & Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex*, *45*(1), 4–17.

Metzger, W. (1934). Beobachtungen über phänomenale identität. *Psychological Research*, *19*(1), 1–60.

Moiseff, A., & Copeland, J. (2010). Firefly synchrony: a behavioral strategy to minimize visual clutter. *Science*, *329*(5988), 181–181.

Morris, M. E., Iansek, R., Matyas, T. A., & Summers, J. J. (1996). Stride length regulation in parkinson's disease. *Brain*, *119*(2), 551–568.

Morsella, E. E., Bargh, J. A., & Gollwitzer, P. M. (2009). *Oxford handbook of human action.* Oxford University Press.

Mörtl, A., Lorenz, T., Vlaskamp, B. N., Gusrialdi, A., Schubö, A., & Hirche, S. (2012). Modeling inter-human movement coordination: synchronization governs joint task dynamics. *Biological cybernetics*, *106*(4-5), 241–259.

Mujber, T. S., Szecsi, T., & Hashmi, M. S. (2004). Virtual reality applications in manufacturing process simulation. *Journal of materials processing technology*, *155*, 1834–1838.

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., & Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area f5) of the monkey. *Journal of neurophysiology*, *78*(4), 2226–2230.

Murray-Smith, R., Ramsay, A., Garrod, S., Jackson, M., & Musizza, B. (2007). Gait alignment in mobile phone conversations. In *Proceedings of the 9th international conference on human computer interaction with mobile devices and services* (pp. 214–221).

Neetens, A. (1984). *Visual system in myelin disorders* (Vol. 7). Springer Science & Business Media.

Nessler, J. A., De Leone, C. J., & Gilliland, S. (2009). Nonlinear time series analysis of knee and ankle kinematics during side by side treadmill walking. *Chaos: An*

Interdisciplinary Journal of Nonlinear Science, *19*(2), 026104.

Nessler, J. A., & Gilliland, S. J. (2009). Interpersonal synchronization during side by side treadmill walking is influenced by leg length differential and altered sensory feedback. *Human movement science*, *28*(6), 772–785.

Nessler, J. A., & Gilliland, S. J. (2010). Kinematic analysis of side-by-side stepping with intentional and unintentional synchronization. *Gait & posture*, *31*(4), 527–529.

Nessler, J. A., Gutierrez, V., Werner, J., & Punsalan, A. (2015). Side by side treadmill walking reduces gait asymmetry induced by unilateral ankle weight. *Human movement science*, *41*, 32–45.

Nessler, J. A., Kephart, G., Cowell, J., & De Leone, C. J. (2011). Varying treadmill speed and inclination affects spontaneous synchronization when two individuals walk side by side. *Journal of applied biomechanics*, *27*(4), 322–329.

Nessler, J. A., McMillan, D., Schoulten, M., Shallow, T., Stewart, B., & De Leone, C. (2013). Side by side treadmill walking with intentionally desynchronized gait. *Annals of biomedical engineering*, *41*(8), 1680–1691.

Nexus 2.0. (2016). (`http://www.vicon.com/products/software/nexus`)

Norman, J. F., Payton, S. M., Long, J. R., & Hawkes, L. M. (2004). Ageing and the perception of biological motion. *Psychology and aging*, *19*(1), 219.

Oliveira, A., Campos, G., Dias, P., Murphy, D. T., Viera, J., Mendonça, C., & Santos, J. (2013). Real-time dynamic image-source implementation for auralisation. In *Proceedings of the 16th international conference on digital audio effects:* (pp. 368–372).

Oullier, O., De Guzman, G. C., Jantzen, K. J., Lagarde, J., & Kelso, J. S. (2008). Social coordination dynamics: Measuring human bonding. *Social neuroscience*, *3*(2), 178–192.

Oztop, E., Franklin, D. W., Chaminade, T., & Cheng, G. (2005). Human–humanoid interaction: is a humanoid robot perceived as a human? *International Journal of Humanoid Robotics*, *2*(04), 537–559.

Pastore, R. E., Flint, J. D., Gaston, J. R., & Solomon, M. J. (2008). Auditory event perception: The source-perception loop for posture in human gait. *Attention, Perception, & Psychophysics*, *70*(1), 13–29.

Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the action simulation for auditory prediction (asap) hypothesis. *Frontiers in Systems Neuroscience*, *8*.

Patel, M., Magnusson, M., Kristinsdottir, E., & Fransson, P.-A. (2009). The contribution of mechanoreceptive sensation on stability and adaptation in the young and elderly. *European journal of applied physiology*, *105*(2), 167–173.

Pearson, K., & Gordon, J. (2000). Locomotion. In E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), *Principles of neural science* (chap. 37). New York: McGraw-Hill New York.

Pellegrino, G. d., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental brain research*, *91*(1), 176–180.

Peters, B. T., Haddad, J. M., Heiderscheit, B. C., Van Emmerik, R. E., & Hamill, J. (2003). Limitations in the use and interpretation of continuous relative phase. *Journal of biomechanics*, *36*(2), 271–274.

Peuskens, H., Vanrie, J., Verfaillie, K., & Orban, G. A. (2005). Specificity of regions processing biological motion. *European Journal of Neuroscience*, *21*(10), 2864–2875.

Pikovsky, A., Rosenblum, M., & Kurths, J. (2001). A universal concept in nonlinear sciences. *Self*, *2*, 3.

Pizzera, A., & Hohmann, T. (2015). Acoustic information during motor control and action perception: A review. *Open Psychology Journal*, *8*(3), 183–191.

Press, C., Gillmeister, H., & Heyes, C. (2007). Sensorimotor experience enhances automatic imitation of robotic action. *Proceedings of the Royal Society of London B: Biological Sciences*, *274*(1625), 2509–2514.

Preuss, N., Hasler, G., & Mast, F. W. (2014). Caloric vestibular stimulation modulates affective control and mood. *Brain stimulation*, *7*(1), 133–140.

Prochazka, A., & Ellaway, P. (2012). Sensory systems in the control of movement. *Comprehensive Physiology*.

Prochazka, A., Gritsenko, V., & Yakovenko, S. (2002). Sensory control of locomotion: reflexes versus higher-level control. In *Sensorimotor control of movement and* *posture* (pp. 357–367). Springer.

Python 2.0. (2016). (`https://www.python.org`)

Quesque, F., Ruggiero, G., Mouta, S., Santos, J., Iachini, T., & Coello, Y. (2016). Keeping you at arm's length: modifying peripersonal space influences interpersonal distance. *Psychological Research*, 1-12.

Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychonomic bulletin & review*, *12*(6), 969–992.

Repp, B. H., & Keller, P. E. (2004). Adaptation to tempo changes in sensorimotor synchronization: Effects of intention, attention, and awareness. *Quarterly Journal of Experimental Psychology Section A*, *57*(3), 499–521.

Repp, B. H., & Keller, P. E. (2008). Sensorimotor synchronization with adaptively timed sequences. *Human movement science*, *27*(3), 423–456.

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(5), 1085.

Repp, B. H., & Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological research*, *68*(4), 252–270.

Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: a review of recent research (2006–2012). *Psychonomic bulletin & review*, *20*(3), 403–452.

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human movement science*, *26*(6), 867–891.

Richardson, M. J., Marsh, K. L., & Schmidt, R. C. (2005). Effects of visual and verbal interaction on unintentional interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(1), 62.

Riley, M. A., Richardson, M., Shockley, K., & Ramenzoni, V. C. (2011). Interpersonal synergies. *Frontiers in psychology*, *2*, 38.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–92.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1997). The space around us.

Science, *277*(5323), 190–191.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*(9), 661–670.

Rochester, L., Baker, K., Hetherington, V., Jones, D., Willems, A.-M., Kwakkel, G., … Nieuwboer, A. (2010). Evidence for motor learning in parkinson's disease: acquisition, automaticity and retention of cued gait performance after training with external rhythmical cues. *Brain research*, *1319*, 103–111.

Rochester, L., Hetherington, V., Jones, D., Nieuwboer, A., Willems, A.-M., Kwakkel, G., & Van Wegen, E. (2005). The effect of external rhythmic cues (auditory and visual) on walking during a functional task in homes of people with parkinson's disease. *Archives of physical medicine and rehabilitation*, *86*(5), 999–1006.

Rochester, L., Nieuwboer, A., Baker, K., Hetherington, V., Willems, A.-M., Chavret, F., … Jones, D. (2007). The attentional cost of external rhythmical cues and their impact on gait in parkinson's disease: effect of cue modality and task complexity. *Journal of Neural Transmission*, *114*(10), 1243-1248.

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.

Schaffert, N., & Mattes, K. (2016). Influence of acoustic feedback on boat speed and crew synchronization in elite junior rowing. *International journal of Sports Science & Coaching*, *11*(6), 832–845.

Schmidt, R. C., Fitzpatrick, P., Caron, R., & Mergeche, J. (2011). Understanding social motor coordination. *Human movement science*, *30*(5), 834–845.

Schmidt, R. C., Nie, L., Franco, A., & Richardson, M. J. (2014). Bodily synchronization underlying joke telling. *Frontiers in human neuroscience*, *8*, 633.

Schmidt, R. C., & O'Brien, B. (1997). Evaluating the dynamics of unintended interpersonal coordination. *Ecological Psychology*, *9*(3), 189–206.

Schmidt, R. C., & Richardson, M. J. (2008). Dynamics of interpersonal coordination. In *Coordination: Neural, behavioral and social dynamics* (p. 281-308). Springer.

Schöner, G. (2016). *Email correspondence.* personal communication.

Schulze, H.-H., & Vorberg, D. (2002). Linear phase correction models for synchronization: Parameter identification and estimation of parameters. *Brain and Cognition*, *48*(1), 80–97.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in cognitive sciences*, *10*(2), 70–76.

Sejdić, E., Fu, Y., Pak, A., Fairley, J. A., & Chau, T. (2012). The effects of rhythmic sensory cues on the temporal dynamics of human gait. *PloS one*, *7*(8), e43104.

Semjen, A., Vorberg, D., & Schulze, H.-H. (1998). Getting synchronized with the metronome: Comparisons between phase and period correction. *Psychological Research*, *61*(1), 44–55.

Seymour, N. E. (2008). Vr to or: a review of the evidence that virtual reality simulation improves operating room performance. *World journal of surgery*, *32*(2), 182–188.

Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport*, *16*(17), 1923–1927.

Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current opinion in neurobiology*, *11*(4), 505–509.

Shockley, K., Baker, A. A., Richardson, M. J., & Fowler, C. A. (2007). Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 201.

Shockley, K., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 326–332.

Siegel, A., & Sapru, H. N. (2006). *Essential neuroscience.* Lippincott Williams & Wilkins.

Silva, C. C., Mendonça, C., Mouta, S., Silva, R., Campos, J. C., & Santos, J. (2013). Depth cues and perceived audiovisual synchrony of biological motion. *PloS one*, *8*(11), e80096.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995.

Stanley, J., Gowen, E., & Miall, R. C. (2007). Effects of agency on movement interference during observation of a moving dot stimulus. *Journal of Experimental*

Psychology: Human Perception and Performance, *33*(4), 915.

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

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(4), 255–266.

Steinicke, F., Visell, Y., Campos, J., & Lécuyer, A. (2013). *Human walking in virtual environments: perception, technology, and applications.* Springer Science & Business Media.

St George, R. J., & Fitzpatrick, R. C. (2011). The sense of self-motion, orientation and balance explored by vestibular stimulation. *The Journal of physiology*, *589*(4), 807–813.

Strogatz, S. H. (2003). *Sync.* Hyperion Books.

Strogatz, S. H., & Stewart, I. (1993). Coupled oscillators and biological synchronization. *Scientific American*, *269*(6), 102–109.

Su, Y.-H. (2014). Visual enhancement of auditory beat perception across auditory interference levels. *Brain and cognition*, *90*, 19–31.

Su, Y.-H., & Pöppel, E. (2012). Body movement enhances the extraction of temporal structures in auditory sequences. *Psychological research*, *76*(3), 373–382.

Suteerawattananon, M., Morris, G., Etnyre, B., Jankovic, J., & Protas, E. (2004). Effects of visual and auditory cues on gait in individuals with parkinson's disease. *Journal of the neurological sciences*, *219*(1), 63–69.

Suzuki, Y., Sakamoto, S., & Gyoba, J. (2004). Effect of auditory information on self-motion perception. In *Proceedings of ica.*

Thaut, M. H., Trimarchi, P. D., & Parsons, L. M. (2014). Human brain basis of musical rhythm perception: Common and distinct neural substrates for meter, tempo, and pattern. *Brain Sciences*, *4*(2), 428–452. doi: 10.3390/brainsci4020428

Thomas, J. P., & Shiffrar, M. (2010). I can see you better if i can hear you coming: Action-consistent sounds facilitate the visual detection of human gait. *Journal of vision*, *10*(12), 14–14.

Thomas, J. P., & Shiffrar, M. (2013). Meaningful sounds enhance visual sensitivity to human gait regardless of synchrony. *Journal of vision*, *13*(14), 8–8.

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), 12–12.

Thompson, D. M. (2005). *Understanding audio: getting the most out of your project or professional recording studio.* Hal Leonard Corporation.

Thornton, I. M. (1998). The visual perception of human locomotion. *Cognitive Neuropsychology*, *15*(6-8), 535–552.

Thornton, I. M., Rensink, R. A., & Shiffrar, M. (2002). Active versus passive processing of biological motion. *Perception*, *31*(7), 837–853.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *nature*, *381*(6582), 520.

Torre, K., & Balasubramaniam, R. (2009). Two different processes for sensorimotor synchronization in continuous and discontinuous rhythmic movements. *Experimental Brain Research*, *199*(2), 157–166.

Torre, K., & Delignières, D. (2008). Distinct ways of timing movements in bimanual coordination tasks: Contribution of serial correlation analysis and implications for modeling. *Acta Psychologica*, *129*(2), 284–296.

Torre, V., Ashmore, J., Lamb, T., & Menini, A. (1995). Transduction and adaptation in sensory receptor cells. *The Journal of Neuroscience*, *15*(12), 7757–7768.

Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of vision*, *2*(5), 2.

Troje, N. F. (2008). Biological motion perception. *The senses: A comprehensive reference*, *2*, 231–238.

Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a ”life detector”. *Current Biology*, *16*(8), 821–824.

Van der Burg, E., Cass, J., Olivers, C., Theeuwes, J., & Alais, D. (2009). Efficient visual search from nonspatial auditory cues requires more than temporal synchrony. *Temporal multisensory processing and its effects on attention*, 63–84.

van der Zwan, R., MacHatch, C., Kozlowski, D., Troje, N., 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.

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

van Ulzen, N. R., Lamoth, C. J., Daffertshofer, A., Semin, G. R., & Beek, P. J. (2008). Characteristics of instructed and uninstructed interpersonal coordination while walking side-by-side. *Neuroscience Letters*, *432*(2), 88–93.

van Ulzen, N. R., Lamoth, C. J., Daffertshofer, A., Semin, G. R., & Beek, P. J. (2010). Stability and variability of acoustically specified coordination patterns while walking side-by-side on a treadmill: Does the seagull effect hold? *Neuroscience letters*, *474*(2), 79–83.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*(3), 598–607.

Varlet, M., Coey, C. A., Schmidt, R. C., & Richardson, M. J. (2012). Influence of stimulus amplitude on unintended visuomotor entrainment. *Human movement science*, *31*(3), 541–552.

Varlet, M., Filippeschi, A., Ben-sadoun, G., Ratto, M., Marin, L., Ruffaldi, E., & Bardy, B. G. (2013). Virtual reality as a tool to learn interpersonal coordination: Example of team rowing. *PRESENCE: Teleoperators and Virtual Environments*, *22*(3), 202–215.

Varlet, M., & Richardson, M. J. (2015). What would be usain bolt's 100-meter sprint world record without tyson gay? unintentional interpersonal synchronization between the two sprinters. *Journal of Experimental Psychology: Human Perception and Performance*.

Vatakis, A., & Spence, C. (2006). Audiovisual synchrony perception for music, speech, and object actions. *Brain research*, *1111*(1), 134–142.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Attention, Perception, & Psychophysics*, *72*(4), 871–884.

Warren, W. H. (2006). The dynamics of perception and action. *Psychological review*, *113*(2), 358.

Warren, W. H., Kay, B. A., & Yilmaz, E. H. (1996). Visual control of posture during walking: functional specificity. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 818.

Welch, R. B., DutionHurt, L. D., & Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Perception & Psychophysics*, *39*(4), 294–300.

Wheat, J. S., & Glazier, P. S. (2005). Measuring coordination and variability in coordination. In K. Davids, K. Bennet, & K. Newell (Eds.), *Movement system variability.* Human kinetics.

Willems, A.-M., Nieuwboer, A., Chavret, F., Desloovere, K., Dom, R., Rochester, L., … Jones, D. (2007). Turning in parkinson's disease patients and controls: the effect of auditory cues. *Movement disorders*, *22*(13), 1871–1878.

Wing, A. M., Doumas, M., & Welchman, A. E. (2010). Combining multisensory temporal information for movement synchronisation. *Experimental brain research*, *200*(3-4), 277–282.

Wing, A. M., Endo, S., Bradbury, A., & Vorberg, D. (2014). Optimal feedback correction in string quartet synchronization. *Journal of The Royal Society Interface*, *11*(93), 20131125.

Wing, A. M., & Kristofferson, A. B. (1973). Response delays and the timing of discrete motor responses. *Perception & Psychophysics*, *14*(1), 5–12.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880.

Wright, R. L., & Elliott, M. T. (2014). Stepping to phase-perturbed metronome cues: multisensory advantage in movement synchrony but not correction. *Frontiers in human neuroscience*, *8*.

Wuerger, S., Meyer, G., Hofbauer, M., Zetzsche, C., & Schill, K. (2010). Motion extrapolation of auditory–visual targets. *Information fusion*, *11*(1), 45–50.

Zahorik, P., Brungart, D. S., & Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica*, *91*(3), 409–420.

Zélanti, P. S., & Droit-Volet, S. (2012). Auditory and visual differences in time perception? an investigation from a developmental perspective with neuropsychological tests. *Journal of experimental child psychology*, *112*(3), 296–311.

Zelaznik, H. N., Spencer, R., & Ivry, R. B. (2002). Dissociation of explicit and implicit timing in repetitive tapping and drawing movements. *Journal of Experimental*

Psychology: Human Perception and Performance, *28*(3), 575.

Zivotofsky, A. Z., Gruendlinger, L., & Hausdorff, J. M. (2012). Modality-specific communication enabling gait synchronization during over-ground side-by-side walking. *Human movement science*, *31*(5), 1268–1285.

Zivotofsky, A. Z., & Hausdorff, J. M. (2007). The sensory feedback mechanisms enabling couples to walk synchronously: An initial investigation. *Journal of neuroengineering and rehabilitation*, *4*, 28.

Appendix A

Appendix

A.1 Coupled Oscillator Model

The coupled oscillator model is the dynamical system that is assumed to underlie the synchronization of movements. It describes the rate of change of the relative phase as a function of a) the angular frequencies of the two oscillators, b) coupling coefficients, and c) actual phase. There are a large number of studies demonstrating that these oscillator models might be valid accounts for describing the coordination dynamics of interacting individuals in social contexts (Schmidt & Richardson, 2008). Synchronization emerges when the net value of the coupling coefficients (K1 and K2) exceeds the angular frequency difference (w1-w2). An individual might modify coupling coefficients and/or angular frequency. So, the synchronization of both oscillators results from the intrinsic dynamics of the system (Warren, 2006; Kelso, 1997; Schmidt et al., 2011; Strogatz & Stewart, 1993).

Equation 1: Simple model of two coupled harmonic oscillators (see Mörtl et al., 2012).

$$\dot{\theta_1} = \omega_1 + K1 sin(\theta_2 - \theta_1)$$
$$\dot{\theta_2} = \omega_2 + K2 sin(\theta_1 - \theta_2),$$

where $\dot{\theta}$ is the change in phase, ω is the angular frequency, K is the coupling function, and θ is the phase of the oscillator. Note that there were proposed many different models (Kay, Kelso, Saltzman, & Schöner, 1987; Eisenhammer, Hübler, Packard,

& Kelso, 1991; Beek, Peper, & Stegeman, 1995). We used here the coupling function of the Kuramoto model for the illustration because it is one of the most popular and best-described models.

A.2 The Control of Walking

Because walking is the coordination of repetitive and stereotypic movements of body segments, the activity of most involved motor neurons can be controlled by the regulation of neural circuits within the spinal cord (Loeb, 1989), called CPG (Pearson & Gordon, 2000).

For a simple illustration of the CPG mechanism, one may consider a CPG that is constituted of just two groups of neurons. Both neurons can receive external input and are connected via inhibitory synapses. Due to a mechanism called reciprocal inhibition, the firing pattern of the more excited neurons suppresses the less excited neurons more than vice versa. Due to reciprocal inhibition and refractory periods after depolarization, self-sustained oscillations are generated. If such a CPG projects onto motor neurons that intervene extensor and flexor muscles, these would be alternately stimulated.

Similar but more complex mechanisms are supposed to underpin the stepping pattern during walking (Cruse, Dean, Heuer, & Schmidt, 1990). Evidence is provided by experimental studies with cats that had a transected spinal cord. Although the transection isolates the hind limb from the superior CNS, cats relearned walking on treadmills. This indicates that CPGs may function without receiving input from higher layers of the CNS. In addition, it was shown that CPG activity maintains when deprived of sensory input (see Pearson & Gordon, 2000). It is still not clear whether CPGs in humans operate in the same way as in animals. Yet, its existence in the human CNS is widely accepted (see Pearson & Gordon, 2000; Duysens & Van de Crommert, 1998). Although its functioning is autonomous, CPG activity can be modified by efferent signals from higher-level brain areas and afferent signals from several sensory systems (Duysens & Van de Crommert, 1998).

Afferent input. Many afferent signals affect walking directly by stimulating motor neurons, by affecting CPG activity, and their integration within supra-spinal processes affects the two former pathways, in a feedback loop. Nerves from sensory receptors project onto alpha motor neurons that innervate muscle groups. Sufficient stimulation of corresponding receptors causes an involuntary muscle contraction, called reflex. Depending on the receptor that is stimulated, a reflex can be as fast as 40ms, or faster (Kurtzer, 2014).

This "direct" control is mainly exerted via the somatosensory system. Somatosensory neurons have receptors for proprioception and cutaneous perception. Proprioceptive receptors are muscle spindles and Golgi tendon organs that are located within the muscle. Muscle spindles signal length changes of extrinsic muscles and contraction of intrafusal muscle fibers. Golgi tendon organs signal tension of muscles. For example, a contracting muscle stimulates muscle spindles that activate alpha motor neurons of the same muscle so that stretch is resisted (Prochazka & Ellaway, 2012). This so-called stretch reflex automatically controls muscle length during walking and illustrates that proprioceptive feedback continuously regulates the activity of the muscles during walking. It enables a quick and automatic force and muscle length adjustments in response to environmental stimuli. In addition, proprioceptive signals can act on the CPG. It was shown that the stimulation of both receptors alters the timing of swing and stance phase mediated via CPG activity (see Cruse et al., 1990). The cutaneous system is a structure of nerve endings with an elastic capsule. The relevant receptors here in place are the Meissner's corpuscles and Merkel cells, which are located on the skin surface (epidermis). They are sensitive to the mechanical force of touch and pressure. Cutaneous receptors project onto alpha motor neurons and on the CPG (Prochazka et al., 2002).

Efferent input. Sensory signals also indirectly affect the control of walking by being integrated within supra-spinal and brain stem processes that influence CPG and muscle activity. The activity of the primary motor cortex and premotor areas exert influence on the motor pathways that descend from the brain stem to the spinal cord. Furthermore, there are circuits between the cerebellum, basal ganglia and cortex areas that intervene in the control from the cortex and brain stem (see e.g., Pearson & Gordon, 2000; Ghez & Krakauer, 2000).

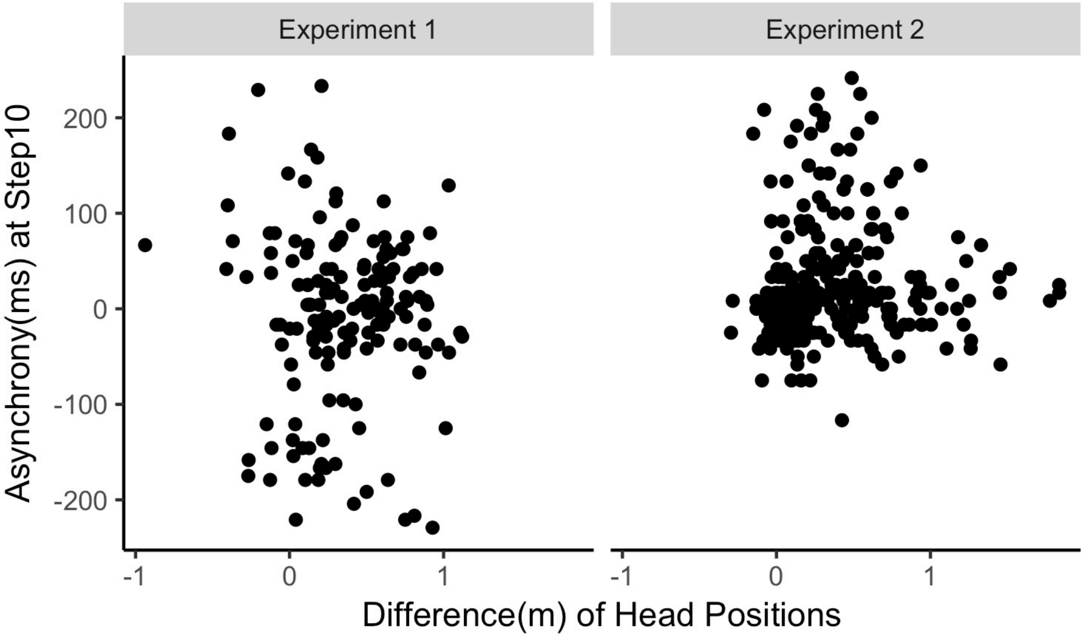
Visual information is mainly used by the primary motor cortex for navigation in space. The cerebellum receives haptic and proprioceptive information. There it is compared with signals from the motor cortex (e.g., to a lesser extent visual information, efferent copies, goals states). The motor cortex integrates particularly visual signals, cutaneous and proprioceptive signals and signals from other subcortical structures like the hippocampus. The cerebellum then sends signals to the brain stem. The brain stem integrates corrective signals with vestibular signals and sends an updated corrective response to the spinal cord. Signals from various sensory systems (visual, vestibular, proprioceptive, efferent) converge in vestibular nuclei and cerebellum for postural control. Moreover, supra-spinal input alters the properties of gamma neurons that intervene in reflex pathways (Ghez & Krakauer, 2000; Loeb, 1989).

A.3 The Vestibular System

The vestibular receptors are located in the semicircular canals indicating angular accelerations and in the otolithic organs indicating linear accelerations. The latter provides more crucial information about the gait pattern if no head rotation and tilt are involved. Receptors in the otolithic organ are stimulated depending on up-down, right-left, and forward-backward movements. But the otolithic organs are also able to detect the direction of the constant linear acceleration due to gravity (Lysakowski & Goldberg, 2004), as the force produced by the action of gravity on the mass of the otoliths generates a static position signal by moving the tips of the kinocilia and stereocilia toward the ground (Coelho & Balaban, 2015). Since the otoliths are stimulated both by linear movements and by the constant linear acceleration gravitational forces, these signals might in some situations cause ambiguity (Cullen, 2012).

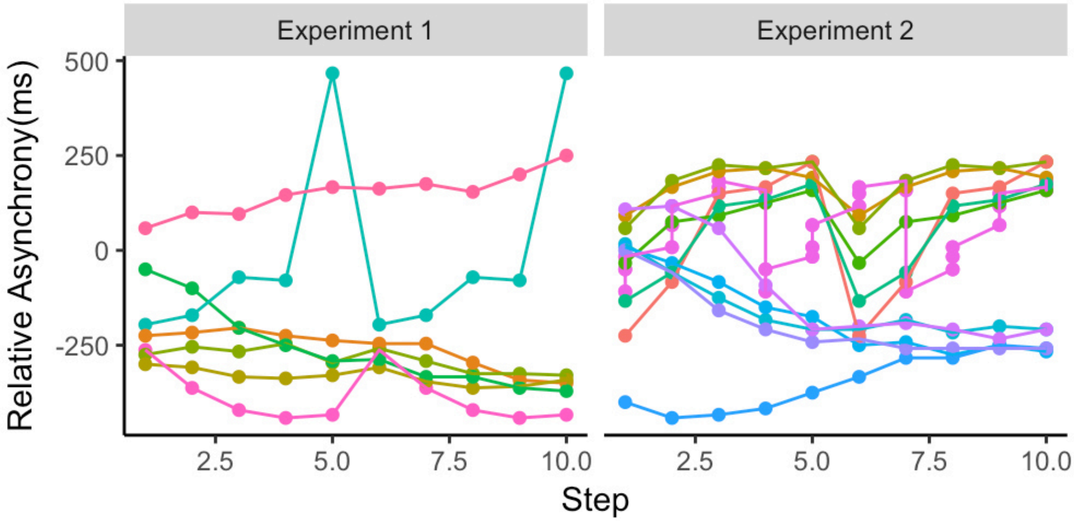
A.4 Spatiotemporal Relations

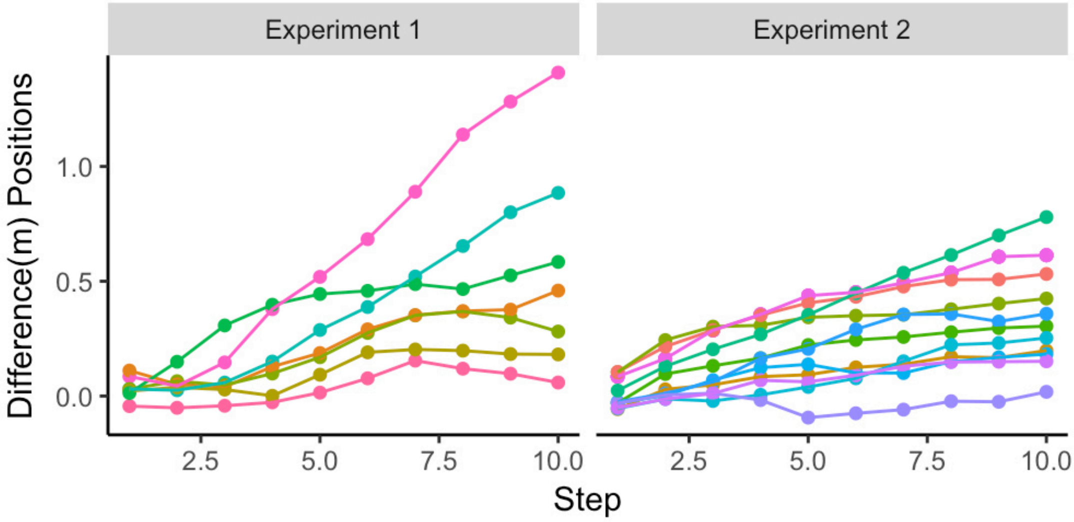
The asynchrony(ms) as a function of difference(m) between the head positions at step 10 of Experiment 1 and Experiment 2. The plot illustrates that there is no relation between both measures.



A.5 Asynchrony Outliers

There may exist a weak relation between asynchrony outliers and an increased distance between the head positions. However, the asynchrony pattern is quite confusing demonstrating some shifts that are biomechanically difficult (e.g., green). Other asynchrony series converge to very large values (e.g., blue). We, therefore, assume that many outlier series may be the result of methodological problems rather than that their cause can be attributed to the particular behavior of the participant.

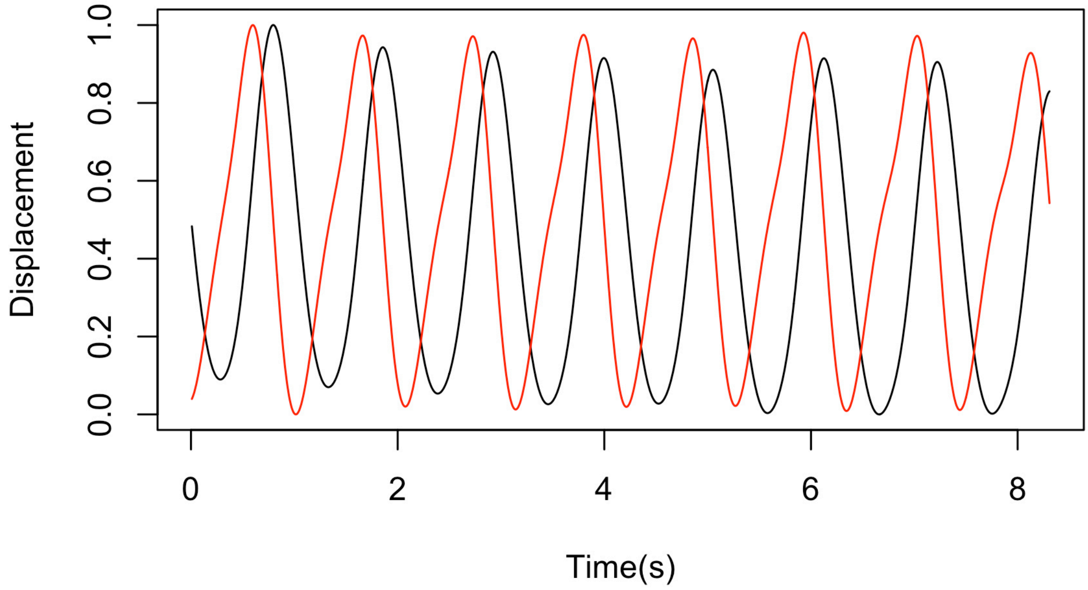




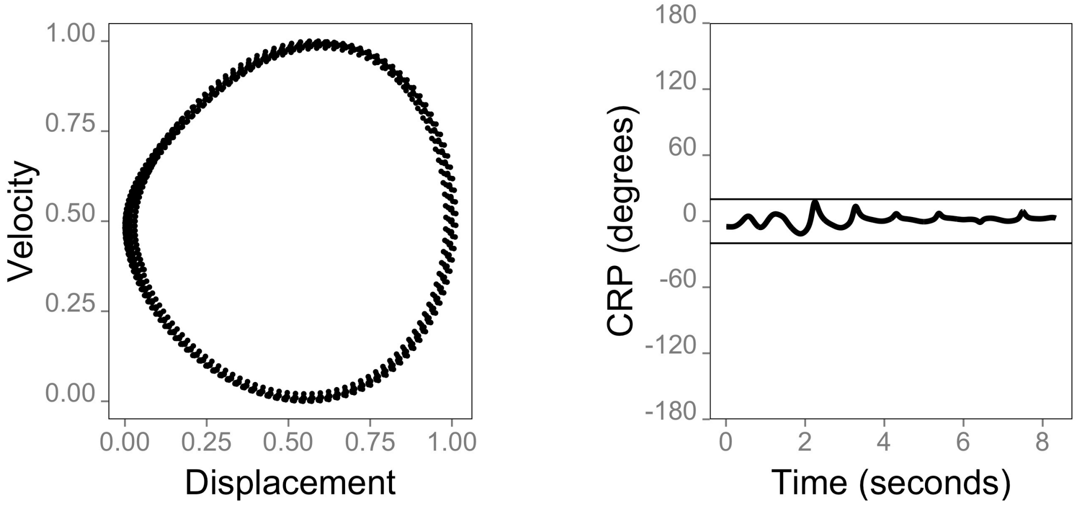
A.6 Continuous Relative Phase

The relative phase is a concept that originated within the dynamical system framework for capturing the coordination between two oscillating entities. The relative phase relationship (e.g., in-phase & anti-phase) is the difference of the obtained phases and can be used as a parameter that describes the dynamic state established by the motion of the oscillators. It is called order parameter because it indicates the order of the movement in space and time (Warren, 2006).There are many ways to calculate the phase, like the spectral method based on Hilbert transform and or the Hybrid Method (see Mörtl et al., 2012; van Ulzen et al., 2008). The most traditional way, called the phase-state method, is to calculate the continuous phase of the two oscillators by creating "phase plane portrait" plots between displacement and its derivate (here instantaneous velocity, see Figure A.6) and taking the arctangent of the instantaneous velocity and displacement in order to obtain the phase angle for each time instance. This procedure works for sinusoidal signals (linear phase increase). In case this requirement is not satisfied, several normalization procedures for the displacement were suggested. The type of normalization may depend on the signal properties (Peters et al., 2003).

We here provide a practical example: First, a Butterworth low-pass filter with a cut-off frequency of 2.4Hz was applied to the displacement of the heels. Second, the instantaneous velocity was computed. Third, displacement and velocity were normalized for each half cycle (Peters et al., 2003). Subsequently, phase plane portraits were constructed and the phase was obtained by transforming the Cartesian coordinates to polar coordinates (i.e., $Phase_i = Arctan(\frac{Velocity_i}{Displacement_i}) * \frac{180}{\pi})$ (Wheat & Glazier, 2005). Then, the CRP phase was computed by subtracting the phase of one oscillator by the phase of the other oscillator for each time instance.



(a) Normalized and filtered vertical displacement (black) and velocity (red) of the heel as a function of time.



(b) Left: Phase Plane Portrait of the time series example displayed above. Right: Continuous Relative Phase as a function of time.

A.7 95% Confidence Intervals of DT Thresholds in Experiment 3

	Sensory	Shift	CI lower.D	CI upper.D	CI lower.E	CI upper.E	CI lower.F	CI upper.F
3	A-Shift	50	1.89	3.46	2.51	Inf	2.58	Inf
4	V-Shift	50	2.59	Inf	Inf	Inf	2.51	Inf
5	A-Shift	100	1.25	1.72	1.84	2.39	1.45	2.23
6	V-Shift	100	2.56	Inf	2.63	Inf	2.46	Inf
7	A-Shift	150	0.76	1.27	1.57	1.98	1.13	1.66
8	V-Shift	150	2.16	7.89	2.68	Inf	2.43	Inf
9	A-Shift	200	0.91	1.31	1.39	1.81	0.71	1.15
10	V-Shift	200	1.90	3.98	2.62	Inf	2.30	Inf
11	A-Shift	250	1.07	1.35	1.31	1.79	0.98	1.61
12	V-Shift	250	1.81	2.50	2.57	Inf	2.24	Inf

A.8 Principle Component Analysis

Here we briefly describe the implemented PCA. In order to obtain PC1, first, the temporal variation of each segment(6) in each spatial dimension(3) was represented as a column vector in a matrix (i.e., 18 columns). Second, each column vector was centralized around zero. Third, the translational component was removed because it would explain already 99% of the variability of the original data set. Fourth, the covariance matrix including the covariances between all variable pairs was computed. Fifth, eigenvalues and unit eigenvectors for this matrix were obtained. The first eigenvector is the dimensions of most co-variation. Its respective eigenvalue reflects how much variance is explained by the vector.