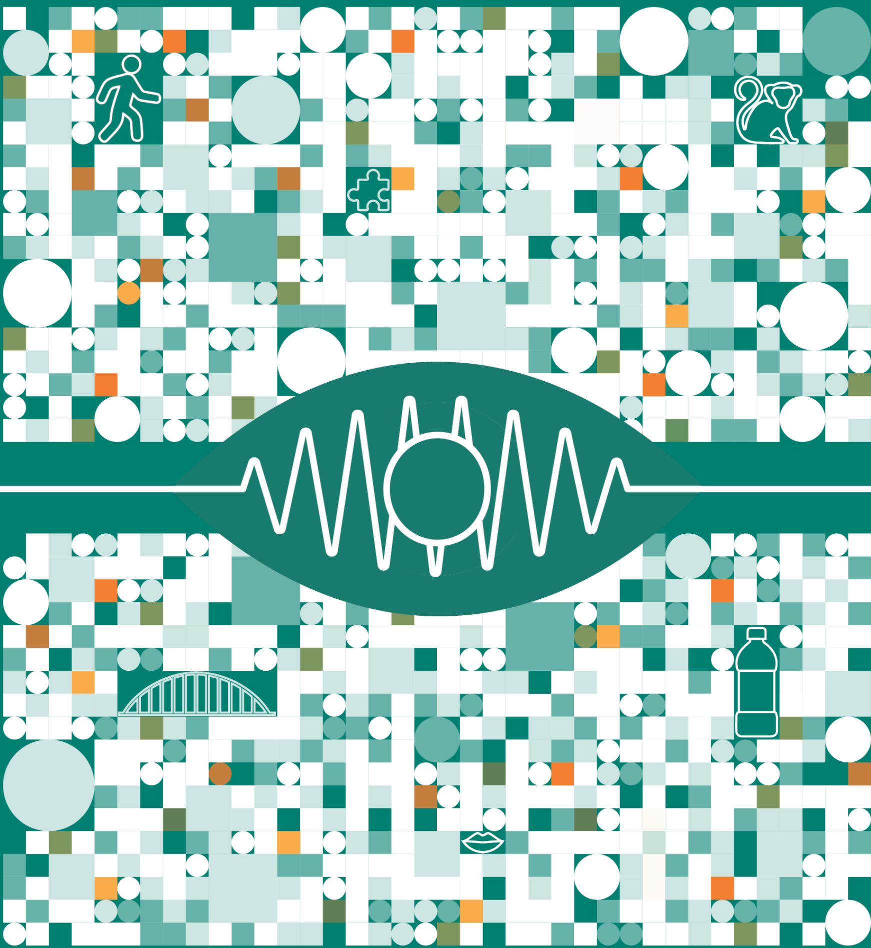
MAX PLANCK INSTITUTE
FOR PSYCHOLINGUISTICS

Perceptual Experience Shapes How Blind and Sighted People Express Concepts in Multimodal Language



EZGI MAMUS

Perceptual experience shapes how blind and sighted people express concepts in multimodal language

Ezgi Mamus

Funding body
This research was funded by Centre for Language Studies, Radboud University.

International Max Planck Research School (IMPRS) for Language Sciences
The educational component of the doctoral training was provided by the International Max Planck Research School (IMPRS) for Language Sciences. The graduate school is a joint initiative between the Max Planck Institute for Psycholinguistics and two partner institutes at Radboud University – the Centre for Language Studies, and the Donders Institute for Brain, Cognition and Behaviour. The IMPRS curriculum, which is funded by the Max Planck Society for the Advancement of Science, ensures that each member receives interdisciplinary training in the language sciences and develops a well-rounded skill set in preparation for fulfilling careers in academia and beyond. More information can be found at www.mpi.nl/imprs

The MPI series in Psycholinguistics
Initiated in 1997, the MPI series in Psycholinguistics contains doctoral theses produced at the Max Planck Institute for Psycholinguistics. Since 2013, it includes theses produced by members of the IMPRS for Language Sciences. The current listing is available at www.mpi.nl/mpi-series

ISBN: 978-94-92910-57-8
Cover design and artwork: Dilara Albayrak
Layout and design: Douwe Oppewal
Printed and bound by Ipskamp Drukkers, Enschede

 The research reported in this thesis was carried out at the Centre for Language Studies, Radboud University and the Max Planck Institute for Psycholinguistics, in Nijmegen, the Netherlands.

Perceptual experience shapes how blind and sighted people express concepts in multimodal language

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.M. Sanders,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 13 mei 2024
om 14.30 uur precies

door

Ayşe Ezgi Mamus
geboren op 2 april 1990
te Istanbul, Turkije

Promotoren:
Prof. dr. H. A. Özyürek
Prof. dr. A. Majid

Copromotor:
Dr. L. J. Speed

Manuscriptcommissie:
Prof. dr. A. S. Meyer (voorzitter)
Prof. dr. A. C. Küntay (Koç Universitesi, Turkije)
Prof. dr. O. Collignon (UCLouvain, België)
Prof. dr. S. Kita (University of Warwick, Verenigd Koninkrijk)
Dr. S. M. Brouwer

To Özcan...

Contents

Chapter 1.	General introduction	9
Chapter 2.	The effect of input sensory modality on the multimodal encoding of motion event	25
Chapter 3.	Lack of visual experience affects multimodal language production: Evidence from congenitally blind and sighted people	45
Chapter 4	Gestures reveal how visual experience shapes concepts in blind and sighted individuals	69
Chapter 5.	General discussion	93
	References	107
	Appendices	
	A. Research data management	125
	B. Supplementary materials for Chapter 4	127
	English Summary	128
	Nederlandse samenvatting	130
	Curriculum Vitae	133
	Publications	134
	Acknowledgements	136

1

General introduction

General Introduction

We experience the world through different perceptual modalities and each modality offers unique affordances that possibly shape our overall understanding of objects and events and therefore our concepts. For example, while observing a car passing by, we also hear the whooshing noise the car makes, which together informs us about the speed of the car. When we communicate about concepts—typically in face-to-face settings—we use spoken words, manual gestures, and facial expressions to convey our multimodal experience and to enhance social interactions (Holler & Levinson, 2019; McNeill, 1992). As with simple sensory experiences, each communication channel has its own affordances and restrictions. For instance, as discussed in Levinson and Holler (2014), spoken languages have a limited and unvarying vocabulary to precisely describe spatial relationships, whereas gesture affords precise depictions of shape, orientation, and the relative location of two objects. Thus, the ability to convey multisensory objects and events through multimodal language may vary based on the affordances of both the input and output channels and their combinatorial possibilities. The aim of the current thesis is to investigate to what extent perceptual experience influences both multimodal language production in speech and gesture as well as the underlying conceptual representations that give rise to these visible behaviors.

Theories of language and cognition, including those focused on multimodal language, differ in their view about the extent to which conceptual representations and language processing are connected to the bodily experience. Although theories can be said to vary on a continuum (see Meteyard et al., 2012 for a review), there is a traditional dichotomy between two views: relatively embodied theories propose that language, including both speech and gesture, is strongly influenced by our sensory and motor experience as this experience is simulated when we process language (e.g., Barsalou, 2016; Hostetter & Alibali, 2008, 2019; Pouw et al., 2014; Pulvermüller, 2013; Wilson, 2002). On the other hand, relatively disembodied symbolic theories suggest that language processing relies on abstract representations that are independent of specific sensory experience (e.g., Levelt, 1989; Mahon & Caramazza, 2008). However, more recent theories propose that our understanding of concepts rely on a combination of simulated and distributional abstract linguistic information (e.g., Connell, 2019; Dove, 2022), and the specific contribution of each type of information depends on various factors specific to an individual and a particular moment in time (Casasanto & Lupyan, 2015; Connell & Lynott, 2014). Therefore, the extent to which perceptual experience shapes conceptual representations remains a matter of ongoing debate (e.g., Davis & Yee, 2021; Dove, 2023; Kemmerer, 2023). Within this intellectual landscape, the aim of this thesis is to examine the degree to which visual experience shapes multimodal language representation and use in blind and sighted individuals.

One way to explore the influence of visual experience on language is to investigate whether and how congenital blindness—i.e., blindness from birth—shapes conceptual representations and how it relates to multimodal language production. A few studies have begun to address this question. While some studies suggest the absence of visual experience does not influence the comprehension and production of language among blind individuals (e.g., Kim et al., 2021; Landau & Gleitman, 1985; Mahon et al., 2009; Özçalışkan et al., 2016b), there is also contradicting evidence (e.g., Connolly et al., 2007; Iverson, 1999; Iverson & Goldin-Meadow, 1997; Shepard & Cooper, 1992). In light of these conflicting results, however, under what circumstances visual experience influences conceptual representations and multimodal language production still remains unknown. One of the aims of the current thesis is to investigate the influence of visual experience (i.e., being sighted or congenitally blind) on conceptual representations and multimodal language production.

Another way to examine the role of visual experience in multimodal language use is to look at to what extent the modality of perceptual input (e.g., visual versus auditory) influences multimodal language use. In fact, research on the influence of visual experience on multimodal language use of sighted individuals is fairly limited. The majority of prior studies focusing on multimodal language have relied heavily on visual modality as their stimulus (e.g., Akhavan et al., 2017; Gullberg et al., 2008; Kita & Özyürek, 2003; Ter Bekke et al., 2022; Ünal et al., 2022). This also suggests existing multimodal language production theories with a perspective based on data derived from a single-modality only may be limited (de Ruiter, 2000, 2007; Hostetter & Alibali, 2008, 2019; Kita, 2000; Kita & Özyürek, 2003; Krauss et al., 2000; McNeill, 1992; McNeill & Duncan, 2000). Considering the nature of our everyday experience and interactions, this creates a serious gap in understanding of multimodal communication. These theories share a common assumption that gestures arise (at least partially) from visuospatial imagery. Although these theories do not explicitly preclude the contribution of non-visual (e.g., auditory) information to gestures, the predominant focus on visual input has resulted in very few studies that explore gestures derived from non-visual information (although see, e.g., Holler et al., 2022; Özçalışkan et al., 2016b, 2018). Thus, another aim of the current thesis is to examine whether perceiving events through single or multiple channels (i.e., vision, audition, or both together) shapes how people encode events in speech and gesture.

In order to bridge some of these gaps, this thesis investigates whether and how multimodal language production and conceptual representations are affected by the modality of perceptual input, with a focus on the productions of congenitally blind people. I address three research questions in this thesis which examine both momentary as well as long-term effects of the perceptual modality of input on object and event concepts and their encoding in multimodal language production:

1) Does the perceptual modality of input (i.e., visual, auditory, or audiovisual) influence the way sighted individuals encode spatial events in speech and gesture? (Chapter 2)
2) Does having access to visual experience (i.e., being sighted or congenitally blind) shape how people encode spatial events in speech and gesture? (Chapter 3)
3) Does having access to visual experience (i.e., being sighted or congenitally blind) affect how concepts are mapped onto silent gestures? (Chapter 4)

In the following sections I critically review existing theories of multimodal language production and studies of language production and comprehension conducted with blind individuals.

Multimodal language production

Human communication typically takes place in face-to-face settings, involving the interactional exchange of multimodal signals (Holler & Levinson, 2019; Perniss, 2018). One of these signals is hand gestures accompanying speech, which play a significant role in cognition and communication (Church et al., 2017; Kendon, 2004; McNeill, 2000). When expressing events, people spontaneously use gestures tightly integrated with speech, both semantically and temporally (McNeill, 1992). Various types of co-speech gestures exist: iconic, deictic, beat, metaphorical, pragmatic, and emblems, and these gestures offer different affordances than speech in terms of representing and organizing event components. For example, iconic gestures are used to depict information about actions (e.g., drinking, writing), motions (e.g., running upstairs, moving closer), object characteristics like size and shape, and spatial relationships between objects (e.g., on, next to). So, they rely on varying degrees of iconic links between their form and meaning conveyed in speech (e.g., Kita et al., 2017), which is less prevalent in speech alone (e.g., Perlman & Cain, 2014; Perniss et al., 2010, but see, e.g., Nielsen & Dingemanse, 2021; Perlman & Lupyan, 2018 for recent evidence suggesting iconic vocalizations are more prevalent than initially thought). On the other hand, deictic or pointing gestures are used to indicate specific concrete or abstract referents. These gestures serve the purpose of guiding attention towards an object or locating an object in gesture space during communication (McNeill, 2000). For the aims of this thesis, I am interested in the use of iconic and, to some extent, deictic co-speech gestures, which are effective tools to think and communicate about spatial information (e.g., Alibali, 2005).

Gesture production theories differ in how they view the relationship between speech and gesture (mostly for iconic gestures), but they have a common assumption that gestures arise from visuospatial imagery (de Ruiter, 2000, 2007; Hostetter & Alibali, 2008, 2019; Kita, 2000; Kita & Özyürek, 2003; Krauss et al., 2000; McNeill, 1992; McNeill & Duncan, 2000). Some theories put the main emphasis on visuospatial and motor imagery (rather than speech) in gesture production (e.g., Gesture as Simulated Action Framework, Hostetter & Alibali, 2008, 2019). According to the interface model (Kita & Özyürek, 2003), however, co-speech gestures derive from the interface between the linguistic conceptualization involved in speech production as well as the visuospatial and motor imagery involved in gesture production. This is based on the fact that speakers tend to focus on different components and convey corresponding aspects of events in speech depending on their language backgrounds (Slobin, 1996; Talmy, 1985). Moreover, the way event components are encoded in gestures differs in ways tightly connected to how they are linguistically encoded in speech (Kita & Özyürek, 2003).

In particular, motion events—consisting of different semantic components—are well-suited to identify cross-linguistic variation in multimodal encoding. Motion events, such as a woman running into the house, are fundamental in our daily lives. They involve the movement of an object, often referred to as the "figure" (in this case, the woman), relative to a reference point or object known as the "ground" (the house). These events encompass both a path or trajectory (into), and a manner in which the movement occurs (running). Languages such as English and Dutch, classified as satellite-framed languages according to Talmy (2000)'s framework, generally convey how an action is performed (manner of motion) through the main verb (e.g., *run*). They use additional elements like prepositional phrases to indicate the path of motion (e.g., *into*). Consequently, in these languages, descriptions of both the path and manner of motion are often combined within a single clause, as in "she ran into the house". In contrast, verb-framed languages such as Turkish, Spanish, Greek, and Japanese primarily express the path of motion through the main verb (e.g., *enter*). Details about how the action is performed, or the manner of motion, are typically found in adverbial phrases or subordinate verbs (e.g., *running*). Therefore, in verb-framed languages, descriptions of path and manner of motion are usually distributed across separate clauses, as in "she entered the house running".

It has been found that gesture patterns of speakers of typologically different languages differ with these linguistic patterns of encoding motion events. Similar to the distinctions in speech, speakers of satellite-framed languages (e.g., English) typically conflate path and manner in a single gesture as they encode path and manner of motion in a single verbal clause in speech. On the other hand, speakers of verb-framed languages (e.g., Turkish and Japanese) typically tend to produce separate gestures for path and manner as they also distribute path and manner information into different clauses in speech (e.g., Kita & Özyürek, 2003). Gestural representations of events reflect the language typology but only when gestures accompany speech (i.e., co-speech gesture). For example, when people are asked to describe motion events only through gesture in the absence of speech (i.e., silent gesture), both Turkish and English speakers combine path and manner into a single gesture, deviating from the expected patterns of Turkish based on its linguistic typology (Özçalışkan et al., 2016a). Over the past two decades, researchers have gathered abundant evidence suggesting that co-speech gestures emerge through the interplay between the linguistic conceptualization involved in speech production and the visuospatial imagery involved in gesture production (Akhavan et al., 2017; Choi & Lantolf, 2008; Chui, 2009; Gullberg et al., 2008; Kita et al., 2017; Özçalışkan et al., 2016b, 2016a; Özyürek et al., 2005; Ter Bekke et al., 2022; Ünal et al., 2022), corroborating the Interface Model (Kita & Özyürek, 2003).

While visuospatial imagery is theorized to be key in all gesture production theories, it is unclear whether the modality of perceptual input influences multimodal language production. Most previous cross-linguistic studies investigating motion event descriptions have predominantly focused on the visual modality as their eliciting stimulus, such as video clips, cartoons, line drawings, and paintings (e.g., Gennari et al., 2002; Gullberg et al., 2008; Papafragou et al., 2002; Ter Bekke et al., 2022, with a few exceptions see Özçalışkan et al., 2016a, 2016b for haptic input). This poses a potential limitation to existing theories that emphasize the role of visuospatial imagery in gesture production since the evidence comes predominantly from visual input (de Ruiter, 2000, 2007; Hostetter & Alibali, 2008, 2019; Kita, 2000; Kita & Özyürek, 2003; Krauss et al., 2000; McNeill, 1992; McNeill & Duncan, 2000).

It is critical to investigate the role of perceptual modality because we receive spatial information via channels beyond vision and each channel has unique affordances. For instance, vision has the advantage of providing simultaneous and constantly accessible information about the characteristics of objects and events whereas audition provides sequential but precise temporal information (Recanzone, 2003; Thinus-Blanc & Gaunet, 1997). Although gesture theories do not explicitly rule out the possible contribution of non-visual spatial information in gestures, only a limited number of studies explore the spatial affordances of non-visual information conveyed through gestures (e.g., Holler et al., 2022). However, it is possible that the perceptual modality of input (e.g., visual versus auditory) may shape how an event is encoded and expressed in multimodal language. For example, it is possible that the sequential format of speech is best suited to express event information perceived through the auditory modality, whereas gestures might best express information from the visual modality—e.g., due to the ease of mapping visuospatial information to gesture, visual input may elicit gestures more frequently than audio input. Or, as vision is widely considered as the primary source for spatial perception (Alais & Burr, 2004; Eimer, 2004; Ernst & Bülthoff, 2004), speakers who receive visual input could provide richer motion event descriptions in both speech and gesture, such as mentioning manner more often, compared to speakers who receive only audio input. Thus, conducting an empirical examination of multimodal event descriptions with different perceptual inputs is necessary to address this gap. In Chapter 2, I present a study that investigated whether perceptual modality of input (vision versus audition) affects event descriptions in speech as well as the frequency and type of gestures accompanying speech.

Multimodal language production of blind people

In addition to manipulating the modality of input, another approach to explore the role of perceptual experience is to examine how sensory loss, for example in congenital blindness, influences multimodal language production for events. Although there is a wide range of cross-linguistic studies on how sighted people describe events multimodally, only a handful of studies have examined blind people's language production for route descriptions and motion events—especially in the context of multimodal language production (Iverson, 1999; Iverson & Goldin-Meadow, 1997; Özçalışkan et al., 2016b, 2018).

Earlier studies found that multimodal expressions of spatial events differ between blind and sighted speakers (Iverson, 1999; Iverson & Goldin-Meadow, 1997). When describing a familiar route in their schools, blind children segmented the path into multiple landmarks in speech, whereas sighted and blindfolded children described paths more holistically, mentioning fewer landmarks but incorporating more gestures alongside their speech. For example, a blind child's description would be: "Turn left, walk north, then you'll see the office, then you'll see 106, then 108, then 110, 112, then there's a doorway. Then there's a hall…" whereas a sighted child might say: "When you get near the staircase, you turn to the left" (Iverson & Goldin-Meadow, 1997, p.463). Furthermore, when children segmented path descriptions in speech, they gestured less often regardless of their visual status. So, it appears that gesture is better-suited to holistic than segmented descriptions, probably because its visual structure is less suited for sequential representation than speech. Because gesture does not need to be structured linearly to the same degree that speech does, it has been characterized as conveying meaning in a more "holistic" way using analog, iconic, and gradual representations (McNeill, 1992; McNeill & Duncan, 2000). Notably, these findings provide support for the notion that spatial cognition in blind people tends to be more sequential compared to sighted people (e.g., Cattaneo & Vecchi, 2011; Iachini et al., 2014; Ruggiero et al., 2021; Vercillo et al., 2018). Thus, visual experience may influence spatial language production via differing underlying spatial cognition. Accordingly, sequential information coming from non-visual input might shape how people express other spatial event components such as manner in speech and co-speech gesture.

More recent investigations into spatial language have, on the other hand, shown similarities in event descriptions between blind and sighted people (Özçalışkan et al., 2016b, 2018). For example, a cross-linguistic study examined motion event descriptions of congenitally blind, sighted, and blindfolded speakers of Turkish and English (Özçalışkan et al., 2016b). Blind and blindfolded speakers explored static scenes depicting motion with figurines—e.g., dolls in various postures indicating running—through touch, while sighted speakers explored them visually. The results showed that speakers' verbal and gestural descriptions of events followed the linguistic patterns of their respective languages regardless of whether speakers were blind, blindfolded, or sighted. These findings suggest that visual experience does not have a substantial impact on how speakers express events through speech or co-speech gestures, with language typology playing a more critical role.

Given the limited and contradictory findings, the role of visual experience in event construal and multimodal language production remains unclear. Some of these differences might arise from the fact that the studies have certain methodological differences which make them difficult to synthesize. For example, while some studies focused on pre-existing spatial representations such as familiar routes (Iverson, 1999; Iverson & Goldin-Meadow, 1997), others relied on novel spatial scenes (Özçalışkan et al., 2016b, 2018). As a result, input modality varied across studies and participants: the studies examining familiar routes did not control the type of input, i.e., how participants learned routes (Iverson, 1999; Iverson & Goldin-Meadow, 1997), and others used different input modalities, i.e., visual input for sighted participants but haptic input for blind and blindfolded participants (Özçalışkan et al., 2016b, 2018).

To address these limitations, I conducted an empirical study with blind, blindfolded, and sighted people where all participants received the same auditory input for everyday motion events. This has the advantage of presenting stimuli that is ecologically relevant for both blind and sighted people, as listening to the sounds of events is a part of both groups' real-world experience. In Chapter 3, I present a study that investigated how blind and sighted individuals encode motion events depicted via sound in speech and co-speech gesture.

The conceptual representations of blind people

In addition to examining the multimodal production of events, looking at blind people's representations of individual object concepts can provide an additional perspective into the role of visual experience in language and cognition. Since the seminal work of Landau and Gleitman (1985), there have been a number of empirical studies of blind people's conceptual representations as a way of understanding the role of perception in shaping concepts more generally. In particular, studies have explored blind people's conceptual representations of colors (Connolly et al., 2007; Kim et al., 2021; Landau & Gleitman, 1985; Marmor, 1978; Saysani et al., 2018; Shepard & Cooper, 1992), other visual properties, such as light emission as expressed in visual verbs (Bedny et al., 2019; Elli et al., 2021; Landau & Gleitman, 1985; Lenci et al., 2013), visual metaphors (Minervino et al., 2018), animals (Kim et al., 2019a), as well as concrete and abstract concepts from diverse semantic categories (Crollen & Collignon, 2020; Lenci et al., 2013; Mahon et al., 2009; Saysani et al., 2021; Striem-Amit et al., 2018). Many of these studies found no differences between blind and sighted people's concepts—for instance, blind children at the age of three demonstrate a comparable understanding of the semantic aspects of vision-related words, such as "look" and "see," to their sighted counterparts (Elli et al., 2021; Landau & Gleitman, 1985). Thus, blind people seem to gain substantial knowledge about concepts through indirect experience derived from language. These results suggest that visual experience has little to no effect on conceptual representations, even for concepts that are primarily related to vision.

However, there is also evidence showing qualitative differences in the conceptual representations of blind and sighted people (Connolly et al., 2007; Kim et al., 2019a; Lenci et al., 2013; Marques, 2010; Shepard & Cooper, 1992). For example, even though blind people have rich knowledge about object colors (e.g., *apples are red*), they seem to assign less importance to this information compared to sighted people when making semantic similarity judgments about fruits and vegetables—e.g., to decide whether an apple is more similar to a strawberry or banana (Connolly et al., 2007). Similarly, Kim et al. (2019a) found that blind people rely to a greater extent on taxonomic knowledge compared to sighted people when making judgments about attributes of animal appearance such as size, shape, and color. In another study asking blind and sighted people to list core features of concepts, blind people overall generated fewer perceptual features for concrete concepts compared to sighted people (Lenci et al., 2013). Together, these studies indicate that certain visual aspects of concepts are less readily accessible to blind people, implying that visual experience can influence conceptual representations to a certain degree.

One important thing to note about previous studies, in addition to the conflicting results, is that most of them have investigated blind people's conceptual representations using language-based measures as a window into concepts—such as semantic judgments based on linguistic stimuli (e.g., Connolly et al., 2007; Kim et al., 2019a, 2021) and feature listing (e.g., Lenci et al., 2013). A novel contribution to this literature could be that gesture offers another way to explore conceptual representations, as it is now well accepted that co-speech and silent gestures can also provide insights into conceptual representations. Multimodal language theories with an embodied perspective claim that gestures are generated through sensorimotor simulations which are critical for conceptual representation (Goldin-Meadow & Beilock, 2010; Hostetter & Alibali, 2008, 2019). Consequently, the forms of gestures reflect the gesturers' specific sensorimotor experiences with objects and events (Beilock & Goldin-Meadow, 2010; Cook & Tanenhaus, 2009; Pouw et al., 2020).

Earlier studies examining how people depict individual concepts in co-speech (Masson-Carro et al., 2016, 2017) or silent gesture (e.g., Ortega & Özyürek, 2020b, 2020a; van Nispen et al., 2017) have revealed certain regularities in the gestural strategies used across communities demonstrating a link between gestural representations and sensorimotor experiences. For example, concepts associated with motor experience, such as manipulable objects, like tools, often lead to the use of an acting strategy, which involves reenacting a physical action. Conversely, when it comes to concepts with limited manipulability, such as objects and animals, sighted individuals tend to use either a drawing strategy (i.e., tracing the object's outline) or a personification strategy (i.e., embodying the movement of a non-human entity using their body), respectively.

If gesture forms depicting concepts emerge from sensorimotor experience, it is reasonable to assume that motor experience contributes to the use of the acting strategy, while visual experience plays a role in the use of drawing and personification strategies. These regularities found in the sighted population suggest that gestures can be used to investigate the extent to which experience shapes conceptual representations. If conceptual representations of blind and sighted people are different, it is expected that gestures would also differ between blind and sighted people. To address this, in Chapter 4 I present an empirical study that examined blind and sighted peoples' gestures of object concepts from different semantic categories as well as feature-listing of the same concepts as an ancillary language-based measure.

Thesis summary

In this thesis, I intend to bridge the fields of multimodal language production and conceptual representations studied from the perspective of blind people who have a different embodied experience of the world. This will enhance our understanding of the interplay between language, perception, and cognition. The current thesis examines whether and how perceptual experience influences conceptual representations as well as multimodal language expressed in speech and gesture, with a focus on multimodal language production of congenitally blind people.

To do so, this thesis focuses on the analysis of speech, co-speech gesture, and silent gesture of congenitally blind, blindfolded, and sighted Turkish speakers. Focusing on Turkish, classified as a verb-framed language by Talmy (2000)'s framework, means this work can be contextualized within the context of prior research on motion events conducted in this language (e.g., Aktan-Erciyes et al., 2022; Allen et al., 2007; Kita & Özyürek, 2003; Özçalışkan et al., 2016b, 2018; Ter Bekke et al., 2022). These studies collectively offer a valuable corrective to the prevalence of research conducted in the English language, as noted in the literature (Blasi et al., 2022; Thalmayer et al., 2021).

In the current thesis, I will present three empirical studies, which together examine whether perceptual experience influences multimodal language production and conceptual representations. More specifically, I investigate three research questions:

1) Does the perceptual modality of input (i.e., visual, auditory, or audiovisual) influence the way sighted individuals encode spatial events in speech and gesture? (Chapter 2)
2) Does having access to visual experience (i.e., being sighted or congenitally blind) shape how people encode spatial events in speech and gesture? (Chapter 3)

3) Does having access to visual experience (i.e., being sighted or congenitally blind) affect how concepts are mapped onto silent gestures? (Chapter 4)

The following summary outlines the structure of each chapter in addressing these research questions.

Chapter 2 investigates whether the perceptual modality of input (vision, audition) influences event descriptions in speech and co-speech gesture. I build on earlier work on the multimodal description of motion events together with theoretical accounts of multimodal language production. This literature shows that co-speech gestures derive from the interface between the conceptualization involved in speech production and the visuospatial imagery involved in gesture production. Here I take a next step by examining the role of the perceptual modality of input in multimodal event descriptions. This approach poses some challenges to existing theories of gesture production, which typically focus on visual input alone and do not consider the spatial affordances of non-visual information. Therefore, in this chapter I discuss the spatial affordances of different perceptual modalities and why it is critical to study multimodal language using non-visual perceptual stimuli too. With this aim in mind, I focus on motion events as there is a large body of speech and gesture production studies to build upon (as summarized briefly in the ***Multimodal language production*** section).

I conducted a comparison of Turkish sighted speakers' speech and co-speech gesture for path (i.e., trajectory of the movement) and manner (i.e., how the movement is performed) of motion events presented as audio-only, visual-only, or multimodal (visual + audio) input. Although my primary objective was to make a comparison between audio-only and visual-only input, I included a multimodal input condition to explore any potential boost that multiple types of information may offer. Thus, I compared the visual-only condition to the multimodal condition to determine if auditory information contributes additional spatial information to language production. There were a number of specific predictions I made based on the affordances of the input modalities as well as multimodal encoding of motion events. First, I predicted that if vision plays a critical role in the linguistic encoding of motion events, speakers in the visual conditions (i.e., visual-only and multimodal conditions) would produce a greater number of motion event descriptions compared to speakers in the audio-only condition. Second, I predicted more mentions of path in speech in the audio-only condition than in the visual conditions, consistent with the proposal that there should be more segmented event descriptions due to auditory input providing more sequential information. Third, I predicted fewer mentions of manner in speech in the audio-only condition than in the visual conditions, given the possibility that vision provides richer information about manner than audition. Lastly, regarding co-speech gestures, I predicted that due to the affordances of the visual modality and the ease of mapping visuospatial information from vision to gesture, visual input might be more suited to elicit gestures than audio input, leading to more path and manner gestures in the visual conditions than the auditory condition.

Chapter 3 investigates whether visual experience shapes how people encode different components of motion events for multimodal language production by comparing people with congenital blindness to sighted counterparts. Here I go beyond the approach used in Chapter 2 by examining the impact of a lifetime of experience of being blind on language use. I discuss the conflicting results of prior studies examining the multimodal language production of blind people (as summarized in the section, ***Multimodal language production of blind people***). To mitigate the limitations and address the gaps in prior research (e.g., not controlling for and equating the type of input at event encoding), I equated the input modality and presented motion events auditorily to all speakers (e.g., hearing footsteps of a person walking into a room). So, this chapter provides a systematic and experimentally controlled investigation of the role of visual experience on language production by using stimuli ecologically-relevant to blind and sighted people.

I compared verbal and spontaneous gestural descriptions of Turkish blind, blindfolded, and sighted speakers for path and manner of motion events. In addition, I focused on how blind and sighted people refer to landmarks (i.e., source and goal of events) in their speech and gesture. Evidence from spatial cognition research, which investigates how congenitally blind people learn spatial layouts, indicates that blind people tend to use an egocentric frame of reference (i.e., relating locations to their own position in space) rather than an allocentric one, which is based on external objects regardless of their relative positions (e.g., Cattaneo & Vecchi, 2011; Iachini et al., 2014). Thus, I had specific predictions concerning speech and gesture based on the spatial language and spatial cognition literature. First, I predicted that if blind people rely more on egocentric than allocentric frame of reference to encode spatial information, as found in previous spatial cognition studies, blind people would describe spatial locations more in relation to their own position in space. So, blind speakers would use more self-anchored landmarks (i.e., referring to the speaker's body such as to/from my left) than blindfolded and sighted speakers. Furthermore, as blind people are often better than sighted people at localizing sounds, as shown in previous spatial cognition studies, I predicted blind speakers would produce more pointing gestures than non-blind speakers. Thus, differences in spatial cognition due to blindness might influence descriptions in speech and gesture.

With regard to path and manner encoding, if non-visual input leads to more sequential encoding of path in speech (as also shown by the results of Chapter 2), I predicted that blind speakers would produce more path descriptions in speech than non-blind (both blindfolded and sighted) speakers. And, if vision provides richer information about manner than audition, I predicted that blind speakers would produce fewer manner descriptions in speech than non-blind speakers. Regarding path and manner encoding in co-speech gesture, if gesture mainly derives from visuospatial imagery, I predicted fewer spontaneous iconic gestures among blind than sighted and blindfolded speakers for both path and manner in motion event descriptions. These results would have implications for gesture production theories which claim that gestures are partially driven from visuospatial imagery. One the other hand, if multimodal language production is driven mostly from abstract linguistic representations, I would not predict differences between sighted and blind individuals.

Chapter 4 investigates how object concepts are mapped onto gestures in both congenitally blind and sighted individuals. As such, this chapter examines the influence of visual experience in conceptual representations by focusing on object concepts rather than event components. This chapter builds on earlier work on conceptual representations in blind people. Previous studies have commonly studied conceptual representations via language-based measures such as semantic judgements and feature listing (as summarized in ***The conceptual representations of blind people***). Here, I propose a different approach to investigate conceptual representations using the production of silent gestures (i.e., gesture in the absence of speech) as an alternative method to traditional language-based measures. As gestures are driven by iconic mappings of visuospatial and motoric experience, they can reveal some aspects of concepts at a finer-grained level than language-based measures.

I compared gesture strategies used by blind and sighted people for concepts that rely on visual (i.e., non-manipulable objects and animals) vs. motor (i.e., manipulable objects) information to different extents. Participants were presented with pre-recorded spoken concepts and asked to produce silent gestures to convey individual concepts. I analyzed the different gesture strategies participants used to depict concepts. Additionally, I collected feature-listing for the same concepts, as an auxiliary language-based measure to explore to what extent the results from both measures are comparable. The features listed by participants were coded as perceptual (i.e., information gained through a primary sensory channel such as size, shape, appearance, sound, body parts, and kinematic information) or non-perceptual (e.g., functional, taxonomic, or encyclopedic). By using both measures, I could triangulate the degree of consistency between the results obtained from silent gestures and feature-listing. Overall, I predicted an interaction between visual experience and semantic category on the frequency of gestures using the drawing, personification, acting, and representing strategies, as well as for the perceptual and non-perceptual features listed. More specifically, if visuospatial and motor cues drive gesture, I first predicted fewer gestures in the blind than sighted group for concepts that rely more on visual (i.e., non-manipulable objects and animals) than motor (i.e., manipulable objects) information. Second, I expected fewer gestures with the drawing strategy for non-manipulable object concepts and fewer gestures with the personification strategy for animal concepts in blind compared to sighted participants as these two strategies rely on visual experience. Third, as the acting strategy relies on motor experience, I did not predict a difference between blind and sighted participants. As the representing strategy is rarely observed among sighted people in earlier studies, I did not have a specific prediction for this strategy. Regarding the feature listing, based on earlier work, I expected fewer perceptual and more non-perceptual features of concepts reported by blind compared to sighted participants. This chapter offers new perspectives in the conceptual representations of blind individuals and the field of embodied cognition.

Finally, in **Chapter 5**, I synthesize the findings of the three experimental studies described in Chapters 2-4 of this thesis. Based on the insights I have gained, I discuss the findings in light of the theoretical implications for multimodal language production, gesture production, and embodied cognition theories. I then discuss the methodological contributions and limitations of the work, as well as directions for future research.

2

The effect of input sensory modality on the multimodal encoding of motion events

This chapter is based on:
Mamus, E., Speed, L. J., Özyürek, A., & Majid, A. (2023). The effect of input sensory modality on the multimodal encoding of motion events. *Language, Cognition and Neuroscience, 38*(5), 711-723. https://doi.org/10.1080/23273798.2022.2141282

Abstract

Each sensory modality has different affordances: vision has higher spatial acuity than audition, whereas audition has better temporal acuity. This may have consequences for the encoding of events and its subsequent multimodal language production—an issue that has received relatively little attention to date. In this study, we compared motion events presented as audio-only, visual-only, or multimodal (visual + audio) input and measured speech and co-speech gesture depicting path and manner of motion in Turkish. Input modality affected speech production. Speakers with audio-only input produced more path descriptions and fewer manner descriptions in speech compared to speakers who received visual input. In contrast, the type and frequency of gestures did not change across conditions. Path-only gestures dominated throughout. Our results suggest that while speech is more susceptible to auditory vs. visual input in encoding aspects of motion events, gesture is less sensitive to such differences.

Introduction

We usually receive spatial information via multiple channels. For example, while seeing someone walking away, we may also hear the fading sound of footsteps echoing in the corridor. Each sensory modality has different affordances that contribute to our overall experience of an event. At the same time, we can express events in language using different modalities, as in the verbal and manual modalities, each of which has its own channel restrictions. It is possible, therefore, that the expressibility of multisensory events into multimodal language may differ according to the constraints of both input and output channels. To test this, we investigate whether perceiving events through vision or audition influences the way we express spatial events in speech and gesture.

Vision has the unique advantage of providing simultaneous (i.e., holistic) information about features of objects and events in both close and distant space (e.g., Eimer, 2004; Thinus-Blanc & Gaunet, 1997). It is continuously accessible and thus allows perceivers to update information about motion, location, and spatial relations. Like vision, audition is a distant sense, however, it provides better temporal information than vision across locations. Audition is found to dominate in temporal processing, such as discriminating rhythmic changes (e.g., Recanzone, 2003; Repp & Penel, 2002; Shams et al., 2000; Spence & Squire, 2003), and in contrast to the holistic nature of visual information, auditory information is sequential. Even though audition provides information about objects and events, vision typically dominates over conflicting auditory information in spatial perception (e.g., Alais & Burr, 2004; Howard & Templeton, 1966). Therefore, vision is widely considered the primary source of spatial perception (e.g., Ernst & Bülthoff, 2004; Welch & Warren, 1980).

It has been claimed that language reflects this asymmetry between vision and audition. Vision appears to have privileged status, especially in languages of Western societies (e.g., Levinson & Majid, 2014; Lynott et al., 2020; Majid et al., 2018; San Roque et al., 2015; Speed & Majid, 2017; Viberg, 1983). This is reflected in the fact that vision-related verbs (e.g., *see, look*) are more frequent and numerous than non-vision related verbs (e.g., *smell*, *feel*) in the perceptual lexicons of languages of the world (e.g., Floyd et al., 2018; Lynott et al., 2020; San Roque et al., 2015; Speed & Majid, 2017; Winter et al., 2018). Although we see differences in the number and frequency of words across the senses, no study has experimentally investigated the role of input modality on the language used to describe events. Moreover, there is little known about its multimodal expression, particularly co-speech gesture.

From first principles, one might speculate the sequential format of speech is best suited to express event information perceived through the auditory modality, while gesture might best express information from the visual modality. Gesture production theories do indeed share an assumption that gesture derives from visuospatial imagery (Sketch Model, de Ruiter, 2000; Postcard Model, de Ruiter 2007; Gesture as Simulated

Action Framework, Hostetter & Alibali, 2008, 2019; Information Packaging Hypothesis, Kita, 2000; Interface Model, Kita & Özyürek, 2003; Lexical Retrieval Hypothesis, Krauss, Chen, & Gottesman, 2000; Growth Point Theory, McNeill, 1992; McNeill & Duncan, 2000), with iconic gestures in particular considered an effective tool to convey visuospatial information (Alibali, 2005; Hostetter & Alibali, 2008, 2019). While there is nothing in these theories precluding the expression of auditory information in gesture, the emphasis on the "visual" has meant there are very few studies that have investigated the spatial affordances derived from non-visual information and expressed through gesture (although see, e.g., Holler et al., 2022).

To be able to address the question of whether input sensory modality affects multimodal language production, it is important to situate this work in the broader study of motion events and language typology. This is important as speakers of different languages package the same spatial experience in different ways focusing on, and conversely omitting, certain event components in speech and gesture. Slobin (1996) proposed that speakers encode aspects of events depending on distinctions in their language. For example, unlike a satellite-framed language such as English, Turkish is considered a verb-framed language, which primarily encodes path in the main verb and optionally encodes manner in a subordinated verb or adverbial phrases (Talmy, 1985). Turkish speakers use path and manner in separate clauses (e.g., *koşarak eve girdi 'she entered the house running'*, see Table 1), whereas English speakers conflate these in a single clause (e.g., *she ran into the house*) with manner as the main verb.

These language-specific patterns in speech are also reflected in co-speech gesture (Kita, 2000; Kita & Özyürek, 2003; Özçalışkan et al., 2016b, 2018). Turkish speakers gesture path and manner separately, whereas English speakers are more likely to produce conflated gestures. In addition, given the focus on path in verb-framed languages, Turkish speakers have a tendency to gesture only about path, even in cases where they mention both path and manner in speech (Özyürek et al., 2005; Ünal et al., 2022; for a similar tendency in Farsi, Mandarin Chinese, and French respectively, see also Akhavan et al., 2017; Chui, 2009; Gullberg et al., 2008). To account for this, Kita and Özyürek (2003) proposed that gesture derives partly from language typology and partly from visuospatial imagery in their interface model.

With respect to our main research question concerning the role of input modality on the expressibility of motion events, most previous studies have relied overwhelmingly on visual stimuli as input (e.g., video-clips, cartoons, line drawings, paintings; Gennari et al., 2002; Gullberg et al., 2008; Papafragou et al., 2002; Slobin et al., 2014; Ter Bekke et al., 2022). A notable exception is the work of Özçalışkan et al. (2016b) who examined cross-linguistic differences in motion event descriptions in congenitally blind, sighted, and blindfolded speakers of Turkish and English. In order to elicit descriptions, blind and blindfolded participants explored scenes haptically while sighted speakers explored them visually. Scenes consisted of landmark objects (e.g., toy house, crib), where static dolls in different postures were posed to create the impression of motion (e.g., a girl running into a house). All participants were instructed to describe the scenes and were explicitly encouraged to gesture at the same time. Özçalışkan et al. (2016b) found that both blind and sighted speakers (blindfolded or not) of Turkish and English expressed events in speech and co-speech gesture according to the typology of their language. In a follow-up study, (Özçalışkan et al., 2018) showed that blind and sighted speakers of Turkish and English do not display typological differences in gesture when produced without speech (i.e., silent gesture), in line with the claim that only co-speech gesture reflects language-specific packaging (Goldin-Meadow et al., 2008).

These findings suggest sensory modality (in this case, visual vs. haptic) does not strongly influence the way speakers express events in speech or co-speech gesture, with language typology playing a more critical role. However, this conclusion may be premature. While Özçalışkan et al. (2016b, 2018) developed a clever paradigm to compare people with and without visual access to stimuli, the conditions were not controlled in all respects. People could have spent longer exploring haptic scenes than visual ones, and this could have affected descriptions. Moreover, there was no direct comparison between descriptions of blindfolded and sighted speakers, so it is possible that within language there were differences between visual and haptic conditions. Finally, in both Özçalışkan et al. (2016b, 2018) speakers were explicitly asked to gesture while describing events. Encouraging gesturing might affect the encoding of events and possibly increased speakers' gesture frequency (e.g., Cravotta et al., 2019). Therefore, it remains unclear whether sensory modality of input affects the rate and type of spontaneous gesture production.

There is, in fact, evidence that sensory input could affect multimodal language production for spatial scenes (Iverson, 1999; Iverson & Goldin-Meadow, 1997), which in turn could have implications for motion event encoding. Iverson and Goldin-Meadow (1997), for example, compared blind and sighted English speakers during a route description task and found blind children described path in a more segmented fashion with more landmarks in their speech than sighted children. For example, a blind child described a route description as: *"Turn left, walk north, then you'll see the office, then you'll see 106, then 108, then 110, 112, then there's a doorway. Then there's a hall..."*, whereas a sighted child said: *"when you get near the staircase you turn to the left"* (p.463). Interestingly, when children gave segmented verbal descriptions, regardless of their visual status, they produced fewer gestures. Iverson and Goldin-Meadow (1997) claimed that gesture frequency decreases with segmented speech due to the process of gesture generation. As gestures express an image as "a global whole" (McNeill, 1992), when speech is represented sequentially, it is not as well-suited for gesture. So, while speech might be more suitable for expressing information from non-visual input, gesture might be less well suited to do so.

To summarize, previous studies provide contradictory evidence about whether sensory modality could influence the way information is expressed in speech and gesture (Iverson, 1999; Iverson & Goldin-Meadow, 1997; Özçalışkan et al., 2016b, 2018). However, no study has directly varied the input sensory modality of motion events—while also controlling for duration and event type—to test whether it affects speech and gesture.

The present study

We explore the effect of sensory modality of input on multimodal language use by focusing on motion events. Motion events provide a good test bed as there is a large body of previous speech and gesture production studies to build upon (Akhavan et al., 2017; Brown & Chen, 2013; Chui, 2009; Gennari et al., 2002; Gullberg et al., 2008; Papafragou et al., 2002). Importantly, path and manner components of motion events can be perceived from both visual and auditory inputs (Geangu et al., 2021; Mamus et al., 2019) and each may be differentially mapped to speech and gesture. Focusing on Turkish in particular allows us to situate our results with respect to previous studies in this language (e.g., Aktan-Erciyes et al., 2022; Allen et al., 2007; Kita & Özyürek, 2003; Özçalışkan et al., 2016b, 2018; Ter Bekke et al., 2022) which together provide an important corrective to the dominance of English language studies in the literature (cf. Thalmayer et al., 2021).

We compared Turkish speakers' speech and gesture for path and manner of motion events that were presented as audio-only, visual-only, or multimodal (visual + audio) input. Our main goal was to compare audio-only to visual-only input. Including a multimodal condition allowed us to examine the additional boost, if any, multiple sources of information provide. In particular, it is interesting to compare the visual-only to the multimodal condition to see if auditory information provides additional spatial information to language production processes.

In speech, there are a number of specific predictions we can make. First, based on the observation that vision dominates in the perceptual lexicons of languages (e.g., San Roque et al., 2015; Winter et al., 2018), it is possible that vision also influences linguistic encoding for motion events. If so, we would predict that participants in the visual conditions (i.e., visual-only and multimodal conditions) would provide more motion event descriptions than participants in the audio-only condition.

In addition, we can make specific predictions about the encoding of path vs. manner in speech. With regard to path, if the previously attested differences in encoding of path information from non-visual input (i.e., segmented path descriptions in blind vs. non-blind , Iverson, 1999; Iverson & Goldin-Meadow, 1997) are caused by the sensory modality of input at encoding, we would predict that participants in the audio-only conditions would describe path of motion in a more segmented fashion than in other conditions because auditory input is more sequential. This would lead to more mentions of path within each description in speech in the audio-only condition than in the visual conditions.

As for manner, it is possible that vision is of advantage here too. For example, in order to differentiate particular manners, such as walk vs. run, vision provides richer information than audition about biomechanical properties (e.g., Malt et al., 2014), as well as providing information about speed and direction of motion. So, participants in the visual conditions might describe manner more often than participants in the audio-only conditions. On the other hand, audition is also good at providing temporal information—such as rhythm of motion (e.g., Recanzone, 2003; Repp & Penel, 2002), so it is also possible that auditory information might be as rich as visual information and lead to a comparable manner encoding of motion.

Regarding co-speech gesture, there are two main possibilities that can be predicted from the previous literature, either visual input is also advantageous for gesture or there is no impact of modality on gesture production. There are three reasons to expect gesture frequency for manner and path gestures would be higher for visual conditions than the auditory condition. First, gestures—due to the affordances of the visual modality and the possibilities of more easily mapping visuospatial information from vision to gesture—might be more suited for expressing visual information than auditory information (Macuch Silva et al., 2020). For example, signing children use more manner and path expressions in Turkish sign language than their Turkish speaking peers because of the visually motivated linguistic forms available to sign languages (Sümer & Özyürek, 2022). Second, one might expect gesture to parallel speech patterns (e.g., Kita & Özyürek, 2003; Özyürek et al., 2005), thus leading to more manner gestures in the visual conditions than the auditory condition. Finally, path gestures might be more difficult to produce with the segmented speech predicted for the auditory condition because gestures are less suited for segmented expressions (Iverson, 1999; Iverson & Goldin-Meadow, 1997), leading to higher rates of path gestures in the visual conditions. For all these reasons, visual input may be particularly suited to elicit gestures.

On the other hand, it is possible that there is no difference in the frequency of gesture production between different input conditions. Gesture production theories focusing on the role of mental imagery in gesture, such as the GSA framework, have suggested that "*any form of imagery* [such as auditory or tactile imagery] *that evokes action simulation is likely to be manifested in gesture*" (Hostetter & Alibali, 2019, p.726). This suggests the type of input does not matter for how much gesture is elicited, as long as spatial imagery can be generated. Thus, on this account, participants in all conditions could produce comparable gestures.

Method

Participants

We recruited 90 native Turkish speakers with normal or corrected-to-normal vision from Boğaziçi University. We randomly assigned 30 participants to each of three conditions: audio-only (M = 21 years, SD = 2, 17 female), visual-only (M = 22 years, SD = 3, 16 female, 1 nonbinary), and multimodal (M = 21 years, SD = 2, 10 female, 2 nonbinary). We tested participants in a quiet room on Boğaziçi University campus. They all received extra credit in a psychology course for their participation and provided written informed consent in accordance with the guidelines approved by the IRB committees of Boğaziçi and Radboud Universities.

Stimuli

We made video- and audio-recordings of locomotion and non-locomotion events with an actress. We created 12 locomotion events by crossing 3 manners (walk, run, and limp) with 4 paths (to, from, into, and out of) in relation to a landmark object (door or elevator)—such as "*someone runs into an elevator*". So, participants either only listened to the sound of someone running into an elevator or watched the event with or without the sound. A video and audio recorder were placed next to the landmark objects. For *to* and *into* events, the actress moved towards landmarks, with the path direction approaching the audio recorder. For *from* and *out of* events, the actress moved away from landmarks, with the path direction away from the audio recorder.

We created 12 non-locomotion events with the same actress performing two-participant "transitive" actions with different objects (e.g., cutting paper, eating an apple), and the video and audio were recorded across from her at a fixed distance. Locomotion events served as the critical items, whereas non-locomotion events were included as fillers. Thus, we did not investigate the non-locomotion events.

There were 24 trials per person, including a total of 12 locomotion (M_{duration} = 11.3s, $SD_{duration}$ = 3.6) and 12 non-locomotion (M_{duration} = 7.7s, $SD_{duration}$ = 2.3) events presented in different random orders across participants (see Appendix I as supplementary data for a list of all events and their durations). All stimuli are also available at https://osf.io/qe7dz/.

The experiment used a between-subjects design with three levels of input modality (audio-only vs. visual-only vs. multimodal).

Procedure

Using a laptop and Presentation Software (Version 20.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com), events were presented as audio-clips to participants in the audio-only condition, as silent video-clips to participants in the visual-only condition, and as video + audio clips to participants in the multimodal condition. All participants regardless of the condition wore headphones during the task. The instructions were the same across the conditions except the opening sentence (i.e., *in this task, you will "watch video clips" / "listen to sound clips"*). Participants were then asked to describe each event at their own pace without any instructions about gesture use. They were told other participants would watch their descriptions and watch/listen to the same events in order to match descriptions with events.

At the beginning of the experiment, participants performed two practice trials with non-locomotion events. Further clarification was provided, if necessary, after the practice trials. Event descriptions were recorded with a video camera that was approximately 1.5m across from participants. The experimenter sat across from participants and next to the camera. After each event description, participants proceeded with the next trial at their own pace by pressing a button on the laptop. Participants also filled out a demographic questionnaire on another laptop after the event description task. The experiment lasted around 15 minutes.

Coding

Speech

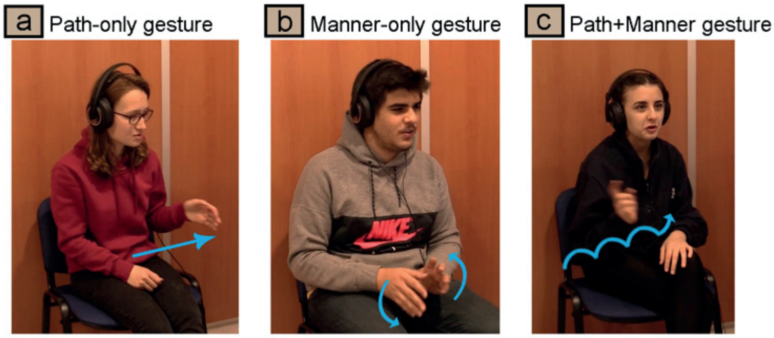
All descriptions of locomotion and non-locomotion events were annotated by two native Turkish speakers using ELAN (Wittenburg et al., 2006), but only descriptions for the locomotion events were transcribed and coded. These descriptions were then split into clauses. A clause was defined as a verb and its associated arguments or a verb with gerund phrases. Clauses including locomotion descriptions (e.g., *someone is walking towards the door*) were coded as relevant, whereas clauses including a transitive event—such as *opening a door* or *ringing the bell*—or other information—such as *a person is wearing high heels*—were coded as irrelevant to the target event. Each relevant clause was coded according to the type of information it contained: (a) path (trajectory of motion), and (b) manner (how the action is performed)—see Table 1 for an example. We calculated the Interclass Correlation Coefficient (ICC) between two coders to measure the strength of inter-coder agreement for path and manner of motion in speech (Koo & Li, 2016). Agreement between coders was .97 for path and .94 for manner of motion.

Table 1. An example of coding a description

	Clause 1			Clause 2		
Turkish description	*koş*	*-arak*	*ev*	*-e*	*gir*	*-di*
Glossing	run	connective	house	dative	enter	past
Coding	manner				path	
English translation	'while running'			'(someone) entered the house'		

Gesture

Participants' spontaneous iconic gestures were coded for each target motion event description. We coded gesture strokes (i.e., the meaningful phase of a gesture) that co-occurred with descriptions (Kita, 2000). Each continuous instance of hand movement



SPEECH

Cl.1 *Asansörden çıktı*
'[someone] exited from the elevator'

Cl.2 *ve binanın içinde aksayarak*
'and in the building limping'

Cl.3 *yürüyor*
'walks'

Cl.1 *Asansörden inip*
'[someone] getting off the elevator'

Cl.2 *böyle aksayarak*
'limping like this'

Cl.3 *koridorun sonuna doğru yürüyor*
'walks towards the end of the corridor'

Cl.1 *Bir kadın koridor boyunca yürüyerek*
'a woman walking along the corridor'

Cl.2 *binadan çıkıyor*
'exited from the building'

Figure 1. Example gestures depicting (a) path only, (b) manner only, and (c) both path and manner. The full event descriptions are split into clauses (Cl.) and translations are given under each gesture example. The gesture stroke occurred during the underlined speech.

was coded as a single gesture. Iconic gesture representing trajectory and/or manner of motion were further classified into the following categories (see Figure 1 for gesture examples):

(a) ***path-only*** gestures depict trajectory of movement without representing manner
(b) ***manner-only*** gestures show the style of movement without representing trajectory
(c) ***path + manner*** gestures depict both trajectory and manner of movement simultaneously

We calculated the ICC between two coders to measure the strength of inter-coder agreement for identifying a gesture and coding each type of gesture. Agreement between coders was .98 for identifying gestures and between .92–.95 for type of gesture (i.e., .95 for coding path only, .92 for manner only, and .95 for path + manner gestures).

Results

To analyse the data, we used linear mixed-effects regression models (Baayen et al., 2008) with random intercepts for participants, items, path type, and manner type, using the packages lme4 (Version 1.1–28; Bates et al., 2015) with the optimizer *nloptwrap* and lmerTest (Version 3.1–3; Kuznetsova et al., 2017) to retrieve *p*-values in R (Version 4.1.3; R Core Team, 2022). We conducted linear mixed effects models on distinct motion elements (path and manner) in speech and gesture. To assess statistical significance of the fixed factors and their interaction, we used likelihood-ratio tests with χ^2, comparing models with and without the factors and interaction of interest. For post-hoc comparisons and to follow-up interactions, we used *emmeans* (Version 1.7.3; Lenth, 2022). Data and analysis code are available at https://osf.io/qe7dz/.

Speech

Overall differences in the amount of speech produced for visual and auditory motion events

First, we tested whether participants differed in the speech they produced for motion events based on audio-only, visual-only, or multimodal input. Table 2 provides the descriptive statistics for the average number of all clauses, motion event clauses, all gestures, and relevant gestures.

Table 2. The average number (M) of clauses and gestures across participants with standard deviations (SD, in parentheses)

Group	All clauses M (SD)	Motion event clauses M (SD)	All gestures* M (SD)	Relevant gestures M (SD)
Audio-only	35.83 (12.77)	20.33 (5.25)	12.03 (6.61)	9.57 (5.40)
Visual-only	32.83 (9.48)	25.53 (5.43)	11.13 (8.86)	8.83 (7.91)
Multimodal	32.40 (8.41)	25.27 (4.65)	11.00 (8.38)	9.07 (7.28)

*All iconic gestures (relevant or irrelevant) produced within a motion event clause

We ran a glmer model with the fixed effect of input modality (audio-only, visual-only, multimodal), the fixed effect of manner type (walk, run, limp), and their interaction term on binary values for mention of motion event clauses in speech (0 = no, 1 = yes) as a dependent variable. See Appendix II as supplementary data for the model summary table. It revealed an effect of input modality, χ^2 (2) = 42.43, $p < .001$, $R^2 = .042$. Participants in the audio-only condition had fewer motion event descriptions compared to participants both in the visual-only (β = -1.07, SE = .170, z = -6.32, $p < .001$, $R^2 = .031$) and multimodal (β = -1.07, SE = .170, z = -6.29, $p < .001$, $R^2 = .031$) conditions. There was no difference between participants in the visual-only and multimodal conditions, (β = .006, SE = .178, z = .032, p = .99). Figure 2 shows the ratio of motion event descriptions (i.e., clauses including locomotion descriptions) in all descriptions.

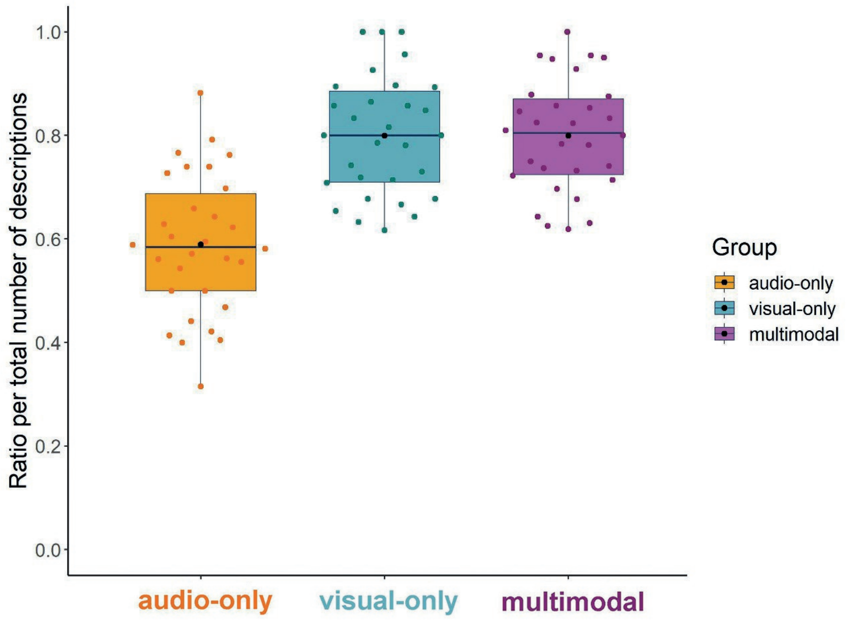


Figure 2. Ratio of motion event descriptions. Coloured dots represent the data for each participant. Black dots represent the group mean.

The model also revealed an effect of manner type, χ^2 (2) = 7.77, p = .021, R^2 = .002. Participants had more motion event descriptions for the run than limp events (β = 0.29, SE = .102, z = 2.83, p = .013, R^2 = .002). But, there was no difference between the walk and limp events (β = 0.09, SE = .102, z = .91, p = .63) and the run and walk events (β = 0.20, SE = .107, z = 1.82, p = .16) in terms of the motion event descriptions. The model did not reveal a significant interaction between input modality and manner type, χ^2 (2) = 9.46, p = .051.

Differences in reference to path and manner in speech

Next, we examined whether participants differed in how much they expressed path and manner in speech. To account for baseline differences in the number of motion event descriptions produced, we calculated the ratio of mention of path and manner per motion event description for each participant and item. We ran a lmer model with the fixed factors of input modality (audio-only, visual- only, multimodal) and type of description (path vs. manner) and their interaction term using the ratio of mention of path and manner per motion event description as the dependent variable (see Figure 3). The model revealed no fixed effect of input modality, χ^2 (2) = 1.37, p = .50, but a fixed effect of type of description, χ^2 (1) = 15.95, p < .001, R^2 = .008, showing that manner was mentioned more than path in speech. However, the model also revealed an interaction between input modality and type of description, χ^2 (2) = 31.25, p < .001, R^2 = .023. To follow-up the interaction, we first used *emmeans* function to compare the use of path vs. manner within each group. There was more mention of manner than path in the

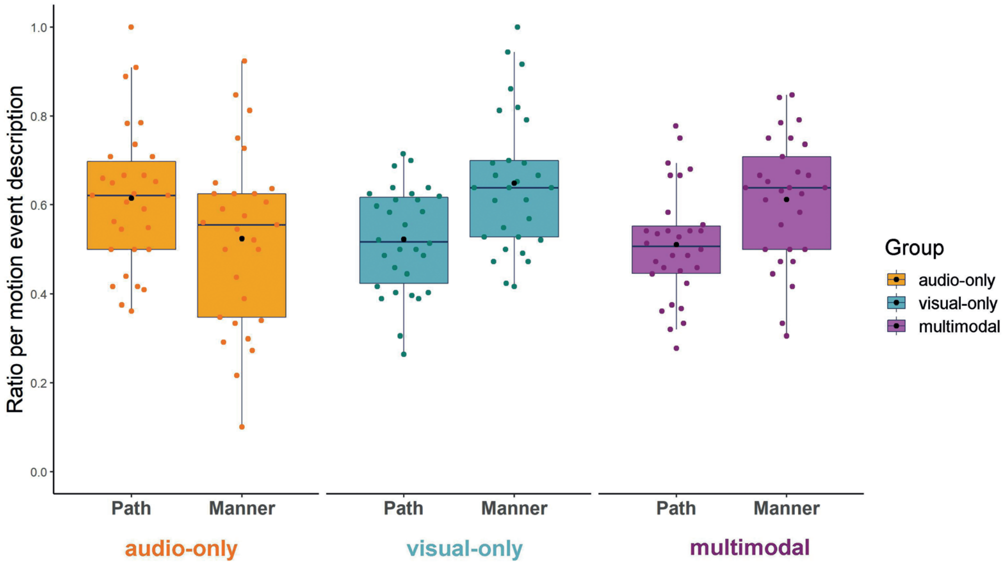


Figure 3. Path and manner in speech. Coloured dots represent the average data for each participant. Black dots represent the group mean.

visual-only (β = .141, SE = .028, t = 5.03, p < .001) and multimodal conditions (β = .115, SE = .028, t = 4.11, p < .001), but more reference to path than manner in the audio-only condition, β = .068, SE = .029, t = 2.33, p = .020. That is, manner and path were differentially salient in the visual versus auditory conditions.

Second, to follow-up the interaction effect, we also compared reference to manner and path separately across input modalities. Path was mentioned more often in the audio-only than visual-only (β = .090, SE = .029, t = 3.15, p = .005) and multimodal (β = .101, SE = .029, t = 3.51, p = .002) conditions. Conversely, manner was mentioned less often in the audio-only than visual-only (β = -.12, SE = .029, t = -4.15, p < .001) and multimodal (β = -.08, SE = .029, t = -2.89, p = .011) conditions. There was no difference between the visual-only and multimodal conditions for references to path (β = .010, SE = .028, t = 0.36, p = .93) or manner (β = .036, SE = .028, t = 1.29, p = .41). See Appendix III as supplementary data for the summary of post-hoc comparisons with *emmeans*.

Gesture

Overall differences in the amount of gesture produced for visual and auditory motion events

We investigated whether participants differed in how much they gestured about different elements of motion events based on input modality (see Table 2 for the descriptive statistics). Because the amount of gesture changes as a function of the rate of motion event descriptions, we first calculated the gesture ratio per motion event description. We compared the groups in terms of their overall gesture ratio using a one-way between-participants ANOVA. There was no significant difference in the gesture ratio between participants in the audio-only (M = 0.59, SD = 0.28), visual-only (M = 0.44, SD = 0.32), and multimodal (M = 0.42, SD = 0.30) conditions; $F(2,87)$ = 2.67, p = .08 (Figure 4).

Differences in path and manner gestures

To investigate the type of iconic gestures participants produced, we again calculated the ratio of path only, manner only, and path + manner conflated gestures per motion event description for each participant and item. For these calculations, total counts of path only, manner only, and path + manner gestures were divided by the number of motion event descriptions for each trial. The data was analysed in the same way as for speech. We ran a lmer model with fixed factors of input modality (audio-only, visual-only, and multimodal) and type of description (path-only, manner-only, and path + manner) using the ratio of path and manner gestures per motion event description as dependent variable (see Figure 5). The model revealed a fixed effect of type of description, χ^2 (2) = 531.82, p < .001, R^2 = .156. Regardless of input modality, speakers produced more path-only gestures than manner-only (β = .230, SE = .011, z = 20.59, p < .001, R^2 = .107) and path + manner gestures (β = .236, SE = .011, z = 21.14, p < .001, R^2 = .113). There was no difference between manner-only and path + manner gestures (β = .006, SE = .011, z =

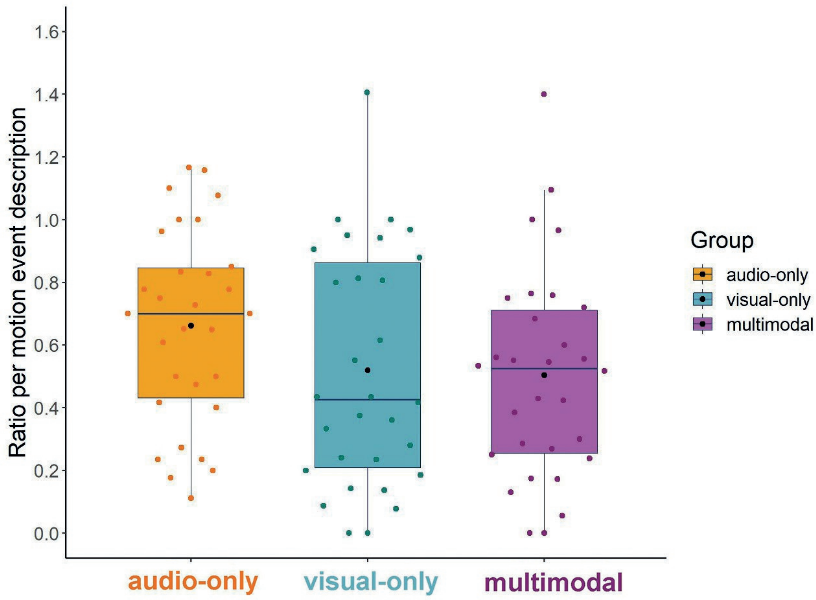


Figure 4. Ratio of gesture for motion event descriptions. Coloured dots represent the data for each participant. Black dots represent the group mean.

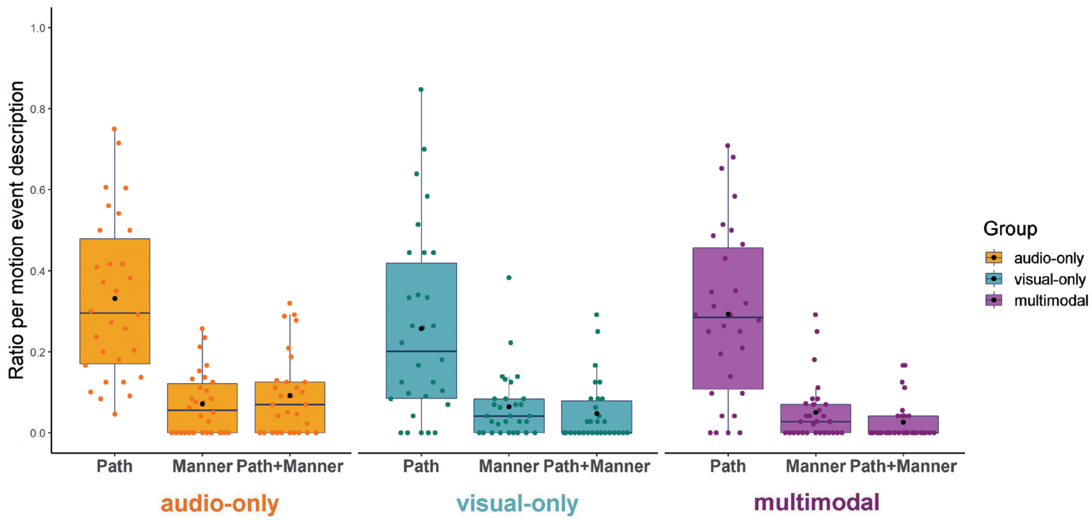


Figure 5. Path and manner gestures for motion event descriptions. Coloured dots represent the average data for each participant. Black dots represent the group mean.

.55, $p = .85$). The model revealed no fixed effect of input modality, χ^2 (2) = 3.64, $p = .16$, and no significant interaction between input modality and type of description on path and manner gestures, χ^2 (4) = 9.29, $p = .054$. See Appendix IV as supplementary data for the model summary table.

Discussion

Our goal was to investigate whether sensory modality of input influences the multimodal linguistic encoding of spatial information in motion events in speech and co-speech gesture. To determine this, we first examined the quantity of motion event descriptions in speech to establish whether the dominance of vision shown in perception lexicons (e.g., San Roque et al., 2015; Winter et al., 2018) is reflected in the linguistic encoding of motion events under experimental conditions. We found speakers produced more motion event descriptions when they watched events—either multimodal or visual-only—in comparison to when they only listened to events, i.e., audio-only. So, speakers provide richer linguistic information about spatial components of motion events when visual information is available. There was no difference in the amount of motion event descriptions between the visual-only and multimodal conditions, which suggests having auditory input on top of visual input does not further enrich speakers' motion event descriptions. These findings support the proposal that vision dominates in language, extending it to the domain of motion events.

There was, however, a qualitative difference in the verbal expressions of different spatial aspects of motion drawn from visual vs. auditory input. Speakers within the visual conditions mentioned manner more than path of motion, whereas speakers within the auditory condition mentioned path more often than manner. In addition, in the audio-only condition speakers mentioned path more often than they did in the visual conditions. This finding is in line with earlier studies of space showing non-visual input at encoding might lead to segmented path descriptions when describing routes (e.g., Iverson, 1999; Iverson & Goldin-Meadow, 1997). This might arise from the fact that non-visual spatial information is represented sequentially in contrast to holistic visual information. It is also possible that auditory input foregrounded path more than manner because information about manner of motion is less accessible without visual information. Although audition can provide high temporal acuity to differentiate rhythmic changes of movements (e.g., Recanzone, 2003; Repp & Penel, 2002), it might not provide detailed information to differentiate manners of motion to the same degree as vision (Malt et al., 2014). On the other hand, we used only three simple manners—i.e., *walk*, *run*, and *limp*, which may have been difficult to discriminate between based on auditory input alone. Our findings showed that participants, regardless of the condition, had more difficulty describing the limp than run events. A study using a more diverse set of manners could better test the affordances of audition vs. vision.

Interestingly, Turkish speakers in the visual conditions mentioned manner more often than path in their speech. Considering the typology of Turkish, this is interesting since Turkish speakers might be expected to omit manner more often in motion event descriptions (Kita & Özyürek, 2003; Özçalışkan et al., 2016b, 2018; Slobin, 1996; Talmy, 1985). Our findings suggest there may be universal processes at work, such that vision always provides more detailed information about manner of motion than audition, and therefore manner of motion might be more salient in visual input, even in a path language like Turkish. This suggests the sensory modality of input could influence speakers' encoding of spatial event components independently of the well-established tendencies of speaking a particular language (e.g., Slobin, 1996; Talmy, 1985). Future cross-linguistic studies could tease apart these possibilities systematically.

Although the finding that speakers in the visual conditions mentioned manner more than path seems discrepant with the usual typological patterns, we are not the first to report a reversed speech pattern in Turkish (Allen et al., 2007; Ter Bekke et al., 2022). Recently, Ter Bekke et al. (2022) also found that Turkish speakers used more manner than path when describing motion events presented as silent videos. To explain their findings, they highlighted the fact that they used salient manners—such as *tiptoe*, *twirl*, and *hop*—that are not "default" ways of changing location. Yet, this explanation does not hold for our findings, as the manners in our study were not particularly salient—i.e., *walk*, *run*, and *limp*. Alternatively, Allen et al. (2007) claimed that Turkish speakers are more likely to omit manner in larger discourse and when it does not simultaneously occur with path in motion events, as used in earlier studies. When manner and path are simultaneously present in motion events—as in the present study—Turkish speakers mention both elements in their event descriptions. Further studies should examine whether the saliency of manner or the ease of expression modulate linguistic encoding of manner, particularly in path-dominant languages (i.e., verb-framed languages; Talmy, 1985).

For gesture, we predicted that gesture frequency for both path and manner might decrease in the audio-only condition compared to the visual conditions because of the affordances of the visual modality. Due to the available mapping between gesture and vision (Macuch Silva et al., 2020), gesture production might be easier in the visual conditions than the audio-only condition. However, this was not the case in the present study. We found auditory input alone can elicit similar gesture frequency and gesture types—path and manner—as visual input. This suggests auditory input can lead to spatial imagery just as visual input does, as explicitly claimed by Hostetter and Alibali (2019). In line with this, Holler et al. (2022) found speakers produce spontaneous co-speech gesture depicting metaphorical spatial features of auditory pitch when describing sounds—e.g., producing a gesture higher in space to depict high pitch notes. Thus, our results support the argument that auditory information can also elicit gesture if it triggers spatial imagery.

Unexpectedly, the difference between path and manner expressions across input modalities found in speech was not reflected in co-speech gesture. Based on prior work (e.g., Iverson, 1999; Iverson & Goldin-Meadow, 1997), if speech for path is segmented, it may be ill-suited for path gesture, and consequently gesture frequency for path may decrease. Contrary to this, we found that although participants in the audio-only condition segmented path of motion more (i.e., made more reference to path in speech) than participants in the visual conditions, the frequency of their path gestures did not differ to those produced in the visual conditions. This discrepancy between our results and earlier findings could arise from the fact that these events only had single paths. So, although speech for path was segmented into smaller units, the amount of segmentation possible might be diminished since we are dealing with smaller-scale paths—as in our motion events—compared to larger-scale route description with multiple paths. Indeed, Iverson (1999) showed segmentation in path descriptions decreases with the diminishing size of a spatial layout.

We found the same discrepancy between speech and gesture for manner. Even though speakers in the visual-only and multimodal conditions mentioned manner more often in speech, there was no increase in the frequency of manner gestures. Regardless of the sensory modality of input, speakers produced more path only gestures than manner gestures, including path + manner, even in cases where they mentioned both path and manner in speech. One might hypothesize that expressing manner in speech was easier than in gesture, and participants might have chosen the modality strategically to avoid confusion for potential addressees who, according to our instructions, would go on to match descriptions to motion events. However, we think this is unlikely since earlier gesture studies of Turkish find that Turkish speakers typically gesture more about path than manner of motion (Aktan-Erciyes et al., 2022; Mamus, Speed, Rissman, et al., 2023; Özyürek et al., 2005; Ünal et al., 2022; although see Ter Bekke et al., 2022). So, the few manner gestures observed in our study fit the broader language typology (e.g., Akhavan et al., 2017; Chui, 2009; Gullberg et al., 2008).

Taken together, our findings are more in line with predictions that language typology is the determining factor in gesture production (e.g., Özçalışkan et al., 2016b, 2018) and that gestures are mostly shaped by language typology during speaking (e.g., Kita & Özyürek, 2003; Özyürek et al., 2005; Slobin, 1996) rather than sensory input. The discrepancy between our speech and gesture findings also suggests that even though speech affects gesture through language typology, gesture does not solely depend on speech contrary to the suggestions of some theories (e.g., Sketch Model, de Ruiter, 2000; Lexical Retrieval Hypothesis, Krauss et al., 2000; Growth Point Theory, McNeill, 1992), but consistent with the proposal that speech and gesture are independent, yet highly interactive systems (e.g., Gesture as Simulated Action Framework, Hostetter & Alibali, 2008; Gesture-for-Conceptualization Hypothesis, Kita et al., 2017).

Although our results imply the sensory modality of input does not affect the gesture of Turkish speakers, results may differ for a satellite-framed language that encodes manner in the main verb—such as English—or an equipollently-framed language—such as Mandarin Chinese (e.g., Brown & Chen, 2013). As manner is usually encoded in speech and co-speech gesture in such languages, the affordances of auditory vs. visual input might be more observable in gestural expressions of manner—e.g., auditory input may lead to fewer manner gestures than visual input. A cross-linguistic investigation is necessary to better understand whether and how co-speech gesture is influenced by the interaction of sensory modality of input and language typology.

Conclusion

The present study examined the role of sensory modality of input on the linguistic expression of motion event components in both speech and co-speech gesture and found they pattern in distinct ways. In comparison to the auditory modality, the visual modality appears to foreground manner more than path in speech, but gestures are generated similarly regardless of the sensory modality of input. These findings suggest the sensory modality of input influences speakers' encoding of path and manner of motion events in speech, but not in gesture.

Acknowledgements

We wish to thank Ayşe Serra Kaya and Şevval Çetinkaya for transcription and annotation, and Sevilay Şengül and Eylül Eski for the reliability coding. We also wish to thank Özcan Vardar for technical support during stimulus preparation, Jeroen Geerts for processing video data, and Maarten van den Heuvel for helping with the experimental setup. In addition, we wish to thank Ercenur Ünal and two anonymous reviewers who contributed valuable feedback and suggestions on an earlier version of this manuscript.

3

Lack of visual experience affects multimodal language production: Evidence from congenitally blind and sighted people

This chapter is based on:
Mamus, E., Speed, L. J., Rissman, L., Majid, A., & Özyürek, A. (2023). Lack of visual experience affects multimodal language production: Evidence from congenitally blind and sighted people. *Cognitive Science*, *47*(1), e13228. https://doi.org/10.1111/cogs.13228

Abstract

Human experience is shaped by information from different perceptual channels, but it is still debated whether and how differential experience influences language use. To address this, we compared congenitally blind, blindfolded, and sighted people's descriptions of the same motion events experienced auditorily by all participants (i.e., via sound alone) and conveyed in speech and gesture. Comparison of blind and sighted participants to blindfolded participants helped us disentangle the effects of a lifetime experience of being blind versus task-specific effects of experiencing a motion event by sound alone. Compared to sighted people, blind people's speech focused more on path and less on manner of motion, and encoded paths in a more segmented fashion using more landmarks and path verbs. Gestures followed speech, such that blind people pointed to landmarks more and depicted manner less than sighted people. This suggests visual experience affects how people express spatial events in multimodal language, and that blindness may enhance sensitivity to paths of motion due to changes in event construal. These findings have implications for the claims that language processes are deeply rooted in our sensory experiences.

Introduction

We experience the world through multiple perceptual channels, such as hearing footsteps while watching someone running upstairs. We also express our multimodal experience in language using different modalities, as in speech and gesture. Modern theories of language and cognition, including multimodal language theories, differ in whether they view language as a relatively embodied or disembodied system (see Meteyard et al., 2012, for a review). According to embodied theories, language processes—both speech and gesture—are deeply rooted in sensory and motor experience (e.g., Barsalou, 2016; Hostetter & Alibali, 2008, 2019; Pouw et al., 2014; Pulvermüller, 2013; Wilson, 2002), whereas disembodied symbolic theories suggest language processing relies on abstract, modality independent representations instead, which interface with perceptual representations later during semantic processing (e.g., Levelt, 1989; Mahon & Caramazza, 2008; Patterson et al., 2007).

Congenitally blind people, who do not have typical visual experience, provide an interesting opportunity to explore the relationship between multimodal experience and language. While some studies have claimed lack of visual experience does not change the way blind people understand and use language (e.g., Kim et al., 2021; Landau & Gleitman, 1985; Mahon et al., 2009; Özçalışkan et al., 2016b, 2018), there is also evidence to the contrary (e.g., Connolly et al., 2007; Iverson, 1999; Iverson & Goldin-Meadow, 1997; Shepard & Cooper, 1992). Thus, there is ongoing debate over whether and how experience shapes language (e.g., Barsalou, 2016; Bedny & Saxe, 2012; Mahon & Caramazza, 2008).

On the one hand, three-year-old blind children understand the semantics of vision-related words—such as look and see—in a manner comparable to their sighted peers (Elli et al., 2021; Landau & Gleitman, 1985). Studies on word comprehension also show no difference between blind and sighted people in semantic judgments of object concepts, actions, and vision-related terms (Bedny et al., 2012, 2019; Kim et al., 2021; Mahon et al., 2009; Marmor, 1978; Saysani et al., 2021). Similarly, previous studies of spatial language have emphasized the similarities in language between blind and typically-sighted people. For example, in one study of congenitally blind, sighted, and blindfolded speakers of Turkish and English, participants explored static scenes depicting motion with figurines—e.g., dolls in different postures so as to indicate running (Özçalışkan et al., 2016b). Both blind and blindfolded participants explored scenes haptically, while sighted people explored them visually. All speakers described motion events in speech and co-speech gesture according to the typology of their language. So, Turkish speakers were more likely to mention path (i.e., the trajectory of movement) and manner (i.e., how the movement was performed) in separate clauses (e.g., koşarak eve geldi 'she came to the house running'), whereas English speakers conflated these components into one clause (e.g., she ran to the house). Critically, gestures followed the language-specific patterns regardless of whether people were blind, blindfolded, or sighted. This suggests visual experience plays little role in language use.

On the other hand, there is evidence that there may be differences in language knowledge and use between blind and sighted people (e.g., Connolly et al., 2007; Iverson, 1999; Iverson & Goldin-Meadow, 1997; Kim et al., 2019a; Lenci et al., 2013; Shepard & Cooper, 1992). This holds for spatial language too. For example, English speaking blind and sighted people differ in their descriptions of routes in speech and gesture—especially regarding path expressions (Iverson, 1999; Iverson & Goldin-Meadow, 1997). When describing a familiar route in their school, blind children segmented the path according to several landmarks, whereas sighted and blindfolded children described paths more holistically using fewer landmarks and with more gestures accompanying speech (Iverson, 1999; Iverson & Goldin-Meadow, 1997). So, a blind child described a route as: "*Turn left, walk north, then you'll see the office, then you'll see 106, then 108, then 110, 112, then there's a doorway. Then there's a hall...*", whereas a sighted child said: "*when you get near the staircase you turn to the left*" (Iverson & Goldin-Meadow, 1997, p.463). Compared to gesture, speech is better suited to represent sequential information coming from auditory and haptic input. Since gesture does not require linearization to the same degree that speech does, it has been described as conveying meaning in a more "holistic" manner that is through analogue, iconic, and gradient representations (McNeill, 1992; McNeill & Duncan, 2000). This theory led Iverson and Goldin-Meadow (1997) to suggest that gesture is better suited for holistic than segmented meaning elements since gesture as a visual format, by nature, is not well-suited for linearization. Accordingly, they found that gesture frequency decreases with segmented path descriptions (i.e., "when path is broken up into a series of locations" on Iverson & Goldin-Meadow, 1997, p.463), particularly when the spatial layout is large-scale and includes multiple paths (Iverson, 1999). This is corroboratory evidence from language that spatial cognition in blind people is more sequential than in sighted people (e.g., Cattaneo & Vecchi, 2011; Iachini et al., 2014; Noordzij et al., 2006; Pasqualotto & Proulx, 2012; Ruggiero et al., 2021; Thinus-Blanc & Gaunet, 1997; Vercillo et al., 2018), and lack of visual experience may shape spatial language via altered spatial representations.

In light of these conflicting results, it is unclear what role visual experience plays in multimodal spatial language use. The previous studies, while informative, have some potential drawbacks which make them difficult to synthesize. First, some of these studies examined pre-existing spatial representations—i.e., familiar routes (Iverson, 1999; Iverson & Goldin-Meadow, 1997), whereas others used novel spatial scenes (Iverson, 1999; Özçalışkan et al., 2016b, 2018). Second, some studies did not control the type of input at encoding—i.e., how participants learned routes (Iverson, 1999; Iverson & Goldin-Meadow, 1997), and some did not equate input modalities—i.e., sighted participants explored scenes visually whereas blind and blindfolded participants explored scenes haptically (Iverson, 1999; Özçalışkan et al., 2016b, 2018). In addition, in Özçalışkan et al. (2016b, 2018) time spent exploring scenes visually versus haptically was not controlled, so haptic groups could have taken longer exploring scenes which allowed them to compensate for differential input. Finally, in some studies speakers were explicitly asked to gesture as they described scenes (Özçalışkan et al., 2016b, 2018), which might have affected how scenes were encoded.

The present study

The present study mitigates these limitations by conducting a new experiment with blind and sighted people where all participants receive the same motion event input. Auditory motion events were recorded depicting a person walking, running, or limping to and from landmarks and presented to participants to elicit verbal descriptions and spontaneous co-speech gesture. Our study has the advantage that it includes ecologically relevant stimuli. Hearing sounds of human locomotion is familiar to both blind and sighted people, and previous research has shown that sighted people are able to extract information about path and manner of motion from auditory input alone (Geangu et al., 2021; Mamus et al., 2019; Mamus, Speed, Özyürek, et al., 2023). To better distinguish whether potential differences in the linguistic encoding of spatial information arise from the long-term effect of blindness or are due instead to momentary effects of lack of vision at encoding, we compared blind and sighted people to blindfolded people. It has been shown that closing the eyes while attending to auditory information modulates attention (Wöstmann et al., 2020). By comparing blindfolded to blind participants, we are better able to determine whether any differences between sighted and blind people reside in momentary stimulus affordances.

We had different predictions concerning speech and gesture based on slightly different literatures regarding perceptual language and current theories of multimodal language production. Accordingly, we will consider the predictions regarding speech and gesture in turn.

Speech

A number of studies report that vision dominates in the perceptual lexicons of languages (e.g., Floyd et al., 2018; Levinson & Majid, 2014; Majid et al., 2018; San Roque et al., 2015; Viberg, 1983; Winter et al., 2018) and leads to richer motion event descriptions (more manner distinctions encoded) than auditory information alone (Mamus, Speed, Özyürek, et al., 2023). Together, this suggests descriptions produced by blind people may be different compared to sighted people. Specifically, we predicted that blind people may produce fewer motion event descriptions overall than sighted people. At the same time, blind people are known to rely more extensively on audition than sighted people to localize space, and are often better than sighted people at processing auditory information (e.g., Battal et al., 2020; Gougoux et al., 2004; Röder et al., 1999; Wan et al., 2010). So, blind participants might provide as many motion event descriptions—if not more—than sighted participants.

In addition to overall number of motion event descriptions, we examined speech for landmark use when participants expressed paths. Earlier route description studies found blind people segment path descriptions using landmarks more than sighted people (Iverson, 1999; Iverson & Goldin-Meadow, 1997). Here, we test if this hypothesis is confirmed with experimentally controlled motion events and examine whether blind participants still use more landmarks than blindfolded and sighted participants.

Furthermore, previous spatial cognition studies have found blind people rely mainly on an egocentric rather than allocentric spatial frame of reference (e.g., Cattaneo & Vecchi, 2011; Iachini et al., 2014; Pasqualotto & Proulx, 2012). Accordingly, we predict that spatial locations will be described more in relation to blind people's own position in space. That is, blind people may mention landmarks in relation to their own body (i.e., self-anchored; *from my left*), instead of using external coordinates (e.g., object-anchored; *from the elevator*). Therefore, we also tested whether mentions of landmarks in the blind participants were primarily self-anchored and those of non-blind participants were more object-anchored.

Finally, we examined speech for the encoding of path and manner separately. With regard to path, based on the previously attested differences in the encoding of path (i.e., segmented paths with more landmarks in blind vs. non-blind; Iverson, 1999; Iverson & Goldin-Meadow, 1997), it might be expected that increased segmentation would increase the use of path verbs. So, blind participants may mention path more often within each description in speech. For manner, vision seems to provide richer information about manner than audition (Malt et al., 2014; Mamus, Speed, Özyürek, et al., 2023), so perhaps blind participants will produce fewer manner expressions. On the other hand, earlier studies suggest blind people can differentiate the semantic similarity of actions as well as sighted people (Bedny et al., 2012, 2019), so perhaps there will be no difference between groups.

Gesture

Theories vary in their specification of the interaction between speech and gesture, as well as in how they view the nature of spatial imagery underlying gesture production (de Ruiter, 2000, 2007; Hostetter & Alibali, 2008; Kita & Özyürek, 2003; Krauss et al., 2000; McNeill, 1992; McNeill & Duncan, 2000). Gesture theories typically emphasize the role of visuo-spatial imagery in gesture production (e.g., de Ruiter, 2000; Hostetter & Alibali, 2008; Kita & Özyürek, 2003; Krauss et al., 2000), although studies have shown gesture can be derived from auditory information alone in sighted people too (Holler et al., 2022; Mamus, Speed, Özyürek, et al., 2023). Though, if visuo-spatial imagery is one of the main sources of gesture production, the lack of any visual experience, as in the case of congenital blindness, might lead to differences in how people gesture in relation to spatial events. Indeed, earlier studies found the rate of spontaneous gesturing was lower among blind than sighted people when describing routes (Iverson, 1999; Iverson & Goldin-Meadow, 1997) and motion events (Özçalışkan et al., 2016b, 2018). Based on this, we predicted fewer spontaneous gestures among blind than non-blind people in motion event descriptions.

Second, we examined speakers' pointing gestures used with mentions of landmarks in speech. Pointing gestures can be used to direct attention to an object or place an object in gesture space during communication (e.g., McNeill, 2000). While describing a motion event, speakers can use pointing gestures to locate landmarks to be communicatively clear. We know blind people are good at localizing sounds and often outperform sighted people (e.g., Battal et al., 2020; Lessard et al., 1998; Röder et al., 1999; Voss et al., 2004). So, it might be expected that blind participants would produce more pointing gestures than non-blind participants.

Finally, we examined speakers' iconic gestures for path and manner. Previous studies (Iverson, 1999; Iverson & Goldin-Meadow, 1997) claimed gesture production decreases with segmented speech because gestures are better suited for holistic expression due to their visual format being less-suited for linearization than speech (McNeill, 1992; McNeill & Duncan, 2000). Based on this, if blind participants use more path verbs to segment their descriptions than non-blind participants, we might not expect a similar increase in frequency of path gestures in blind compared to non-blind participants. But, according to speech-gesture interface theories one would also expect gestures to parallel speech patterns and align with speech frequency (e.g., Kita & Özyürek, 2003; Özçalışkan et al., 2016b, 2018). If so, there would be more path gestures in blind than non-blind participants. Similarly for manner gestures, visual experience of human locomotion may be necessary to map the sounds of manner into gesture regardless of speech. If so, blind participants would express manner less often in gesture than non-blind participants. Alternatively, gesture patterns may align with speech and so, if blind participants mention manner in their speech at comparable rates to non-blind participants, we would not expect a difference in manner gestures.

Method

Participants

Twenty-one congenitally blind (M = 28.19 years, SD = 6.56, range = 18–40), 21 blindfolded (M = 27.43 years, SD = 6.10, range = 19–49), and 21 sighted (M = 27.29 years, SD = 6.61, range = 20–41) native Turkish speakers were paid to participate in the experiment. The sample size was determined by access to the special population with the control groups matched to the number of blind participants recruited. At the time of testing, 12 blind participants had light perception and 9 had total blindness (see Table 1 for detailed characteristics of the blind participants). Blindfolded and sighted participants with normal or corrected-to-normal vision were matched for age, gender, and education to blind participants. Participants were tested in a quiet room on Boğaziçi University campus. They all were paid the equivalent of €9 in Turkish Lira for their participation and provided written informed consent approved by the IRB committees of Boğaziçi and Radboud Universities.

Table 1. Blind participants demographic information.

Ss	Gender	Age	Age of Blindness	Cause of Blindness	Residual Light Perception*	Highest level of Education
101	NB	25	Birth	Retinal degeneration	yes	BA (student)
102	F	26	Birth	Retinal degeneration	yes	BA
103	M	19	Birth	Optic nerve atrophy	yes	BA (student)
104	M	25	Birth	Retinitis pigmentosa	yes	BA
105	M	24	Birth	Optic nerve hypoplasia	yes	BA
106	F	25	Birth	Retinitis pigmentosa	yes	BA (student)
108	F	28	Birth	Retinitis pigmentosa	yes	BA
109	F	25	Birth	Optic nerve hypoplasia	yes	BA
112	M	25	7 months	Retinoblastoma	none	BA (student)
114	M	20	Birth	Premature birth	none	BA (student)
115	M	40	Birth	Hereditary/ Unknown cause	yes	PhD
116	M	26	Birth	A genetic disease	yes	MA
117	M	24	Birth	Anophthalmia	none	BA
118	M	36	Birth	Retinitis pigmentosa	none	BA
119	M	30	Birth	Dry optic nerves	none	BA
120	M	33	Birth	Norrie disease	none	BA (student)
122	F	39	Birth	Hereditary/ Unknown cause	yes	MA
123	F	39	6 months	Retinoblastoma	none	MA (student)
125	M	35	6 months	Retinoblastoma	none	BA
126	F	18	Birth	Premature birth/ Retinal Tear	none	BA (student)
127	F	30	Birth	Optic nerve atrophy	yes	MA

*under optimal conditions

Auditory stimuli

We audio-recorded locomotion and non-locomotion events performed by an actress. Locomotion events were the critical items and non-locomotion events were filler items. We created 12 locomotion events by crossing 3 manners (walk, run, and limp) with 4 paths (to, from, into, and out of) in relation to a landmark object (door or elevator)—e.g., "*someone walks from a door*". An audio-recorder was placed next to the landmark objects. For *to* and *into* events, the actress approached the landmarks, so the path direction approaching the audio-recorder—and for *from* and *out of* events, the actress moved away from the landmarks, so the path direction moving away from the audio-recorder. To ensure that landmark objects were recognizable, we created auditory landmarks. For example, for the "elevator" landmark, we recorded the sound of an elevator ring—the tone that is heard when an elevator arrives at its destination. We also recorded the sound of an elevator door opening automatically. Then we created a combined audio-file: the ring (representing the arrival of the elevator) followed by the opening sound.

In addition, we edited the path azimuth angle using Soundtrack Pro audio editing software to vary the path motion. Five movement angles were created in a semicircular space ranging from 90° left to 90° right with 45° intervals. From the right to the left these were: 0° (right), 45° (right-sided), 90° (front), 135° (left-sided), and 180° (left) motions (see Figure 1). We created 12 events with 5 movement angles, resulting in 60 events in total. All locomotion events were exported as 5.1 surround sound.

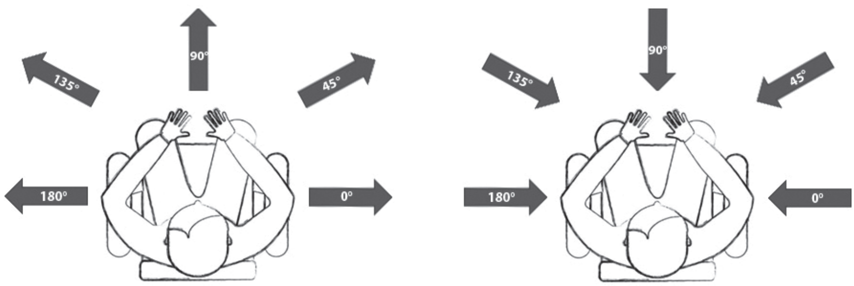


Figure 1. Five movement angles for "*from*" and "*out of*" events (left) and "*to*" and "*into*" events (right). The figure was taken from Mamus et al. (2019).

To create non-locomotion events, the same actress performed "transitive" actions with different objects (e.g., opening a can, chopping a cucumber), and audio was recorded at a fixed distance. We do not examine these items further.

There were 77 trials per person, including a total of 60 locomotion events and 17 non-locomotion events. Locomotion events lasted on average 9s (*SD* = 1.9) and non-locomotion events 8s (*SD* = 2.2). The event list and stimuli are available at https://osf.io/qsr7j/.

Procedure

The procedure was the same for all groups, except that blindfolded participants' eyes were covered with a mask before they entered the room. Five speakers were placed 1.34 m from the participant's head and approximately 95 cm from the ground in a 5+1 surround system configuration. Front left and right speakers were placed 30° off center, and rear left and right speakers were 110° off center. Participants sat in the middle of the speakers. The experimenter stayed in the room to initiate the task and advance trials on a laptop using Presentation Software.

Events were presented aurally and participants were asked to describe each event at their own pace without any instructions about gesture use. They were told that another participant would watch the video recording of their descriptions and listen to the same events to match descriptions with events. At the beginning of the experiment, participants performed two practice trials consisting of one locomotion and one non-locomotion event. Further clarification was provided, if necessary, after the practice trials. Descriptions were recorded with two video cameras. One camera was approximately 1.5 m across from participants and the other recorded the top view of the participants' frontal space so as to capture arm and hand movements. Participants filled out a demographic questionnaire—including questions about blindness history for blind participants—on another laptop after the event description task. The experiment lasted around 45 minutes.

Coding

Speech

Descriptions of locomotion and non-locomotion events were annotated by native Turkish speakers using ELAN (Wittenburg et al., 2006), but only descriptions for the locomotion events were transcribed and coded. Event descriptions were split into sentence units, defined as a verb and its associated arguments (Azar et al., 2020; Özçalışkan et al., 2016b). Sentence units could contain a subordinate clause as well. Sentence units were then coded as motion event descriptions if they referred to locomotion (e.g., someone is running into an elevator); sentence units including a transitive event, e.g., "*opening a door*" or "*ringing a bell*", or other information, e.g., "*wearing high heels*" or "*a wooden floor*", were coded as irrelevant to the target event.

Motion event descriptions were coded for: landmark—either source (start point of movement) or goal (end point of movement), (b) path (trajectory of motion), and (c) manner (how the action is performed). We also coded whether landmarks reference either: (i) external objects (e.g., from/to a door or elevator) or (ii) self-anchored (the speaker's body, e.g., to/from my left)—see Table 2 for an example. We calculated the

Interclass Correlation Coefficient (ICC) between two coders to measure the strength of inter-coder agreement for landmark, path, and manner in speech (Koo & Li, 2016). Agreement between coders was .94 for object-anchored landmark, .96 for self-anchored landmark, .98 for path, and .95 for manner of motion.

Table 2. An illustrative example of a description and its coding.

Turkish description	**Asansör**	**-den**	**çık**	**-ıp**	**sağ**	**- ım**	**-a**	**doğru**	**yürü**	**-yor**
Glossing	elevator	ABL	exit	GER	right	1sPOSS	DAT	towards	walk	PRS.3SG
Turkish description	object-anchored landmark		path		self-anchored landmark				manner	
English translation	'(someone) exited from the elevator walking towards my right'									

Co-speech gesture

Participants' spontaneous representational gestures (pointing and iconic) were identified for each target motion event description (Kita, 2000). We coded gesture strokes (i.e., the meaningful phase of a gesture) that co-occurred with parts of the description. Each continuous instance of hand movement was coded as a single gesture. Pointing gestures were either head or hand pointing gestures to empty locations in gesture space and were coded when they represented a source/goal landmark in speech. For example, if a speaker pointed to a spatial location to indicate the starting point of movement without showing its trajectory, the gesture was coded as a pointing gesture referring to localization of a landmark (e.g., Figure 2). Iconic gestures representing trajectory or manner of motion were further classified into the following categories:

(a) ***path-only*** gestures depict trajectory of movement without representing manner
(b) ***manner-only*** gestures show the style of movement without representing trajectory
(c) ***path + manner*** gestures depict both trajectory and manner of motion simultaneously

We calculated the ICC between two coders to measure the strength of inter-coder agreement for identifying a gesture and coding each type of gesture. Agreement between coders was .88 for identifying gestures and between .82–.93 for type of gesture—i.e., 89 for coding pointing gestures, .89 for coding path only, .93 for manner only, and .82 for path+manner gestures).

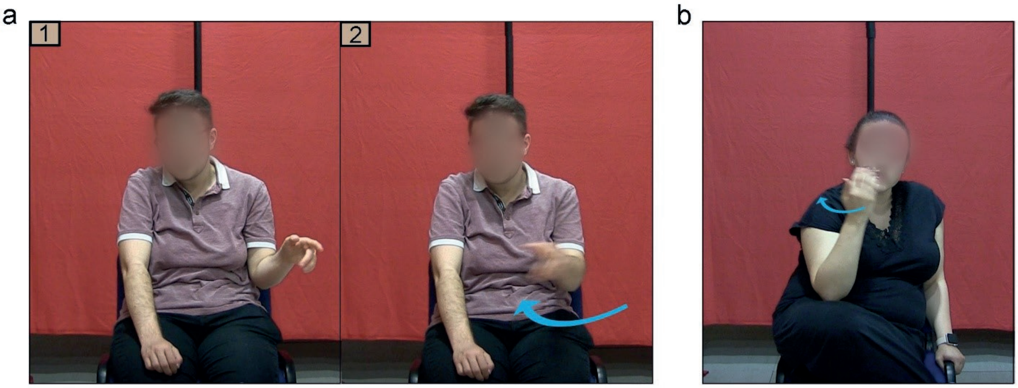


Figure 2. (a) A blind participant produces a pointing gesture (1) to a landmark and (2) then a path gesture while saying *soldan sağa doğru geldi* 'came from the left towards the right'. (b) A sighted participant produces a path gesture (hand moving backwards) while saying *içeri giriyor* 'entering inside'.

Results

To analyze the data, we used linear mixed-effects regression models (Baayen et al., 2008) with random intercepts for participants and items, using the packages *lme4* (Version 1.1–28; Bates et al., 2015) with the optimizer *nloptwrap* and *lmerTest* (Version 3.1–3; Kuznetsova et al., 2017) to retrieve p-values in R (Version 4.1.3; R Core Team, 2022). We conducted linear mixed effects models on the different motion elements in speech and gesture. To assess statistical significance of the fixed factors and their interaction, we used likelihood-ratio tests with χ^2, comparing models with and without the factors and interaction of interest. For post-hoc comparisons and to follow-up interactions, we used *emmeans* (Version 1.7.3; Lenth, 2022). Data and analysis code are available at https://osf.io/qsr7j/.

Speech

We examined speech for the overall amount of motion event descriptions, landmark use, and reference to path and manner.

Overall amount of motion descriptions

First, we tested whether participants differed in the speech they produced for motion events. We ran a glmer model with the fixed factor of group (blind, blindfolded, or sighted) on binary values for mention of motion event description in speech (0 = no, 1 = yes) as a dependent variable. It revealed no effect of group on motion event description, χ^2 (2) = .91, p = .635.

Landmark use in speech

We predicted that blind participants would segment descriptions using more mention of landmarks than blindfolded and sighted participants. To account for baseline differences in the number of motion event descriptions produced, we calculated the ratio of landmark (including all types of landmark) per motion event description for each participant and item. We ran an lmer model with the fixed factor of group using the ratio of mention of landmark per motion event description as the dependent variable (Figure 3). The model revealed an effect of group, χ^2 (2) = 15.41, $p < .001$. Blind participants mentioned landmarks more than blindfolded ($\beta = .421$, $SE = .012$, $t = 3.40$, $p = .003$) and sighted ($\beta = .452$, $SE = .012$, $t = 3.65$, $p = .002$) participants, and there was no difference between blindfolded and sighted participants, $\beta = .032$, $SE = .012$, $t = 0.26$, $p = .97$.

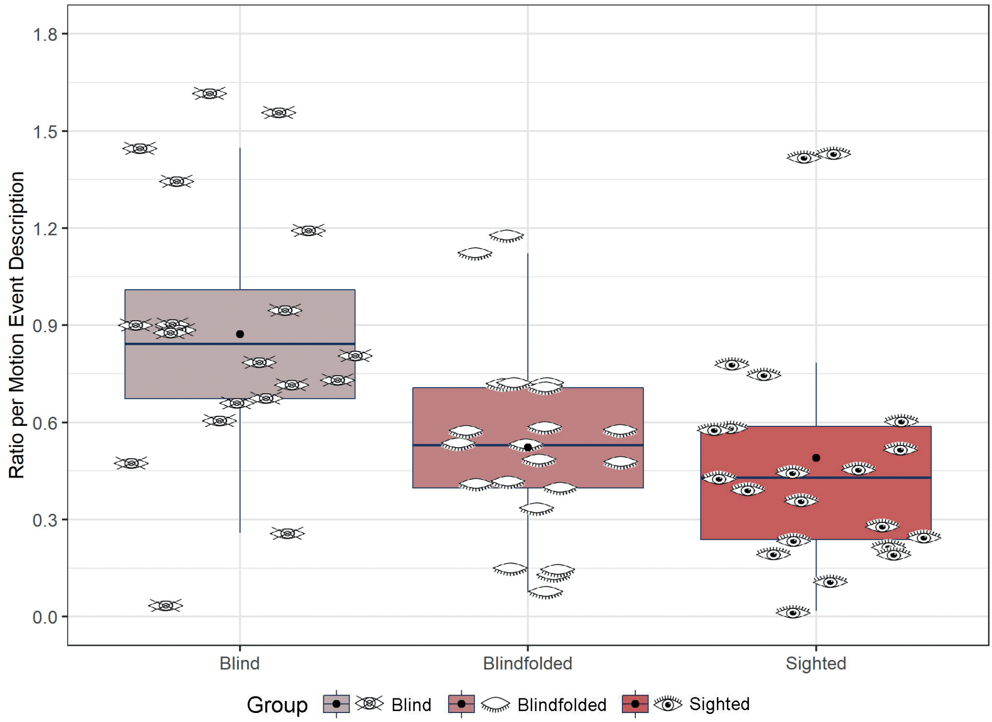


Figure 3. Overall landmarks in speech. Eye icons represent the average ratio for each participant. Black dots represent the group mean.

We further predicted that if blind people rely more on an egocentric frame of reference, they would use more self-anchored landmarks than blindfolded and sighted participants. In contrast, blindfolded and sighted people would use more object-anchored landmarks than blind participants. To test this, we calculated the ratio of mention of self-anchored and object-anchored landmark per motion event description for each participant and item. Then, we ran an lmer model with the fixed factors of group and landmark reference (object- or self-anchored) using the number of mention of landmark per motion event description as the dependent variable (Figure 4). The model revealed an effect of group, χ^2 (2) = 14.98, p < .001, showing that blind participants mentioned more landmarks in their speech than non-blind participants, and an effect of landmark category, χ^2 (2) = 160.33, p < .001, showing that object-anchored landmarks were mentioned more than self-anchored landmarks. Yet, the model also revealed an interaction between group and landmark category, χ^2 (2) = 161.03, p < .001. To follow up the interaction we compared the effect of group separately by landmark category. As expected, blind participants referred to self-anchored landmarks more than blindfolded (β = .292, SE = .053, t = 5.50, p < .001) and sighted (β = .305, SE = .053, t = 5.74, p < .001) participants, and there was no difference between blindfolded and sighted participants (β = .014, SE = .053, t = 0.25, p = .97). But, the groups did not differ in terms of reference to object-anchored landmarks, all ps > .10.

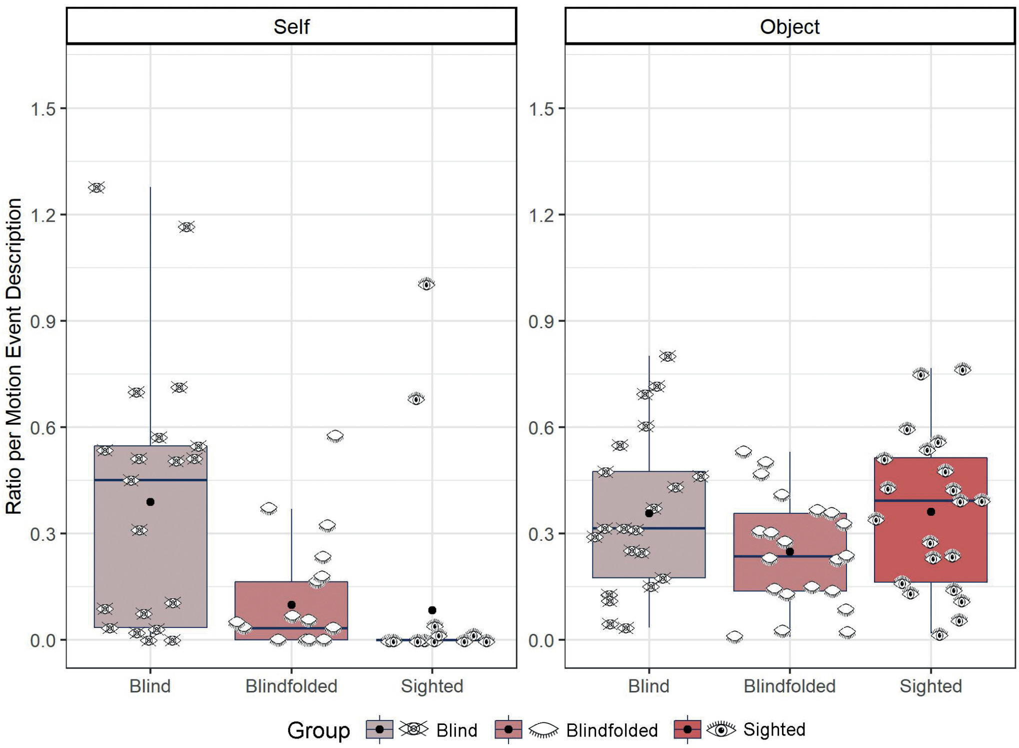


Figure 4. Self and object-anchored landmarks in speech. Eye icons represent the average ratio for each participant. Black dots represent the group mean.

Path and manner use in speech

Next, we examined whether participants differed in how they expressed path and manner in speech. For this, we calculated the ratio of mention of path and manner per motion event description for each participant and item. We ran an lmer model with the fixed factors of group and type of expression (path vs. manner) and their interaction term using the ratio of mention of path and manner per motion event description as the dependent variable (Figure 5). The model revealed no effect of group, χ^2 (2) = 0.68, p = .71, no effect of type of expression, χ^2 (2) = 0.004, p = .95, but an interaction between group and type of expression, χ^2 (2) = 16.31, p < .001. To follow-up the interaction, we used the *emmeans* function to compare the groups for path and manner use separately.

For path, although the interaction was significant in the model, there was no difference in the mention of path between blind and sighted (β = .224, SE = .103, z = 2.18, p = .075), blind and blindfolded (β = .104, SE = .102, z = 1.02, p = .56), or blindfolded and sighted (β = .120, SE = .102, z = 1.17, p = .47). However, the difference between blind and sighted participants (β = .224, SE = .104, t = 2.15, p = .033) was significant when we did not use the conservative p-adjustment in *emmeans*: blind participants mentioned path more than sighted participants.

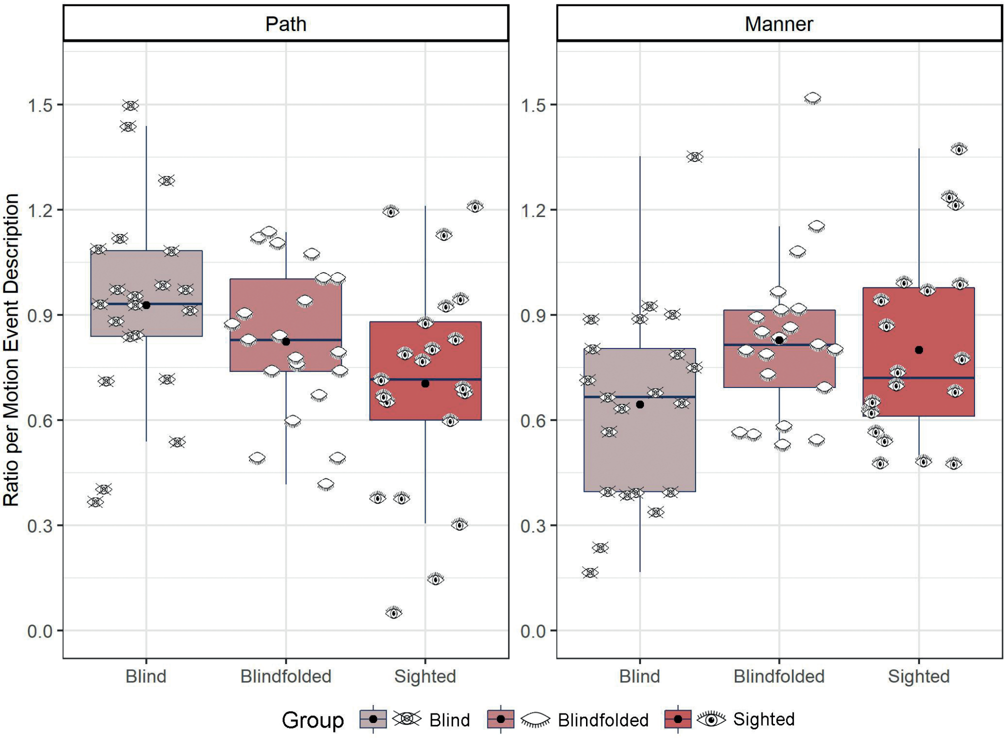


Figure 5. Path and manner in speech. Eye icons represent the average ratio for each participant. Black dots represent the group mean.

For manner, blind participants mentioned manner less often than sighted (β = -.345, *SE* = .104, *t* = -3.32, *p* = .001) but not blindfolded (β = -.183, *SE* = .103, *t* = -1.78, *p* = .08) participants. There was no difference between blindfolded and sighted participants β = -.162, *SE* = .104, *t* = -1.56, *p* = .12). The interaction between group and type of expression can be seen in Figure 5.

Gesture

As with speech, we first examined the overall amount of gesture produced by each group, before comparing landmark gestures, and path and manner gestures. As the amount of gesture changes as a function of the rate of motion event descriptions, we first calculated the gesture ratio per motion event description by dividing the total number of gestures by the total number of motion event descriptions. To further investigate what type of gestures participants produced, we calculated the number of pointing gestures referring to localization of landmark (hand and head pointing combined) and iconic (path-only, manner-only, and path+manner) gestures per motion event description for each participant and item. For these calculations, total counts of pointing gestures, path-only, manner-only, and path+manner gestures were divided by the number of motion event descriptions for each trial. Hand gestures constitute 81.5% of the pointing gestures. The data was analyzed in the same way as speech.

Overall gesture rate

We compared the groups in terms of their overall gesture ratio using a one-way between-participants ANOVA. There was a significant difference in the gesture ratio between blind (*M* = 0.44, *SD* = 0.48), blindfolded (*M* = 0.82, *SD* = 0.53), and sighted (*M* = 0.69, *SD* = 0.47) participants; $F(2,60) = 3.18$, *p* = .049. A post-hoc Tukey test showed that blindfolded participants had more gestures than blind participants (*p* = .041), but there was no difference between sighted and blind (*p* = .25) or blindfolded and sighted participants (*p* = .65).

Pointing gestures to landmarks

We predicted that if blind participants would use more landmarks in their speech than non-blind participants, this might be reflected in more pointing gestures to landmarks (Figure 6), and §*Landmark use in speech* showed blind individuals did mention landmarks more often. To test for differences in gesture, we ran an lmer model with the fixed factor of group using the number of pointing gestures per motion event description as the dependent variable. The model revealed a marginal effect of group, χ^2 (2) = 5.81, *p* = .055. Blind participants produced more pointing gestures than sighted (β = .156, *SE* = .064, *z* = 2.45, *p* = .038) but not blindfolded participants (β = .095, *SE* = .064, *z* = 1.50, *p* = .29). There was no difference between blindfolded and sighted participants (β = .060, *SE* = .064, *z* = 0.95, *p* = .61).

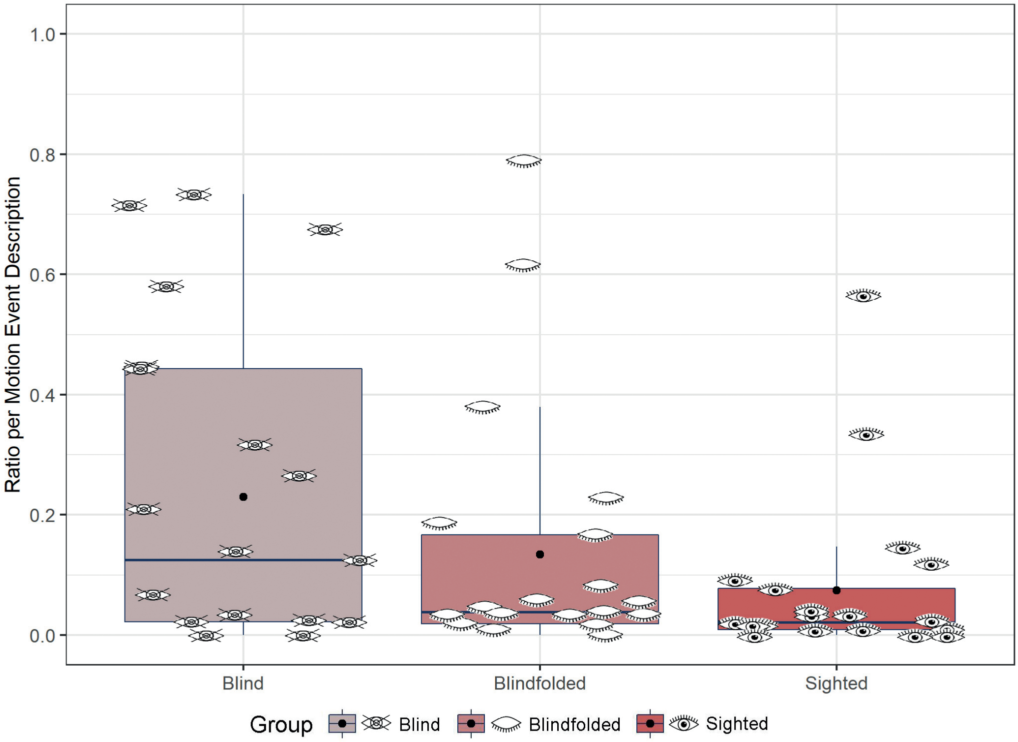


Figure 6. Pointing gestures to landmarks. Eye icons represent the average ratio for each participant. Black dots represent the group mean.

Path and manner gestures

To compare iconic gestures, we ran an lmer model with the fixed factors of group and type of expression (path-only, manner-only, and path+manner) using the ratio of path and manner gestures per motion event description as the dependent variable (Figure 7). The model revealed an effect of group, χ^2 (2) = 10.39, p = .006, an effect of type of expression, χ^2 (2) = 1354.7, p < .001, and an interaction effect of group and type of expression, χ^2 (2) = 52.67, p < .001.

All groups produced more path-only gestures than manner-only (β = .227, SE = .007, t = 30.99, p < .001) or path+manner gestures (β = .253, SE = .007, t = 34.54, p < .001). To follow-up the interaction, we used the *emmeans* function to compare the groups for each gesture type separately. Blind participants produced fewer path-only gestures than blindfolded (β = -.167, SE = .035, t = -4.81, p < .001) and sighted (β = -.119, SE = .035, t = -3.41, p = .001) participants. Also, blind participants produced fewer manner-only gestures than blindfolded (β = -.074, SE = .035, t = -2.15, p = .037) and sighted participants (β = -.096, SE = .035, t = -2.76, p = .008). Blind participants also produced fewer path+manner gestures than blindfolded (β = -.075, SE = .035, t = -2.15, *p* = .036) but not sighted (*β* = -.033, *SE* = .035, *t* = -.94, *p* = .35) participants. There was no difference between blindfolded and sighted participants in terms of path-only (*β* = .048, *SE* = .035, *t* = 1.39, *p* = .17), manner-only (*β* = -.022, *SE* = .035, *t* = -.62, *p* = .54), or path+manner (*β* = .042, *SE* = .035, *t* = 1.21, *p* = .23) gestures.

Overall, then, blind participants produced fewer iconic gestures—both path and manner—than blindfolded and sighted participants, but there was no difference between blindfolded and sighted participants.

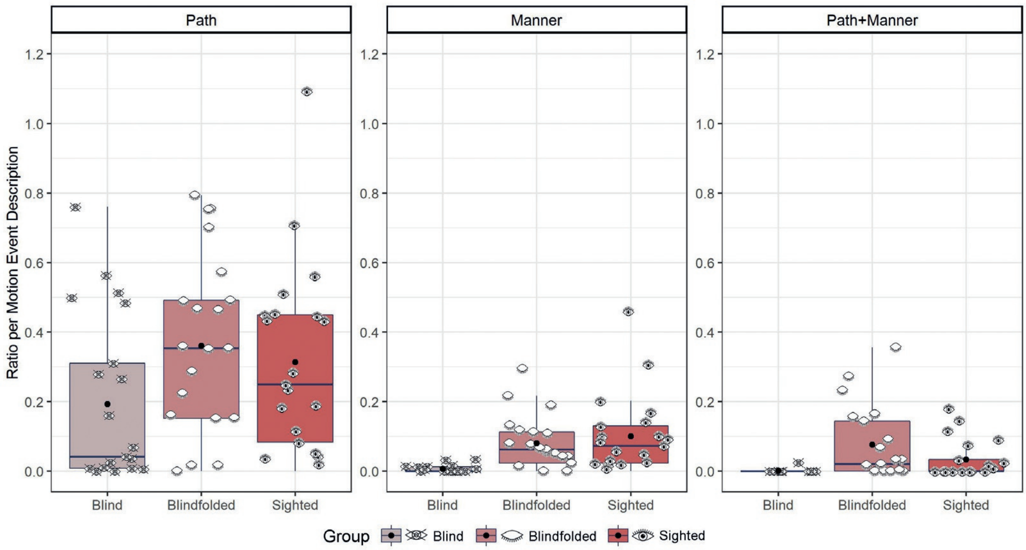


Figure 7. Path and manner gestures for motion event descriptions. Eye icons represent the average ratio for each participant. Black dots represent the group mean.

Discussion

Our findings point to some similarities, but also notable differences between blind people's multimodal language use and their sighted and blindfolded counterparts. All speakers produced a comparable amount of motion event descriptions in their speech, but differed in how they referred to certain aspects of events. In comparison to non-blind (both blindfolded and sighted) speakers, blind speakers were more likely to use landmarks and, in particular, more self-anchored landmarks. In addition, blind speakers tended to talk more about path and less about manner of motion events than sighted speakers. With regard to co-speech gesture, we observed a similar gesture rate between blind and sighted speakers. However, speakers' gesture frequency differed depending on the gesture type: blind speakers produced more pointing gestures with landmarks than sighted speakers, but had fewer path and manner gestures than non-blind speakers (blindfolded and sighted). Even though all speakers' gesture patterns were consistent with the Turkish motion typology (i.e., path dominant gestures), blind speakers produced fewer iconic gestures than non-blind speakers overall. We contextualize and discuss each of these points in more detail.

The fact that Turkish blind and non-blind (blindfolded and sighted) individuals did not differ in the overall amount of verbal descriptions produced is perhaps not surprising given that blind people are good at processing auditory information (e.g., Battal et al., 2020; Gougoux et al., 2004; Röder et al., 1999; Wan et al., 2010). Similarly, we found co-speech gesture rates were comparable between blind and sighted individuals, although blind people gestured less than blindfolded speakers. At first glance, this seems partially inconsistent with what has been reported in earlier studies—i.e., blind speakers produce less gesture than sighted speakers (Iverson, 1999; Iverson & Goldin-Meadow, 1997; Özçalışkan et al., 2016b, 2018). However, this apparent contradiction could be because earlier studies focused only on iconic gesture production, whereas the current study examined different gesture types—both pointing and iconic.

Although overall rates of speech and gesture were comparable across groups, there were notable qualitative differences in the verbal and gestural expressions which merit further discussion. For example, blind speakers mentioned landmarks more than non-blind (blindfolded and sighted) speakers. In particular, when landmarks were mentioned, blind speakers were more likely to refer to them in relation to their own position (e.g., self-anchored; *from my left*). We also found blind speakers had more pointing gestures to posited landmarks in gesture space than sighted speakers. Taken together, this is in line with previous studies that find blind people rely more on egocentric than allocentric frames of reference when learning spatial layouts (e.g., Cattaneo & Vecchi, 2011; Iachini et al., 2014; Pasqualotto & Proulx, 2012; Ruggiero et al., 2021). Thus, our results provide further linguistic evidence for the use of egocentric frame of reference in spatial language (see also Iverson, 1999; Iverson & Goldin-Meadow, 1997).

Blind speakers also used more path verbs than sighted speakers. Previous route description studies (Iverson, 1999; Iverson & Goldin-Meadow, 1997) found blind people use landmarks on routes and suggest this is because blind people segment paths more in order to make routes more navigable. Although our motion events had single paths (i.e., smaller-scale in comparison to earlier route description studies with multiple paths), speakers could still segment paths into smaller units by mentioning landmarks more and, thus, utilizing different path verbs in their descriptions of a single event (e.g., *someone came from my side and went away towards the elevator*). So, this path segmentation is a result of more mentions of landmarks (e.g., "from my side" and "towards the elevator"). Together with the increased landmark use, increased mention of path in speech suggests that blindness may enhance sensitivity to paths due to changes in event construal that arise from altered spatial cognition (e.g., Cattaneo & Vecchi, 2011; Lessard et al., 1998; Röder et al., 1999; Voss et al., 2004). At the same time, blind speakers did not differ from blindfolded speakers, suggesting temporary lack of vision through blindfolding at encoding can also lead to changes in the encoding of path in motion events.

In contrast to speech, blind speakers used fewer path gestures than non-blind (blindfolded and sighted) speakers. Even though there was a mismatch in the frequencies of path in speech and path in gesture, speech and gesture type were still coupled with respect to motion event depictions in Turkish—i.e., separated path and manner use in both speech and gesture (e.g., Kita & Özyürek, 2003; Özçalışkan et al., 2016b, 2018). The reduced frequency of path gestures from blind speakers could arise for a different reason, however, namely because gesture frequency decreases when paths are more segmented in speech, as suggested by earlier studies (Iverson, 1999; Iverson & Goldin-Meadow, 1997). This could be because gestures are better suited for holistic than segmented expression due to their visual format (McNeill, 1992; McNeill & Duncan, 2000).

In contrast to path talk, blind speakers mentioned manner less often in speech than sighted speakers. Earlier language comprehension studies have shown that blind and sighted speakers have similar semantic knowledge of action and motion verbs (e.g., Bedny et al., 2008, 2012, 2019), but our findings suggest semantic knowledge of motion verbs might not be enough to map the sounds of locomotion to manner verbs. In addition, blind speakers had almost no manner gestures except for a very few cases where they represented manner of motion bodily—e.g., imitating a person running using the upper body. The lack of manner in the speech and gesture production of blind individuals could be the result of lack of visual experience; perhaps it is harder to learn manner distinctions from auditory input. However, there is an alternative possibility: Turkish is a verb-framed language and sighted Turkish speakers tend to omit manner more often than speakers of satellite-framed languages, such as English (e.g., Kita & Özyürek, 2003; Özçalışkan et al., 2016b, 2018; Slobin, 1996; Talmy, 1985). So, the paucity of manner in the speech and gesture of blind participants could be the result of language statistics, rather than lack of perceptual access. Further studies could disentangle these possibilities by examining how manner expressions are modulated by both visual experience and language typology, particularly in manner-dominant languages (i.e., satellite-framed languages, such as English).

The comparison of blind and blindfolded speakers enabled us to differentiate the effect of momentary lack of vision from the long-term effect of blindness. Even though blind participants differed from blindfolded participants, there were cases when the blindfolded group was indistinguishable from the blind and sighted groups, while the blind and sighted group differed from each other (e.g., in the use of path in speech and pointing gestures). This could suggest an additional role of momentary lack of vision in the expression of spatial language (see also Mamus et al., 2019), however, additional research is needed to establish this definitively.

Finally, the gestures of congenitally blind speakers offer fresh insights into multimodal language production theories. Our results showed that both blind and sighted speakers' gesture patterns were in line with what we would expect considering the typology of a verb-framed language, i.e., Turkish (e.g., Kita & Özyürek, 2003; Özçalışkan et al., 2016b, 2018; Ter Bekke et al., 2022). All speakers gestured more about path than manner of motion. This supports claims that language typology is the determining factor in co-speech gesture production, even in blind speakers (e.g., Özçalışkan et al., 2016b, 2018). Moreover, the alignment between blind people's speech and gesture (i.e., more landmark mentions with more pointing to landmarks and reduced manner mentions with fewer manner gesture) is in line with integration theories of speech and gesture (e.g., Kita & Özyürek, 2003). The fact that blind people had fewer iconic gestures overall than non-blind people is also in line with theories highlighting the role of visuo-spatial imagery underlying iconic gesture production (e.g., Hostetter & Alibali, 2008, 2019). Possibly, co-speech gesture derives partly from language typology and partly from visuo-spatial imagery (Kita & Özyürek, 2003).

Conclusion

Theories of embodied cognition propose that multimodal language processes are rooted in sensory and motor experience (Barsalou, 2016; Hostetter & Alibali, 2008; Pouw et al., 2014; Wilson, 2002). There is also substantial evidence that spatial cognition differs between blind and sighted people (Cattaneo & Vecchi, 2011; Lessard et al., 1998; Röder et al., 1999; Ruggiero et al., 2021; Voss et al., 2004). Thus, lack of visual experience may shape spatial language via altered spatial cognition. In line with this, we find differences in spatial language use in response to auditory motion events experienced by the blind and sighted individuals. To disentangle the effects of a lifetime experience of being blind versus task-specific effects of experiencing a motion event by sound alone, we included a third condition of sighted individuals who were blindfolded during the task.

Overall, we found blind people were more likely to mention landmarks, especially those in relation to themselves, than both sighted and blindfolded people. They were also more likely to mention path of motion in speech than sighted people, while omitting manner in both speech and gesture. However, based on our current data we cannot rule out the possibility that blind speakers of a satellite-framed language may show more resilience in extracting manner information from sound. Whilst the verbal encoding of path and manner did not differ between blind and blindfolded people, the differences in the gestural encoding of path and manner distinguished blind people from both sighted and blindfolded people. This suggests that beyond merely a temporary lack of sight, a lifetime of blindness changes how these components are represented in gesture. This may be because iconic gestures are more difficult to build upon non-visual information alone.

Although the current data illustrates differences between blind and sighted people, it remains unclear whether the differences in language use occur because blind people's lifetime of perceptual experience influences their conceptualization of spatial events or because blind people extract event information from auditory input for linguistic expressions differently than sighted people. Further research on blind people's language use is needed to uncover precisely how perceptual experience shapes multimodal language.

Taken together, our study illustrates that lack of visual experience affects how people encode spatial events for multimodal language production.

Acknowledgements

We wish to thank Yeşim Özüer, Esma Savaş, and Şevval Çetinkaya for transcription and annotation, and Simge Öksüm for the reliability coding. We also wish to thank Özcan Vardar for technical support during stimulus preparation, Jeroen Geerts for processing video data, and Maarten van den Heuvel for helping with the experimental setup. In addition, we wish to thank two anonymous reviewers who contributed valuable feedback and suggestions on an earlier version of this manuscript.

4

Gestures reveal how visual experience shapes concepts in blind and sighted individuals

This chapter is based on:
Mamus, E., Speed, L. J., Ortega, G., Majid, A., & Ozyurek, A. (revised & resubmitted). Gestures reveal how visual experience shapes concepts in blind and sighted individuals.

Abstract

To what extent experience influences conceptual representations is an ongoing debate. This pre-registered study had a novel approach and tested whether visual experience affects how single concepts are mapped onto gestures rather than only words. Theories claim gestures arise from sensorimotor simulations, reflecting gesturers' experience with objects. Thirty congenitally blind and 30 sighted Turkish speakers produced silent gestures for concepts from three semantic categories that rely on motor (manipulable objects) or visual (non-manipulable objects and animals) experience to different extents. As an ancillary measure of conceptual knowledge, participants listed features for the same concepts. Blind individuals were less likely than sighted individuals to produce a gesture for non-manipulable objects and animals, but not for manipulable objects. Compared to sighted, their gestures relied less on strategies depicting visuospatial features—i.e., tracing of an object (drawing) and embodying a non-human entity (personification). In the language-based task, however, the two groups differed only in the number of perceptual features listed for animals, but not the other categories. Our results suggest gesture might be driven directly by mappings of visuospatial and motoric representations onto body that are not fully accessible through listing features of concepts. Thus, gesture can provide an additional window into conceptual representations, which is not always evident in words.

Introduction

Theories of concepts differ in how they describe the relationship between multimodal sensory experience and mental representations. Traditionally, two contrasting views have been proposed: embodied theories posit concepts are rooted in sensorimotor simulations (e.g., Barsalou, 2016) while amodal theories instead claim abstract, modality-independent representations are necessary for conceptual representation (see Meteyard et al., 2012 for a review). Recent theories, however, have suggested that concepts can rely on both simulation and distributional linguistic processing (e.g., Connell, 2019), but the precise role of each remains underspecified. It therefore remains an ongoing matter of debate to what extent sensory experience shapes conceptual representations. To address these issues, previous studies have compared blind and sighted people's conceptual representations, mostly using language-based measures such as feature listing as a window into concepts (Bedny et al., 2019; Connolly et al., 2007; Kim et al., 2019a, 2021; Landau & Gleitman, 1985; Lenci et al., 2013).

It is now well established that gestures during speech or when no speech is allowed (i.e., silent gestures) are also potential windows into conceptual representations. Gesture provides a unique tool to examine conceptual representations as an alternative to commonly used language-based measures such as semantic judgments (e.g., Bedny et al., 2019; Connolly et al., 2007; Kim et al., 2021) and feature norms (e.g., Lenci et al., 2013; McRae et al., 2005; Vigliocco & Vinson, 2007), which typically require more conscious deliberation from participants.

Multimodal language theories with a relatively embodied perspective claim that gestures arise from sensorimotor simulations underlying concepts (e.g., Goldin-Meadow & Beilock, 2010; Hostetter & Alibali, 2008, 2019). As such, gesture forms are thought to reflect gesturers' specific sensorimotor experience with objects and events (for empirical support of this prediction, see Beilock & Goldin-Meadow, 2010; Cook & Tanenhaus, 2009; Pouw et al., 2020). Therefore, another way of assessing how visual experience shapes conceptual representations can be to investigate gesture forms used to depict different kinds of concepts. In the present pre-registered study, we specifically tested whether differential visual experience (i.e., congenital blindness) affects how different semantic concepts are mapped onto gestures. As an ancillary language-based measure, we also compared blind and sighted people's feature listings for the same set of concepts.

Numerous studies have investigated blind people's mental representations of concepts using mainly language-based measures, such as for color (Connolly et al., 2007; Kim et al., 2021; Landau & Gleitman, 1985; Marmor, 1978; Saysani et al., 2018; Shepard & Cooper, 1992), other visual properties, such as light emission (Bedny et al., 2019; Landau & Gleitman, 1985; Lenci et al., 2013), animals (Kim et al., 2019a), and various concrete and abstract concepts from different semantic classes (Crollen & Collignon, 2020; Lenci et al., 2013). Blind people appear to acquire considerable knowledge about concepts through indirect experience derived from language—such as learning objects' colors and the meaning of vision verbs like *look* and *see* (e.g., Landau & Gleitman, 1985). Many studies found no differences between blind and sighted people's semantic judgments (Bedny et al., 2019; Kim et al., 2021; Landau & Gleitman, 1985; Lewis et al., 2019; Marmor, 1978; Ostarek et al., 2019; Saysani et al., 2018), suggesting the absence of a sensory modality, such as vision, has no demonstrable effect on conceptual knowledge, even for concepts that are primarily related to vision. This supports the proposal that sensory information can be learned from language alone (see Bedny & Saxe, 2012, for a review). Accordingly, blind people could represent visual content in the same way that sighted people do.

However, other studies have revealed qualitative differences in the conceptual representations of blind and sighted people (Connolly et al., 2007; Kim et al., 2019a; Lenci et al., 2013; Marques, 2010; Shepard & Cooper, 1992). For example, Lenci et al. (2013) collected feature norms using a feature listing task for concrete and abstract concepts by asking blind and sighted people to define words using core features of concepts. For concrete concepts (e.g., *pencil*), blind people reported fewer perceptual features (e.g., *being cylindrical*) but more contextual features (e.g., *paper*) than sighted people. Connolly et al. (2007) also found that despite having knowledge about visual properties, such as object color (e.g., *apples are red*), blind people did not use this information to decide whether an apple was more similar to a strawberry or banana. More recently, Kim et al. (2019a) found that blind people rely more on taxonomic knowledge than sighted people to reason about animal appearance, including attributes such as size, shape, and color. They suggested blind people have broadly similar categories to sighted people, but differ in the fine-grained details, especially in properties that cannot be explored non-visually (through touch, for example). Together, these studies suggest that some visual features of concepts are less accessible to blind people. This implies visual experience can affect conceptual representations to some extent.

When it comes to gestures, research has shown that different types of sensorimotor experience could affect gestures in different ways. This can be seen in the different types of gestural strategies used to depict concepts. Studies examining the gestural representation of single semantic concepts (e.g., objects) in co-speech (Masson-Carro et al., 2016, 2017) or silent gestures (Brentari et al., 2015; Ortega & Özyürek, 2020a, 2020b; Padden et al., 2015; van Nispen et al., 2017) have revealed regularities in the gestural strategies used by sighted people. Concepts that trigger motor experience—such as manipulable objects, like tools—result in the use of an acting strategy (i.e., the reenactment of a bodily action with or without an object). Conversely, when there are limited ways to manipulate referents such as for non-manipulable objects and animals, sighted people tend to depict their visual appearance through a drawing strategy (i.e., tracing the outline of an object) or a personification strategy (i.e., embodiment of a non-human entity by mapping its movement onto one's body), respectively. In addition, a representing strategy (i.e., hands representing the partial or full form of an object) is also used across semantic categories, albeit infrequently (see Figure 1 for examples). Thus, if gesture forms arise from sensorimotor simulations, it is likely that motor experience contributes to the use of the acting strategy, while visual experience contributes to the use of drawing and personification strategies (Ortega & Özyürek, 2020a, 2020b). Notably, similar strategies for certain semantic categories are observed across communities (e.g., Dutch and Mexican; Ortega & Özyürek, 2020b).

If gestures for concepts are shaped by different types of sensorimotor experience, they can also offer a new window into blind and sighted people's conceptual representations. This can be assessed by examining how both groups use these different strategies of gestures. This, however, has not been tested before. Additionally, it is not clear to what extent the results from gesture compare with the results of feature listings, a language-based measure of conceptual representations.

Figure 1. Different strategies in silent gesture using (a) drawing for *chimney*, (b) personification for *lion*, (c) acting for *spoon*, and (d) representing for *plane*, from left to right.

The present study

In this study we used a silent gesture task to investigate differences in gesture between blind and sighted people for concepts from three semantic categories (i.e., manipulable objects, non-manipulable objects, and animals) that are known to elicit different strategies for gestures. Additionally, we used a feature listing task as a language-based measure to explore to what extent the results from both tasks are comparable.

If visuospatial and motor cues drive gesture (Hostetter & Alibali, 2008; Masson-Carro et al., 2016; Ortega & Özyürek, 2020a; van Nispen et al., 2017), then we expected to find fewer gestures in the blind than sighted group for concepts that rely more on visual (i.e., non-manipulable objects and animals) than motor information (i.e., manipulable objects). That is, we predicted an interaction between the visual experience of participants and the semantic category they had to gesture. Specifically, we expected fewer gestures with the drawing strategy for non-manipulable object concepts and fewer gestures with the personification strategy for animal concepts in blind compared to sighted people. We did not expect to find a difference between blind and sighted people in the number of gestures with the acting strategy for manipulable object concepts. Lastly, as the representing strategy is rarely observed among sighted people in earlier studies and is not typically used for any specific semantic category (e.g., Ortega & Özyürek, 2020b), we did not have predictions regarding these gestures.

We also entertained the possibility that gesture production does not differ between blind and sighted people for any type of concept. Beyond the single concept level, earlier studies cross-linguistically investigating blind people's gesture for motion events found similar language-specific patterns for path and manner expressions in both speech and gesture across blind and sighted people (Özçalışkan et al., 2016a, 2016b, 2018, but see Mamus, Speed, Rissman, et al., 2023). In this scenario several explanations are possible. It could be that there are no differences because blind and sighted people have similar conceptual representations: blind people are able to extract relevant conceptual information from non-visual sensory experience and through language input (e.g., Bedny et al., 2019; Kim et al., 2021; Landau & Gleitman, 1985; Lewis et al., 2019). Alternatively, there could be conceptual differences between blind and sighted people, but the silent gesture task may have its own limitations—due to the constraints of visual-manual depiction—to detect some differences. For example, salient features depictable in gesture may not be sensitive to detect some features such as colors that might differ across groups.

Considering the manual affordances of gesture for depiction of certain concepts and to explore how the results from gesture compare with the results of a language-based measure, we used a feature listing task as an additional measure of conceptual representation. Here participants simply had to provide any features they consider important for each concept. Feature listing is an established tool to gain insight into conceptual representations as produced features can reveal different aspects of representations from a variety of modalities and are useful to test predictions of semantic representation theories (McRae et al., 2005). However, features are language based and not all aspects of concepts are verbalizable. We therefore cannot assess exactly if gestures and features tap into the same aspects of conceptual representations. However, these two measures enabled us to have different insights into conceptual representations. While gestures can reflect visuospatial and motoric aspects of concepts, a range of information beyond these can be detected via features.

For the feature listing task, based on an earlier feature norming study with blind and sighted people (Lenci et al., 2013), we expected to find fewer perceptual (e.g., features related to size, shape, appearance, sound, kinematic information, and so on) and more non-perceptual features (e.g., how an object is used or its purpose, taxonomic categories, and encyclopedic information such as object substance, and animals' habitat and diet) of concepts reported by blind compared to sighted people. We also expected an interaction between visual experience and semantic category on the frequency of perceptual and non-perceptual features, such that differences between blind and sighted people would be larger for the non-manipulable objects and animals than manipulable objects. However, it is also possible that we might not obtain converging evidence of how blindness influences conceptual representations from the gesture and feature listing results.

Method

Participants

As pre-registered, 30 congenitally blind (11 Female, *M* = 32 years, *SD* = 14.14, range = 19–52) and 30 sighted (14 Female, *M* = 34 years, *SD* = 10.21, range = 18–57) native Turkish speakers were recruited online. At the time of testing, 18 blind participants had light perception and 12 had total blindness (see Table S1 for detailed characteristics of the blind participants, provided as supplementary data at https://osf.io/6j7xr/?view_only=6417551c5125408fabcac75985e590c4). Thirty sighted participants with normal or corrected-to-normal vision were matched for age, gender, and education to blind participants. They all were paid the equivalent of €10 in Turkish Lira for their participation and provided online informed consent approved by the IRB committee of Radboud University.

Stimuli

There were 60 experimental items in total, 20 per semantic category: manipulable objects (e.g., spoon), non-manipulable objects (e.g., bridge), animals (e.g., dog), and 4 practice items (i.e., ice cream, tree, book, and penguin, respectively). The item “castle” in the non-manipulable object category had to be removed in the final analyses, as Turkish *kale* is a homonym of “goal” and participants did not interpret the intended meaning. This left 19 non-manipulable object concepts.

We initially selected 239 concepts for manipulable and non-manipulable objects, and animals from the silent gesture database produced by Ortega and Özyürek (2020a), other gesture studies (Masson-Carro et al., 2016, 2017; van Nispen et al., 2017), and neuroimaging studies of semantic concepts (He et al., 2013; Peelen et al., 2013). In order to select 60 out 239 items, we collected sensorimotor ratings from a separate group of Turkish sighted participants following Lynott et al. (2020). Participants rated each concept on a scale from 1 to 5 on the extent to which a particular concept is experienced through six different sensory modalities (touch, hearing, smell, taste, sight, and introception) and five different action affordances (by performing an action with mouth/throat, hand/arm, foot/leg, head, and torso). The aim was to use touch and hand action ratings to operationalize manipulable vs. non-manipulable objects.

We divided the 239 concepts into three lists, and each participant rated only one list online via Qualtrics (2022, Qualtrics, Provo, UT). Each concept was rated by at least 21 participants, with a maximum of 33 participants.

To select the final set of concepts, we used the sensorimotor ratings as well as word frequency information. First, we chose items with word frequency between 0 and 300 in 1 million words from various written Turkish texts (Göz, 2003). Then, we chose 20 manipulable objects with the highest rating of tactile and hand action experience and 20 non-manipulable objects with the lowest rating of tactile and hand action experience. We replaced an item if a word (e.g., *fork* or *sailboat*) was semantically too similar to other selected words (e.g., *spoon* or *ship*). We did this to increase gesture diversity because we would expect similar gestures for objects such as spoon-fork and ship-sailboat. Furthermore, we replaced the items that we thought would be difficult to express by gesture—such as music and electricity. Finally, we chose 20 animals that had comparable word frequency with the manipulable and non-manipulable objects. A one-way ANOVA showed no difference in word frequency across semantic categories, $F(2,57) = .039$, $p = .96$. See Appendix A as supplementary data for the list of 60 concepts with average word frequency and modality ratings. Audio files were recorded by a female Turkish speaker so that they served as prompt to both blind and sighted participants. The appendices and audio files of stimuli are available at https://osf.io/6j7xr/?view_only=6417551c5125408fabcac75985e590c4.

Procedure

Participants filled out an online informed consent document containing information about the experiment and data sharing, which was sent to participants as online forms via Qualtrics. Participants first performed the silent gesture task and then feature listing task. Participants also filled out a demographic questionnaire—including questions about blindness history for blind participants.

Silent gesture task

We used the Zoom video platform to test participants due to Covid-19 regulations. The experimenter recorded the session through the screen recording feature. Participants were presented with pre-recorded spoken concepts and asked to produce silent gestures to convey single concepts. The instructions were adapted from Ortega and Özyürek (2020a). They were instructed not to speak or point at any objects around them during the task. For instance, if they heard the concept *floor* they were not allowed to point at the floor in the room. Participants were told another participant would watch their gestures later and guess their meaning.

Each trial began with a one-second beep to indicate the upcoming trial, then participants heard the audio recording of each experimental item twice. Following the second repetition, participants had 6 seconds to provide a gesture. After 6 seconds, they heard a one-second beep to indicate the end of the trial, and the initiation of the next trial. Participants performed 4 practice trials in the same order with concepts not included in the experiment. Experimental trials were presented in a different randomized order for each participant. Each session lasted approximately 15 minutes.

Feature listing task

In this task, participants were asked to list at least 5 features that are typically true of each word that they hear in the silent gesture task. The instructions were adapted from Papies et al. (2020)—see Appendix B as supplementary data for the full instructions in English and Turkish. Participants were informed that they would have one minute to list features in a written form to describe the characteristics of each concept. Two examples (i.e., *blender* and *sun*) were provided. On each trial, participants were given a summary of the instructions. This questionnaire was administered via Qualtrics. The questionnaire lasted around 60 minutes.

Coding

Gesture coding

Trials were annotated and coded by native Turkish speakers using ELAN (Wittenburg et al., 2006). We first segmented each meaningful gesture for each concept. Repetition of the same gesture was disregarded. We then classified each gesture according to their strategy (i.e., drawing, personification, acting, representing, and pointing) following Ortega and Özyürek, (2020a).

Two people coded the strategies used in gestures as (a) *drawing*, if the gesture traced the outline of a concept by hand or index finger, (b) *personification*, if the gesturer embodied a non-human animate entity or object by mapping its movement onto body, (c) *acting*, if the gesture imitated a bodily action with or without an object, (d) *representing*, if the hands represented the partial or full form of a concept, (e) *pointing*[1]*, if the gesturer pointed at an imaginary object. See Appendix C as supplementary data for our coding scheme with examples. See Figure 1 for an example of each strategy. There were also cases when one gesture included two strategies simultaneously to depict a concept. For example, to depict the concept cat,* a sighted participant depicted the ears of a cat by placing two fingers on their heads (representing strategy) while enacting a cat meowing with the mouth movements (personification strategy). As these strategies were combined in one gesture simultaneously (not consecutively), we coded them as gestures with double strategies. See § 3.1.3 for the descriptive statistics for simultaneous gestures.

1 Pointing gestures are coded only for complimentary purposes as we followed Ortega and Özyürek (2020a). As this category was not the main interest of the present study, the results are presented as supplementary data.

We calculated the Interclass Correlation Coefficient (ICC) between two coders to measure the strength of inter-coder agreement for identifying how many gestures are present in each trial (Koo & Li, 2016). Agreement in identifying the number of gestures between coders was .94. We also calculated Cohen's Kappa (Cohen, 1968) for coding type of strategy and found .89 unweighted and .91 weighted Kappa, which represent a strong to almost perfect strength of agreement.

Feature coding

The features were first classified as perceptual features or non-perceptual features. The definitions of feature types were adapted from earlier feature norming studies (e.g., McRae et al., 2005; Vinson & Vigliocco, 2008). Perceptual features capture information gained through a primary sensory channel. Perceptual features can depict among other things size, shape, appearance, sounds, body parts (e.g., *has 4 legs, has a tail*), object parts (e.g., *strap, with a handle*), and kinematic information (e.g., *runs fast*).

Non-perceptual features could be of different types, functional—how an object is used or its purpose (e.g., *used for writing*), taxonomic—including superordinate categories (e.g., *animal*, *tool*), encyclopedic—referring to object substance (e.g., *plastic*), animals' habitat and diet (*carnivore*), or object location (e.g., *found in houses*), or other for non-classified features. See Appendix D as supplementary data for the full coding scheme.

All data was coded by a native Turkish speaker. A second native speaker coded 50% of the features for reliability analysis. We found .90 weighted Cohen's Kappa for coding type of features as 2 levels (perceptual or non-perceptual), which represents a strong agreement.

Results

We pre-registered our hypotheses and analyses (available at https://osf.io/57qvn/?view_only=522ab09aa5914ba7a19dd10269db7505). Following the pre-registration, we analyzed both the gesture and feature data using linear mixed-effects and generalized linear mixed-effects regression models (Baayen et al., 2008) with the fixed factor of visual status (blind or sighted), fixed factor of semantic category (manipulable objects, non-manipulable objects, or animal), and their interaction term, together with random intercepts for participants and items, using the packages *lme4* (Version 1.1–31; Bates et al., 2015) with the optimizers *bobyqa* and *nloptwrap*, and *lmerTest* (Version 3.1–3; Kuznetsova et al., 2017) to retrieve *p*-values in R (Version 4.2.2; R Core Team, 2022). To assess statistical significance of the fixed factors and their interaction, we used likelihood-ratio tests with χ^2, comparing models with and without the factors and interaction of interest. For post-hoc comparisons and to follow-up interactions, we used *emmeans* with the Tukey adjustment (Version 1.8.2; Lenth, 2022). Data and analysis code are available at https://osf.io/6j7xr/?view_only=6417551c5125408fabcac75985e590c4.

Gesture

In the pre-registration, we predicted an interaction between visual experience and semantic category in the frequency of gesture production and strategies. Our hypotheses were: (1) blind participants would produce fewer gestures than sighted participants for non-manipulable objects and animals but not for manipulable objects, (2) sighted participants would use the drawing strategy more than blind participants for non-manipulable objects and animals, but not for manipulable objects, and (3) sighted participants would use the personification strategy more than blind participants for animals, but not manipulable and non-manipulable objects. As the acting strategy predominantly relies on motor experience for manipulable objects, we did not expect differences between the two groups for this gesture strategy. Lastly, we did not make predictions about the representing strategy because this was rarely observed among sighted people in earlier studies (Ortega & Özyürek, 2020a, 2020b).

Frequency of gesture production

We first provide descriptive statistics for the total number of gestures produced across trials and average number of gestures per concept (Table 1). The percentages of trials depicted with one or more gestures can be found as supplementary data (Table S2) at https://osf.io/6j7xr/?view_only=6417551c5125408fabcac75985e590c4.

Table 1. Descriptive statistics for blind and sighted participants across manipulable objects, non-manipulable objects, and animals.

	Manipulable objects		Non-manipulable objects		Animals	
	Blind	**Sighted**	**Blind**	**Sighted**	**Blind**	**Sighted**
N gestures	690	1070	461	841	370	699
M gesture per concept (*SD*)	1.15 (0.54)	1.78 (0.87)	0.81 (0.62)	1.48 (0.99)	0.62 (0.62)	1.17 (0.83)

Following our pre-registration, we compared whether blind and sighted people differed in how often they produced at least one gesture across semantic categories (Figure 2). We ran a glmer model on binary values for gesture presence (0 = no, 1 = yes) as a dependent variable. It revealed an effect of visual status, χ^2 (1) = 10.31, p = .001. Overall, sighted people produced more gestures than blind people, β = 1.40, *SE* = .41, z = 3.41, p < .001. The model also revealed an effect of semantic category, χ^2 (2) = 57.32, p < .001, and a significant interaction between visual status and semantic category, χ^2 (2) = 8.55, p = .014.

As hypothesized in the pre-registration, blind people produced fewer gestures than sighted people for non-manipulable objects (β = -1.23, 95% *CI* = -2.08 – -.38, *SE* = .44, *z* = -2.83, *p* = .005) and animals (β = -1.65, 95% *CI* = -2.48 – -.81, *SE* = .43, *z* = -3.86, $p <$ .001), but there was no significant difference for manipulable objects (β = -.62, 95% *CI* = -1.61 – .38, *SE* = .51, *z* = -1.22, *p* = .22). On average, blind participants skipped 2.7 trials in the manipulable objects (13.6%), 6.8 trials in the non-manipulable objects (35.6%), and 10.5 trials in the animal categories (52.5%), while sighted participants skipped 1.7 trials in the manipulable objects (8.5%), 3.3 trials in the non-manipulable objects (17.4%), and 4.5 trials in the animal categories (22.5%).

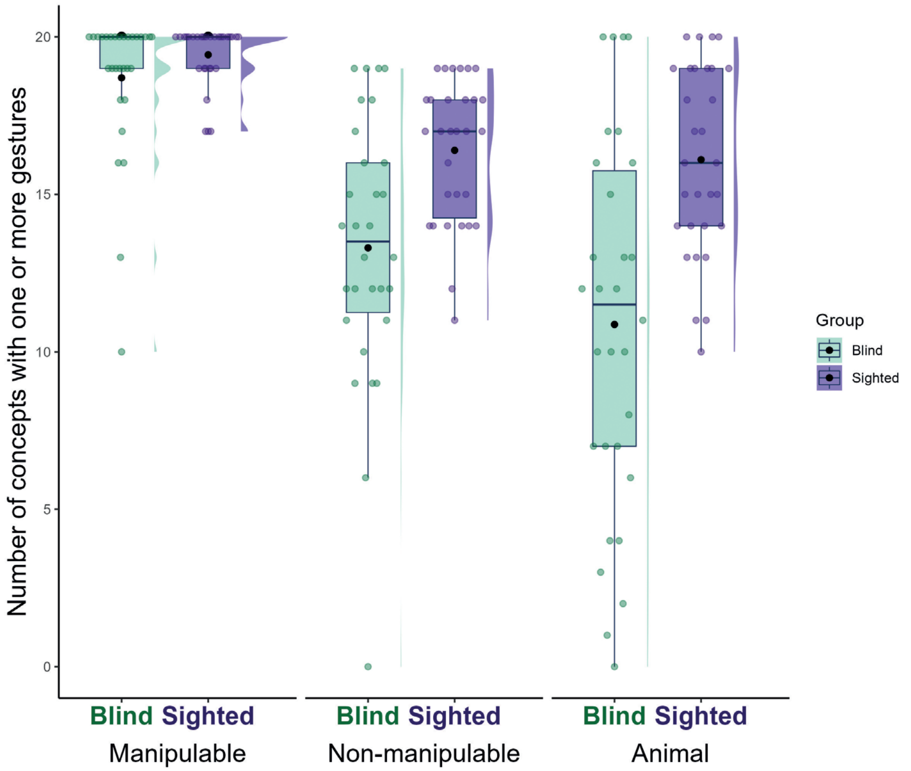


Figure 2. Frequency of gestures for manipulable object (*n* = 20), non-manipulable object (*n* = 19), and animal (*n* = 20) categories. Colored dots represent each participant. Black dots represent the group mean.

Frequency of gesture strategies

As a next step, we compared the groups in terms of the gesture strategies produced per semantic category if they gestured. We analyzed the presence of each gesture strategy (0

= no, 1 = yes) with four pre-registered glmer models (one for each gesture strategy). We present data excluding skipped trials, but note the results are the same when skipped trials are included. Figures 3–6 show the proportion of trials in which each strategy was observed (i.e., number of trials with each strategy divided by total number of trials per semantic category). We present the results for each strategy separately.

Drawing strategy. We found a significant effect of visual status on the production of gestures with the drawing strategy, χ^2 (1) = 35.25, $p < .001$ (Figure 3). Overall, blind people produced fewer gestures with the drawing strategy than sighted people, β = -2.78, *SE* = .45, *z* = -6.21, $p < .001$. The model also revealed an effect of semantic category, χ^2 (2) = 7.20, *p* = .028, and a significant interaction between visual status and semantic category, χ^2 (2) = 7.62, *p* = .022. As predicted, blind people produced fewer gestures with the drawing strategy than sighted people for non-manipulable objects (β = -2.90, 95% *CI* = -3.84 – -1.95, *SE* = .48, *z* = -6.02, $p < .001$) and animals (β = -2.13, 95% *CI* = -3.13 – -1.12, *SE* = .51, *z* = -4.15, $p < .001$). They also produced fewer gestures with the drawing strategy for manipulable objects (β = -3.29, 95% *CI* = -4.37 – -2.22, *SE* = .55, *z* = -5.99, $p < .001$). Critically, the size of the difference between blind and sighted people appears to differ across semantic categories—i.e., the effect of visual status might be the largest for the manipulable objects—which drives the interaction between visual status and semantic category.

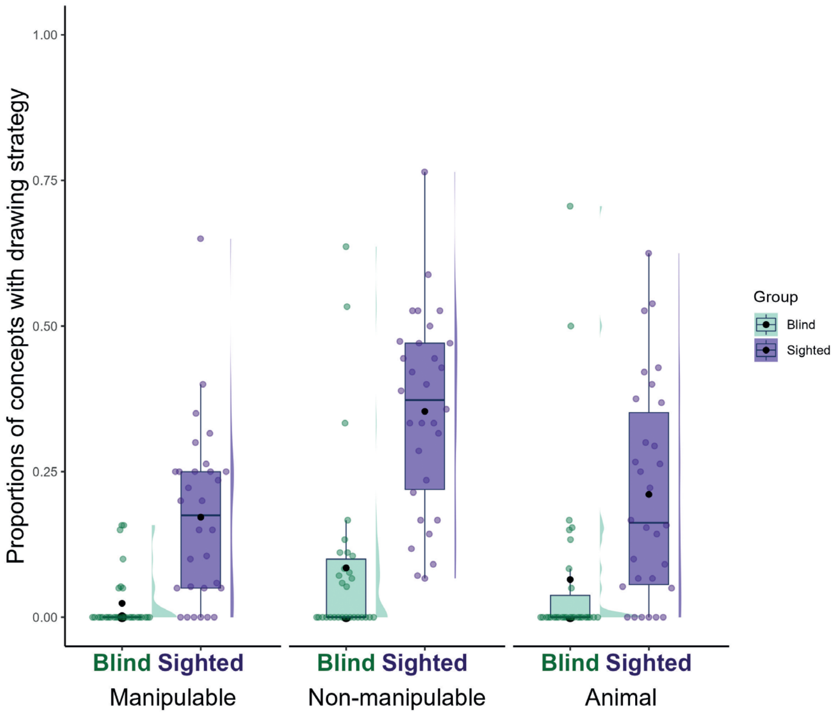


Figure 3. Proportion of trials with a drawing strategy by semantic category and group. Colored dots represent each participant. Black dots represent the group mean.

Personification strategy. Next, we compared whether blind and sighted people differed in their frequency of gesture with the personification strategy (Figure 4). We ran the same glmer model on binary values for the presence of gesture with the personification strategy (0 = no, 1 = yes) as a dependent variable. The model did not reveal an effect of visual status, χ^2 (1) = 2.53, *p* = .11, but an effect of semantic category, χ^2 (2) = 93.61, *p* < .001, and an interaction between visual status and semantic category, χ^2 (2) = 15.15, *p* < .001. As predicted, blind people produced fewer gestures with the personification strategy than sighted people for animals (β = -.94, 95% *CI* = -1.61 – -.26, *SE* = .35, *z* = -2.71, *p* = .007) but not manipulable (β = -14.16, 95% *CI* = -3156 – 3128, *SE* = 1603, *z* = -.01, *p* = .99) or non-manipulable objects (β = .32, 95% *CI* = -.45 – 1.09, *SE* = .39, *z* = .82, *p* = .42).

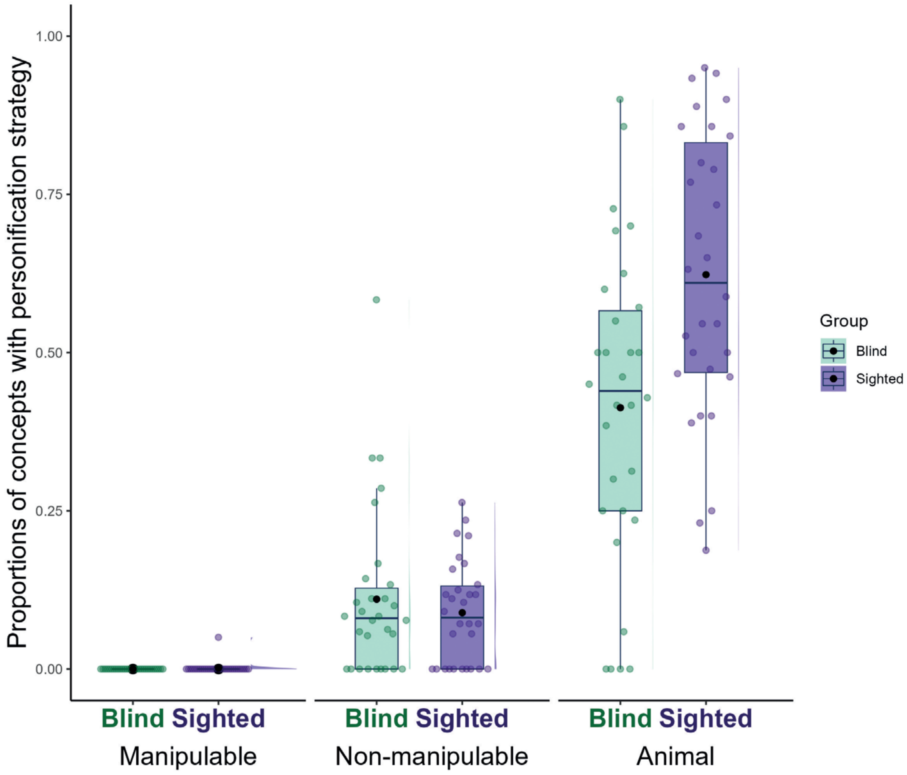


Figure 4. Proportion of trials with a personification strategy by semantic category and group. Colored dots represent each participant. Black dots represent the group mean.

Acting strategy. Although we did not expect a difference in the use of the acting strategy, we conducted exploratory analyses (Figure 5). Unexpectedly, the model revealed an effect of visual status, χ^2 (1) = 17.19, p < .001, and an effect of semantic category, χ^2 (2) = 73.63, p < .001. Overall, blind people produced fewer gestures with the acting strategy than sighted people, β = -1.21, *SE* = .27, *z* = -4.51, p < .001. Importantly, there was a significant interaction between visual status and semantic category, χ2 (2) = 7.86, p = .020. Blind people produced fewer gestures with the acting strategy than sighted people for manipulable objects (β = -1.78, 95% *CI* = -2.60 – -0.96, *SE* = .42, *z* = -4.26, p < .001) and non-manipulable objects (β = -1.28, 95% *CI* = -1.94 – -0.63, *SE* = .34, *z* = -3.83, p < .001), but not animals (β = -0.30, 95% *CI* = -1.18 – 0.58, *SE* = .45, *z* = -0.66, p = .51).

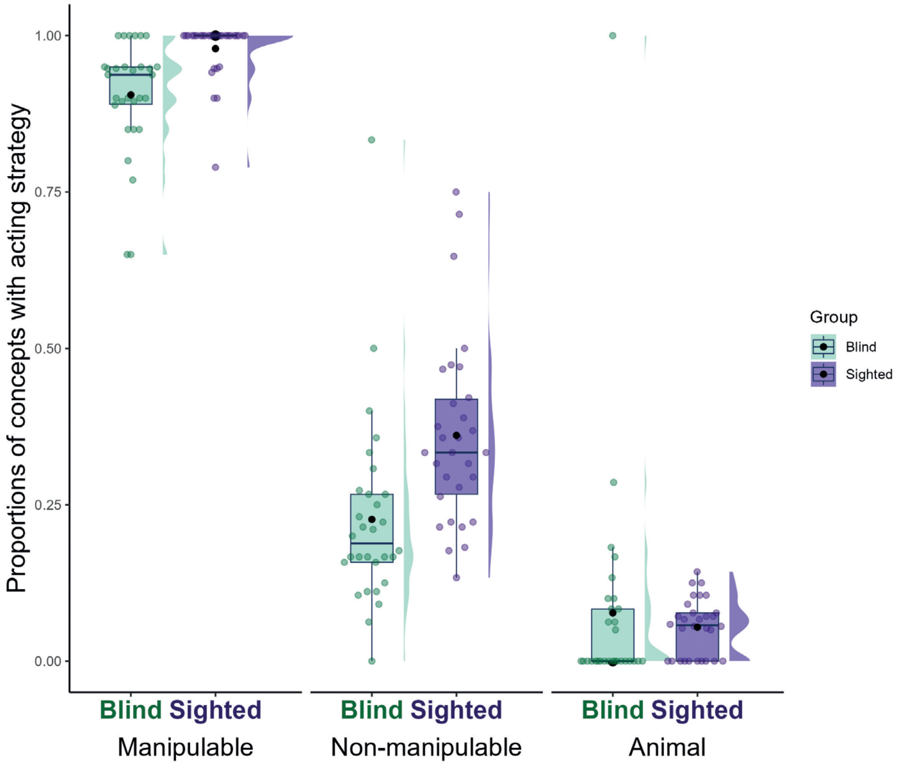


Figure 5. Proportion of trials with an acting strategy by semantic category and group. Colored dots represent each participant. Black dots represent the group mean.

Representing strategy. As we did not have any predictions about the representing strategy, we conducted exploratory analyses of this variable (Figure 6). The model revealed an effect of visual status, χ^2 (1) = 6.72, p = .010. Overall, blind people produced more gestures with the representing strategy than sighted people, β = .65, *SE* = .24, *z* = 2.66, p = .008. There was also an effect of semantic category, χ^2 (2) = 19.41, p < .001, and an interaction between visual status and semantic category, χ^2 (2) = 29.42, p < .001. Blind people produced more gestures with the representing strategy than sighted people for manipulable objects ($\beta = 1.22$, 95% *CI* = .68 – 1.77, *SE* = .28, *z* = 4.38, *p* < .001) and non- manipulable objects (β = .71, 95% *CI* = .15 – 1.27, *SE* = .29, *z* = 2.48, *p* = .013). There was no difference for animals across groups (β = -.10, 95% *CI* = -.67 – .46, *SE* = .29, *z* = -.36, *p* = .72).

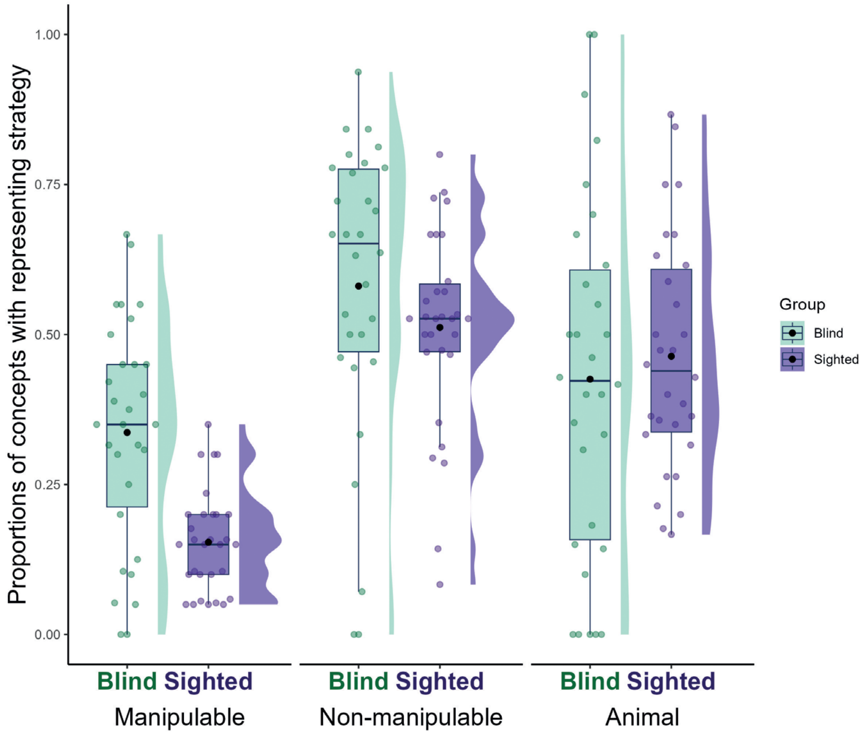


Figure 6. Proportion of trials with a representing strategy by semantic category and group. Colored dots represent each participant. Black dots represent the group mean.

Use of simultaneous strategies

We observed that participants occasionally used two strategies simultaneously to depict a concept. The descriptive statistics indicate that 26 out of 30 blind participants used at least one gesture that included two strategies simultaneously to depict manipulable objects, and 15.1% of the gestures for manipulable objects used acting and representing strategies simultaneously. For example, for *pencil*, 13 blind people depicted its elongated shape with their hands (representing strategy) while simultaneously imitating writing with a pencil (acting strategy; see Figure 7). Other manipulable objects depicted simultaneously by 10 or more blind participants were *broom, clothespin, spoon, key, match, banana,* and *bread*.

Figure 7. A blind participant produces a simultaneous gesture to depict *pencil* with acting (the right hand) and representing (the left hand) strategies.

Twenty-four out of 30 sighted participants used at least one simultaneous gesture. They used simultaneous gestures mostly to depict animals, and 6.6% of the gestures for animals used personification and representing strategies simultaneously. For example, 8 sighted people depicted the trunk of an elephant by placing their whole arm around their nose (representing strategy) while acting like an elephant moving its trunk (personification strategy).

Summary

Overall, as expected, blind people produced fewer gestures than sighted people for non-manipulable objects and animals that rely more heavily on visual information, but not for manipulable objects that rely more on motor information. In line with our predictions, compared to sighted people, blind people produced fewer gestures with the drawing strategy to depict non-manipulable objects and animals and fewer gestures with the personification strategy to depict animals. As an unexpected finding, blind people produced fewer gestures with the acting strategy but more with the representing strategy than sighted people for manipulable as well as non- manipulable objects. These findings indicate that visual experience influences how people produce gestures for concepts, supporting embodied accounts of gesture production.

In the next section, we report the results of the feature listing task to find out whether we obtain converging evidence from a language-based measure.

Features

In the pre-registration, we predicted an interaction between visual experience and semantic category on the frequency of perceptual and non-perceptual features—

see § 2.4.2 for a detailed definition of feature categories. Our hypotheses were: (1) blind participants would produce fewer perceptual features of concepts than sighted participants, (2) blind participants would produce more non-perceptual features of concepts than sighted participants, and (3) the difference between blind and sighted participants for hypotheses 1 and 2 would be bigger for the non-manipulable objects and animals than manipulable objects.

Frequency of feature production

Following our pre-registration, although we did not have a specific prediction, we first compared whether blind[2] and sighted people differed in the number of features they produced for the different semantic categories. The model did not reveal an effect of visual status, χ^2 (1) = .92, p = .34, or of semantic category, χ^2 (2) = 5.33, p = .07, but it did reveal a significant interaction between visual status and semantic category, χ^2 (2) = 27.03, p < .001. There was a trend for a difference between blind and sighted people for non-manipulable objects and animals, but this was not the case for manipulable objects (Table 2). However, the post-hoc tests did not reveal any difference between blind and sighted people for manipulable objects (β = .05, SE = .27, z = .20, p = .84), non-manipulable objects (β = -.40, SE = .27, z = -1.49, p = .14), or animals (β = -.41, SE = .27, z = -1.56, p = .12). These differences were not significant probably because we used the conservative p-adjustment in *emmeans* for having multiple comparisons.

Frequency of perceptual features

We compared whether blind and sighted people differed in the number of perceptual features (Table 2). The model did not reveal an effect of visual status, χ^2 (1) = .008, p = .93, but an effect of semantic category, χ^2 (2) = 20.73, p < .001, and an interaction between visual status and semantic category, χ^2 (2) = 63.01, p < .001. Compared to sighted people, blind people produced fewer perceptual features for animals (β = -.39, 95% CI = -.71 – .07, SE = .16, z = -2.40, p = .017) but not for manipulable (β = .30, 95% CI = -.02 – .62, SE = .16, z = 1.85, p = .064) or non-manipulable objects (β = -.05, 95% CI = -.27 – .37, SE = .16, z = .31, p = .76).

2 Two blind participants did not respond to the feature listing questionnaire. To check whether the missing data from the blind group affected the results, we ran the analyses excluding two sighted participants who were tested as matched controls with those two blind participants. The results did not change, therefore in the main text we report the analyses including the data of two sighted participants.

Table 2. Average number of all, perceptual, and non-perceptual features produced by blind (n = 28) and sighted (n = 30) participants across manipulable objects, non-manipulable objects, and animals. Significant differences across groups are indicated in bold.

	Manipulable objects		Non-manipulable objects		Animals	
	Blind	**Sighted**	**Blind**	**Sighted**	**Blind**	**Sighted**
M all features (*SD*) [Range]	5.23 (1.16) [2.4–7.3]	5.18 (0.81) [3.8–7.6]	4.69 (1.26) [1.6–7.3]	5.09 (0.86) [3.4–7.3]	5.03 (1.35) [2.3–8.1]	5.45 (0.84) [4.0–7.2]
M perceptual (*SD*) [Range]	1.67 (0.74) [0.2–3.6]	1.37 (0.58) [0.5–2.7]	1.28 (0.63) [0.1–2.7]	1.23 (0.44) [0.5–2.5]	**1.89 (0.72) [0.2–3.7]**	**2.28 (0.72) [0.9–4.2]**
M non-perceptual (*SD*) [Range]	3.45 (1.11) [1.7–5.5]	3.75 (0.64) [2.8–5.3]	**3.26 (1.42) [1.2–5.9]**	**3.77 (0.73) [2.6–5.7]**	3.04 (1.22) [1.0–5.3]	3.12 (0.75) [2.1–4.9]

Additionally, we wanted to check whether perceptual features reported by blind and sighted people differed qualitatively although they did not differ much quantitively. So, we examined the top 10 perceptual features per semantic category—see Table A in the Appendix. We found in terms of the types of features produced, there was also substantial overlap across groups, but blind people were more likely to mention sound within the top 10 perceptual features, whereas sighted people mentioned color.

Frequency of non-perceptual features

We compared whether blind and sighted people differed in the number of non-perceptual features produced (Table 2). The model did not reveal an effect of visual status, χ^2 (1) = 1.60, p = .21, but an effect of semantic category, χ^2 (2) = 10.37, p = .006, and an interaction between visual status and semantic category, χ^2 (2) = 15.68, p < .001. Blind people produced fewer non-perceptual features than sighted people for non-manipulable objects (β = -.51, 95% *CI* = -.98 – -.03, *SE* = .24, *z* = -2.10, *p* = .036), but not manipulable objects (β = -.31, 95% *CI* = -.78 – -.16, *SE* = .24, *z* = -1.29, *p* = .20) or animals (β = -.08, 95% *CI* = -.55 – .39, *SE* = .24, *z* = -.34, *p* = .74).

To find out if the features differed qualitatively across groups, we examined the top 10 non-perceptual features per semantic category—see Table B in the Appendix—and there was again substantial overlap across groups. We also conducted exploratory analyses for different types of non-perceptual features (i.e., functional, taxonomic, and encyclopedic), which can be found as supplementary data at https://osf.io/6j7xr/?view_only=6417551c5125408fabcac75985e590c4. In sum, blind and sighted people did not differ in the number of functional and taxonomic features produced in any semantic category, but blind people produced fewer encyclopedic features than sighted people for non-manipulable objects.

Discussion

In the present study, we explored whether visual experience shapes how single concepts are mapped onto gestures as a potential window into conceptual representations. We compared congenitally blind and sighted people's silent gestures for concepts that rely primarily on use of visual (non-manipulable objects and animals) versus motor information (manipulable objects). As an ancillary measure of conceptual representations, we also collected language-based feature lists for the same concepts—a traditional method to gain insight into conceptual representations (McRae et al., 2005; Vinson & Vigliocco, 2008). Gestures reflect visuospatial and motoric aspects of concepts while written features can provide a larger range of information beyond visuospatial and motoric aspects. This way we aimed to gain different insights into conceptual representations using both the silent gesture and feature listing task.

In the silent gesture task, as predicted, we found blind people produced fewer gestures than sighted people for non-manipulable objects and animals, but not for manipulable objects. Across the board, blind people skipped more trials when concepts relied more on visual than motor information.

Regarding the gesture strategies, blind and sighted people were similar in which strategy they mostly used for each semantic category—i.e., the acting strategy for manipulable objects and the representing strategy for non-manipulable objects. For animals, though, sighted people mostly used the personification strategy while blind people used the personification and representing strategies to similar extents. Although the general tendency in the semantic categories was similar across groups, there were differences in the frequency of strategies used by blind and sighted people. As predicted, we found that compared to sighted people, blind people produced fewer gestures with the drawing strategy for non-manipulable objects and animals and fewer gestures with the personification strategy for animals. The drawing strategy relies on the overall shape (i.e., the outline) of objects, while the personification strategy relies on mapping visual and kinematic features of non-human animate entities onto producers' body. Thus, visual experience seems critical to depict these aspects of objects in gestural form. This shows that blind people might have less access to these aspects of conceptual representations, to the extent they can be expressed in gesture.

One unexpected finding was that blind people produced fewer gestures with the acting but more gestures with the representing strategy than sighted people for manipulable objects. To depict the concept *spoon*, for example, sighted participants imitated eating with a spoon (acting strategy) whereas blind participants showed the concave shape of a spoon (i.e., a shape curves inwards) with their hands (representing strategy). The representing strategy might be more accessible to blind people because of their salient tactile experiences with objects, and this might be as important as motor experience. Although we did not predict differences for manipulable objects, the findings align with earlier work that consider gestures as outcomes of our embodied experience (Beilock & Goldin-Meadow, 2010; Cook & Tanenhaus, 2009; Goldin-Meadow & Beilock, 2010; Hostetter & Alibali, 2008, 2019; Pouw et al., 2020). Moreover, blind people used the acting and representing strategies simultaneously in 15% of their gestures depicting manipulable objects. One might argue that simultaneous depictions of different strategies were more informative (for example, to differentiate the action of *writing* from *pencil*—the object of that particular action) than combining two strategies consecutively as the former taps into the meaning of a concept more directly (Emmorey, 2014)—see also Slonimska et al. (2020, 2022) for supporting experimental evidence from sign languages. This should be tested in the future work.

Although our main interest is the comparison of gesture strategies used across groups, we should also note that our results align with some of the earlier findings showing regularities across sighted people in the strategies used for different semantic categories (Ortega & Özyürek, 2020a, 2020b; van Nispen et al., 2017). For example in the sighted group, acting is the most preferred strategy for manipulable objects while personification is the most preferred strategy for animals. Ortega and Özyürek (2020b) found these preferred strategies in Dutch and Mexican Spanish, and we replicate them here in Turkish, an unrelated language. However, regarding non-manipulable objects, Turkish sighted speakers mostly preferred the representing strategy unlike the findings of Ortega and Özyürek (2020b) showing that Dutch and Mexican Spanish speakers mostly preferred the drawing strategy to depict non-manipulable objects. This suggests there are both universal and cultural patterns for mapping semantic categories to visual-manual expressions.

Unlike the differences in the gesture task, in the feature listing task we found no differences across groups except that compared to sighted people, blind people produced fewer perceptual features (i.e., features related to size, shape, appearance, kinematic information and so on) for animals and fewer non-perceptual features for non-manipulable objects. Although the frequency of features differed across groups for these two categories, there was substantial overlap in terms of perceptual and non-perceptual features listed. The most distinguishable difference was that blind people were more likely to mention sound, whereas sighted people mentioned color more. For example, while sighted people frequently reported specific colors (i.e., *yellow*, *red*) of concepts, blind people only mentioned that concepts are *colorful*. We expected to find fewer perceptual features, especially for non-manipulable objects and animals, in the blind than sighted group, but this was only supported for the animal category. This finding is in line with an earlier study showing that blind and sighted people share substantial knowledge of animal appearance, except for colors (Kim et al., 2019a).

Considering our results together, we find that the results from gestures and features produced by blind and sighted people mostly do not overlap, apart from some convergent patterns in the animal category. The gesture results indicate that visual experience influences how people produce gestures for concepts, even for manipulable objects to some extent. Overall, blind people rely less on strategies depicting visual information (i.e., drawing and personification) but more on strategies depicting motor and tactile information (i.e., acting and representing). These adaptations in gesture forms might be a result of differences in conceptual representations gained through perception without vision, supporting the embodied cognition framework (e.g., Barsalou, 2016). Yet, the feature results only partly align with the gesture results, specifically for animal concepts. Thus, we found converging evidence from both tasks that blind and sighted people differ in their representation of animal concepts. However, there was no converging evidence from the gesture and feature lists for manipulable and non-manipulable objects as listed features for these categories were mostly shared while gestures differed across groups.

Together, our results suggest gestures can provide an additional window into conceptual representations as they are driven directly by iconic mappings of visuospatial and motoric representations (Hostetter & Alibali, 2008, 2019), which may not be fully accessible through feature listing because some iconic aspects of concepts may be difficult to verbalize. Earlier studies using language-based measures, such as semantic judgements, usually report no difference between blind and sighted people (Bedny et al., 2019; Kim et al., 2021; Landau & Gleitman, 1985; Lewis et al., 2019; Marmor, 1978; Ostarek et al., 2019; Saysani et al., 2018). This might arise from the fact that such methods are more limited in revealing fine-grained visuospatial and motoric aspects of concepts. For example, both blind and sighted people mention that lions roar, yet verbal description of how a lion moves in order to roar is not simple. However, a gestural description could easily inform us about how people represent a roaring lion (see Figure 1b), suggesting gestures provide different insights into conceptual representations.

An alternative explanation of our gesture findings is that observing other people's gestures may be crucial to learn how to use certain strategies for certain concepts so that gestures are interpretable by others. People often understand what a newly constructed gesture may mean thanks to iconicity (i.e., similarity between the form and the meaning of a referent; Perniss et al., 2010; Taub, 2001) and also to systematicity in the forms of gestures people produce (van Nispen et al., 2017). This gestural experience may ultimately drive the systematicity in silent gestures (see also van Nispen et al., 2017). Thus, even if blind people have similar conceptual representations to sighted people, they still need to transform those representations into gestures, for example, to depict the shape of a tree, so that other people understand the meaning of the gesture. Therefore, visual experience might have an effect on gesture production through differences in communication experience.

In the present study we did not measure the systematicity and comprehensibility of gestures but observed cases such that even when both blind and sighted people use the same gestural strategy for depicting a concept, the content of the gesture differed. For example, two participants produced an acting gesture to depict the concept *bread*, but the sighted gesturer imitated cutting bread whereas the blind gesturer imitated smelling bread. If the systematicity of gestures drive successful comprehension as suggested earlier (Ortega & Özyürek, 2020b; van Nispen et al., 2017), blind people's gestures might be less comprehensible than sighted people's gestures. Indeed, recent work has shown that success in interpretation of gestures is greater for sighted than blind gesturers (Fay et al., 2022), suggesting this may be a fruitful line of inquiry to explore in future research.

Conclusion

Taken together, our study illustrates that visual experience shapes single concepts expressed in gestures reflecting an individual's sensorimotor experience with objects in line with embodied theories of gesture production. Conceptual differences that can be observed across groups through gestures may not be fully accessible via language-based methods. Gestures, therefore, provide additional insight into conceptual representations through direct mappings of visuospatial and motoric aspects of concepts. In conclusion, visual experience influences how concepts are mapped onto gestural and, to some extent, verbal expressions. We believe our current methodology and results open new avenues of research into links between blind and sighted individuals' experience and their language, cognition, and communication.

5

General discussion

General Discussion

In the current thesis, I explored the role of perceptual experience on conceptual representations and multimodal language production through the examination of speech and gesture in both blind and sighted individuals. Specifically, I tested whether the perceptual modality of input (i.e., visual, auditory, or audiovisual) influenced the way people encode motion events in speech and gesture (Chapter 2). Then, I asked whether visual experience (i.e., being sighted or congenitally blind) shaped how people encode spatial events for multimodal language production (Chapter 3) as well as how object concepts are mapped onto silent gestures (Chapter 4). Together, the results presented in this thesis showed that:

1) Perceptual modality of input affects how sighted speakers encode motion events in their speech, in line with the spatial affordances of visual and auditory inputs, but has no influence on gesture type or frequency.

2) Visual experience influences the encoding of motion event components (i.e., path, manner, source, and goal) in both speech and gesture, in line with posited differences underlying spatial cognition in blind and sighted individuals.

3) Visual experience shapes how object concepts are represented in silent gestures, and blind and sighted individuals produce gesture forms which are influenced by their own sensorimotor experience, in line with an embodied cognition framework.

These findings overall suggest that perceptual experience—particularly visual experience—shapes how concepts are expressed in multimodal language. This chapter begins with a summary of the key findings from Chapters 2–4 for the readers' convenience. This is followed by a separate section that provides a comprehensive discussion of the findings' broader theoretical and methodological implications. Additionally, I propose potential avenues for future research and conclude with final remarks.

Summary of key findings

In **Chapter 2**, I investigated whether the perceptual modality of input (i.e., visual, auditory, or audiovisual) influences the way people encode spatial events in speech and gesture. Multimodal language production theories share a common assumption that gestures arise (at least partially) from visuospatial imagery (de Ruiter, 2000, 2007; Hostetter & Alibali, 2008, 2019; Kita, 2000; Kita & Özyürek, 2003; Krauss et al., 2000; McNeill, 1992; McNeill & Duncan, 2000). However, gesture is commonly studied with only visual input, such as video clips and cartoons (e.g., Akhavan et al., 2017; Gullberg et al., 2008; Kita & Özyürek, 2003; Ter Bekke et al., 2022), which leaves the theoretical assumption concerning the nature of imagery underlying gesture production open to scrutiny. To address this, I compared speech and co-speech gestures of Turkish sighted speakers, focusing on the path and manner of motion events which were presented with different perceptual inputs. I used motion events as a testing ground since there is a substantial amount of prior research on speech and gesture production that served as a foundation for this study. Events were presented in three different conditions: audio-only, visual-only, and multimodal (visual + audio). While my main objective was to compare the audio-only and visual-only conditions, I also included the multimodal condition to investigate if the combination of both types of information enhanced spatial language production.

The results of **Chapter 2** revealed that, as predicted, when speakers watched events (in the visual-only or multimodal condition), they produced more motion event descriptions compared to when they only listened to the events (audio-only condition). I observed no statistically significant difference in the amount of motion event descriptions between the visual-only and multimodal conditions. This suggests the addition of auditory information does not further enrich speakers' descriptions of motion events when visual information is available. Furthermore, as predicted, speakers in the visual conditions mentioned manner of motion more often and path of motion less often than speakers in the auditory condition. Contrary to my expectations, however, gesture frequency for path and manner did not differ between the audio-only and visual conditions. Speakers, regardless of the input modality, produced more path-only gestures than manner-only and path + manner gestures, which is consistent with the typology of Turkish. Together, the findings suggest the perceptual modality of input affects how speakers encode path and manner of motion events in speech, but it has no influence on gesture. I discuss the broader implications of the findings below in the ***Theoretical implications*** section.

After establishing the role of perceptual modality on the multimodal encoding of motion events of sighted speakers, in **Chapter 3** I examined whether lack of visual experience, as experienced by congenitally blind people, shapes how individuals encode spatial events for multimodal language production. To address this aim, I compared the speech and co-speech gesture of blind, blindfolded, and sighted Turkish speakers for different components of motion events. As described in Chapter 1, I presented auditory motion events to all participants, ensuring that the input modality was equated for both blind and non-blind individuals (including both blindfolded and sighted) using ecologically relevant stimuli. In addition to the comparison of path and manner descriptions in speech and iconic gesture, I examined how blind, blindfolded, and sighted speakers described landmarks (the source and goal of motion) using both speech and pointing gestures. Evidence coming from studies of spatial cognition, which examine how congenitally blind people learn spatial layouts, has shown that blind people rely more on an egocentric frame of reference (i.e., referring to locations relative to their position in space) than an allocentric one (i.e., referring to locations based on external objects irrespective of their position) (e.g., Cattaneo & Vecchi, 2011; Iachini et al., 2014). In this chapter, I examined whether this tendency of blind people would also be observed when they referred to landmarks in their motion event descriptions.

The results of **Chapter 3** showed that, as predicted, compared to non-blind speakers (both blindfolded and sighted), blind speakers showed a greater tendency to use landmarks, particularly self-anchored landmarks in their verbal descriptions (i.e., referring to locations relative to their own body, such as "to/from my left"). They also placed more emphasis on path and less on manner of motion events compared to sighted speakers. In terms of spontaneous co-speech gestures, there was no difference in the overall gesture rate between blind and sighted speakers. However, speakers' gesture frequency varied depending on the type of gesture. Blind speakers produced more pointing gestures to landmarks than sighted speakers, but had fewer iconic gestures related to path and manner compared to non-blind speakers. While all speakers' gesture patterns aligned with the Turkish motion typology (i.e., all speakers produced predominantly path-only gestures), blind speakers produced fewer iconic gestures overall. Together, the findings suggest that lifetime experience of being blind (as opposed to temporary lack of visual input, as shown in Chapter 2) influences how people encode motion events in both speech and co-speech gesture.

After comparing blind and sighted people's multimodal productions for events, in **Chapter 4**, I focused on the representation of object concepts and explored whether visual experience affects how different types of concepts are mapped onto silent gestures. Gestures, driven by iconic associations with visuospatial and motor experience, can provide a different window into some facets of concepts than language-based measures, such as semantic judgments and feature listing (e.g., Connolly et al., 2007; Kim et al., 2019a, 2021; Lenci et al., 2013), and thus offer an additional approach to study concepts in relation to traditional language-based measures. As described in Chapter 1, I investigated Turkish blind and sighted people's gestures depicting concepts (that were provided to participants as pre-recorded audio clips) from different semantic categories that rely to different extents on visual (i.e., non-manipulable objects and animals) vs. motor (i.e., manipulable objects) information. In the gesture task, different strategies used to depict concepts were coded, which were the drawing, personification, acting, and representing strategies. In the second task, which used a language-based measure, participants listed core features of the same concepts, and these were coded as perceptual (i.e., information gained through a primary sensory channel such as size, shape, appearance, sounds, body parts, and kinematic information) or non-perceptual (e.g., functional, taxonomic, or encyclopedic) features. By using both measures, I could explore the extent to which the outcomes from both measures aligned with each other.

The results of **Chapter 4** revealed that, as predicted, blind participants produced fewer gestures compared to sighted participants for non-manipulable objects and animal concepts. There was no group difference for manipulable objects, as expected.

Overall, blind participants skipped a higher number of trials when concepts were predominantly dependent on visual rather than motor information. While there was a shared trend across participants in the strategies used for each semantic category, the relative frequency of strategies differed across groups. Most critically, blind participants used fewer gestures depicting visuospatial aspects of concepts, so fewer drawing and personification strategies, compared to sighted participants. Thus, the gesture results overall showed that blind participants adjusted their gesture strategies according to their sensorimotor experience with objects, emphasizing the influence of visual experience on gesture production for concepts. Contrary to expectations, however, in the language-based task, blind participants differed from sighted participants only in the frequency of perceptual features listed for animals. Blind participants reported fewer perceptual features for animals and, unexpectedly, fewer non-perceptual features for non-manipulable objects than sighted participants. While the frequency of some features varied across groups for these categories, there was mostly a convergence in the perceptual and non-perceptual features that were listed by blind and sighted participants. Together, the findings suggest that visual experience influences how concepts are represented in gestures and, to some degree, through written features.

Theoretical implications

The primary contribution of the current thesis is that it provides new insights into the role that perceptual input plays in multimodal language and cognition through an examination of the conceptual representations and language use of blind individuals in comparison to sighted individuals. The findings of this thesis have implications for existing theoretical accounts of multimodal language production, gesture production of blind people, as well as debates about the nature of conceptual knowledge, as discussed below.

Perceptual experience shapes the multimodal linguistic encoding of events

The findings of this thesis highlight the importance of perceptual experience in the linguistic encoding of spatial events for both blind and sighted individuals. First, by systematically testing the effect of perceptual input (vision vs. audition) in multimodal language use, I showed that when visual information is present, speakers produce more linguistic information concerning spatial aspects of motion events (Chapter 2). Critically, having auditory information in addition to visual information does not further enrich the linguistic expressions of motion events. These findings support the dominance of vision shown in the perceptual lexicons of languages (Floyd et al., 2018; Levinson & Majid, 2014; San Roque et al., 2015; Viberg, 1983; Winter et al., 2018) and extend it to the domain of spatial events.

Second, the affordances of different input modalities shape how people encode distinct aspects of spatial events for linguistic expression, for example, by foregrounding some types of information over others. The visual modality foregrounds manner of motion more than path of motion in speech (Chapter 2). Speakers watching motion events talked more about how a motion was performed (e.g., someone walked/limped) whereas speakers hearing the sounds of events talked more about the trajectory of the motion (e.g., someone went further away). This is possibly due to the fact that vision offers more detailed information about how an action is performed than audition, thus making manner more salient than path for verbal descriptions. The encoding of path is also influenced by the affordances of input modality (Chapter 2). As mentioned earlier, audition provides sequential spatial information in contrast to holistic visual information (e.g., Thinus-Blanc & Gaunet, 1997). The sequential information coming from audition might force speakers to encode path information in a segmented fashion, e.g., dividing the path into smaller chunks, thus leading to more path verbs in descriptions built upon auditory input. This implies that building holistic spatial representations may be more difficult based on auditory input alone. This finding is particularly interesting in light of the other research question of this thesis that asks whether the absence of visual experience shapes how people encode spatial events for multimodal language production (Chapter 3).

By comparing the descriptions of blind, blindfolded, and sighted speakers for auditory motion events (Chapter 3), I found that when visual input is absent, as experienced by people born blind, speakers' verbal descriptions of event components differ in a way consistent with the findings of Chapter 2, as well as with findings from earlier spatial cognition studies. Blind speakers mentioned path of motion more and manner of motion less than sighted speakers. The results of Chapter 2 suggest auditory input foregrounds path over manner in the descriptions of sighted speakers too, and that the representation of path becomes more segmented due to the affordances of the auditory modality. Chapter 3 provides converging evidence for this proposal by showing that in the absence of visual experience, speakers' path representations become more segmented and path also becomes salient than manner information. This suggests perceptual experience does influence the linguistic encoding of motion events.

Furthermore, blind speakers tended to use more landmarks with segmented path descriptions (Chapter 3), aligning also with earlier route description studies suggesting that landmarks and segmented descriptions make routes more navigable for blind individuals (Iverson, 1999; Iverson & Goldin-Meadow, 1997). Critically, blind speakers were more likely to refer to landmarks in relation to their own position in space (e.g., self-anchored; such as "to/from my left") in contrast to sighted speakers who referred mainly to object-anchored landmarks (such as "to/from the elevator"). These results provide novel linguistic evidence that blind speakers rely on an egocentric frame of reference when they encode and describe spatial events, corroborating previous studies of spatial cognition that examined blind people's route knowledge and navigational skills (e.g., Cattaneo & Vecchi, 2011; Iachini et al., 2014; Noordzij et al., 2006; Pasqualotto & Proulx, 2012; Vercillo et al., 2018).

Together, the results of this thesis provide evidence that spatial representations may be shaped by visual experience. The differential spatial affordances of each input modality shapes how people are able to encode components of spatial events for linguistic purposes (Chapter 2), and a person's lifetime perceptual experience has an influence on spatial event construal as expressed in speech and gesture (Chapter 3), possibly due to differences underlying spatial cognition (e.g., Battal et al., 2020; Cattaneo & Vecchi, 2011; Lessard et al., 1998; Röder et al., 1999; Ruggiero et al., 2021; Voss et al., 2004).

Perceptual experience shapes the conceptual representations of objects and events

The findings in this thesis offer evidence supporting the proposal that a lifetime's perceptual experience influences the representation of spatial events (Chapter 3) as well as the representation of object concepts (Chapter 4). Earlier studies investigated blind people's conceptual knowledge mainly for individual object concepts and used many language-based measures, such as semantic judgments and feature listing (e.g., Bedny et al., 2019; Connolly et al., 2007; Kim et al., 2019a, 2021; Lenci et al., 2013; Marmor, 1978; Saysani et al., 2018, 2021; Shepard & Cooper, 1992). Most of these studies did not report a difference between blind and sighted people's concepts, even for concepts of colors where visual experience is typically thought crucial (Kim et al., 2021; Landau & Gleitman, 1985; Marmor, 1978; Saysani et al., 2018, 2021). Thus, these studies concluded that perceptual experience has no demonstrable effect on conceptual knowledge.

Language provides a valuable resource for blind individuals, forming a foundation for their knowledge of visual properties (e.g., Lewis et al., 2019; van Paridon et al., 2021), but how exactly blind people learn from language is still controversial (e.g., Kim et al., 2019b, 2019c; Ostarek et al., 2019). However, knowledge derived only from language may not hold the same significance or serve the same purpose for blind individuals as it does for sighted individuals. For example, despite having the knowledge about colors (e.g., *apples are red*) blind people, unlike sighted people, do not use color information to make similarity judgments about broader categories such as fruits and vegetables (Connolly et al., 2007). This suggests there may be qualitative differences in the organization of conceptual knowledge of blind and sighted people that previous studies have missed.

As discussed in detail in the preceding section, the absence of visual experience shapes how blind people encode different components of motion events in speech and gesture in line with the spatial affordances of perceptual inputs and differences underlying spatial cognition of blind people (Chapter 3). Regarding object concepts, despite largely sharing a substantial amount of knowledge about perceptual and non-perceptual features of manipulable objects, non-manipulable objects, and animals (Chapter 4), blind people's knowledge of animal concepts for some perceptual features, like colors, seems to differ from sighted people. This is consistent with some earlier findings (Kim et al., 2019a). So, the results of this thesis are consistent with the proposal that language is a good source of information (e.g., Bedny et al., 2019; Landau & Gleitman, 1985; Lewis et al., 2019; van Paridon et al., 2021). Looking into the gestures in more detail, however, revealed that blind people's silent gestures relied less on strategies depicting visuospatial aspects of concepts (such as tracing the outline of a non-manipulable object, like a bridge), and instead relied more on strategies based on motor and haptic experience (Chapter 4). So, blind people do not differ from sighted people in their gestures for concepts that they have first-person experience (motor and haptic), but for concepts that rely more on visual experience, blind people adapt their gestures based on their own sensorimotor experience. This demonstrates that visual experience shapes how object concepts are conveyed through gestures, aligning with embodied theories of language production that emphasize the link between sensorimotor experience and gestural expression (e.g., Barsalou, 2016; Hostetter & Alibali, 2008). Using a gesture task and a feature listing task together showed that differences in conceptual representation evident in gestures might not be captured by language-based approaches (Chapter 4). This might be because the affordances of gesture rely on direct mappings of visuospatial and motoric aspects of concepts through iconicity. Thus, gesture can provide additional insight into conceptual representation.

Together, the results of this thesis suggest that visual experience affects the way concepts are represented in speech and co-speech and silent gesture. While visual experience may not always be essential for acquiring the core features of concepts, the organization of conceptual knowledge of both object concepts and broader spatial event representations is shaped by perceptual experience.

Revisiting theoretical accounts of gesture production

The current thesis also has implications for theoretical accounts of gesture production. First, gesture production theories have a shared assumption that gestures arise from visuospatial imagery (de Ruiter, 2000, 2007; Hostetter & Alibali, 2008, 2019; Kita, 2000; Kita & Özyürek, 2003; Krauss et al., 2000; McNeill, 1992; McNeill & Duncan, 2000). While these theories do not explicitly rule out the involvement of non-visual (e.g., auditory) information in gesture production, the spatial affordances of non-visual information have not been systematically investigated through gesture studies. The findings of Chapter 2 indicate that sighted speakers' gestures based on auditory input alone are comparable to those based on visual input, both in terms of gesture frequency and gesture type, empirically supporting the argument that any form of information can stimulate gestures if it evokes imagery for action simulation (Hostetter & Alibali, 2019).

Critically, however, the findings of Chapter 3 and 4 showed that the lifetime absence of visual input in a person's perceptual experience affects gesture production. Blind people produced fewer iconic gestures than sighted people when they described spatial events (Chapter 3). Blind people also produced fewer silent gestures than sighted people to depict non-manipulable object and animal concepts that rely more on visuospatial imagery, but not for manipulable object concepts that rely more on motor imagery (Chapter 4). Synthesizing the findings across chapters suggests that the perceptual modality of input does not significantly influence gesture production as long as people have the prior perceptual and motor experience to simulate visuospatial and motor imagery.

Second, gesture production theories diverge in their views about the relationship between speech and gesture (de Ruiter, 2000, 2007; Hostetter & Alibali, 2008, 2019; Kita, 2000; Kita & Özyürek, 2003; Krauss et al., 2000; McNeill, 1992; McNeill & Duncan, 2000). Chapter 2 revealed that although the perceptual modality of input influences the encoding of event components in speech, it does not do so to the same extent for gesture. All speakers, regardless of the input modality, produced predominantly path-only gestures, which is in line with the Turkish language typology. The findings of Chapter 3 also showed that the gesture patterns of both blind and sighted speakers followed the Turkish typology although they differed in terms of frequency. This distinction provides new insights into the relation between speech and gesture, suggesting that gesture production is not just dependent on speech, which contradicts certain gesture theories (e.g., Sketch Model, de Ruiter, 2000; Lexical Retrieval Hypothesis, Krauss et al., 2000; Growth Point Theory, McNeill, 1992). Instead, it aligns with the idea that gestures are generated through interactions between the conceptualization underlying speech production and the visuospatial imagery underlying gesture production (The Interface Model, Kita & Özyürek, 2003). Thus, speech and gesture are independent yet interactive systems (e.g., Gesture as Simulated Action Framework, Hostetter & Alibali, 2008; Gesture-for-Conceptualization Hypothesis, Kita et al., 2017).

Understanding the nuances of gesture production in blind individuals

The current thesis also contributes to the field by filling gaps in the literature on the gesture production of blind individuals. Critically, the results of Chapter 3 showed that blind speakers' descriptions of event components differed from non-blind speakers not only in speech but also in gesture. Blind speakers produced fewer spontaneous iconic gestures, both for path and manner, than non-blind speakers. This differs from the findings of Chapter 2, where sighted people's gesture frequency did not differ according to the perceptual modality of input. This suggests that visual experience is critical for the production of gestures, particularly iconic gestures. Iconic gestures—which mimic motion, trace shapes, and depict size and spatial relations of objects—may not be as accessible to blind people.

Chapter 4 provides converging evidence for this proposal. When explicitly asked to gesture to convey individual object concepts, blind people produced fewer gestures than sighted people for non-manipulable objects and animals that are more likely to be experienced via vision due to their limited motor affordances. Moreover, compared to sighted people, blind people's gestures relied less on strategies depicting visuospatial features—i.e., tracing an object shape (the drawing strategy) or embodying a non-human entity (the personification strategy). Critically, however, blind people produced a similar amount of gesture to sighted people for manipulable objects (e.g., spoon, banana) that lean more on motor experience.

Synthesizing the findings from Chapter 3 and 4 suggests that certain features of iconic gestures (e.g., tracing the shape of an object) may be less adaptable when they build solely on non-visual input, while other aspects that rely more on motor imagery (e.g., imitating how to use an object) remain similar in blind and sighted people. This may explain the generally lower frequency of spontaneous iconic gestures in the blind population reported in earlier studies (e.g., Iverson & Goldin-Meadow, 1997; Özçalışkan et al., 2016b, 2018).

It is important to emphasize, however, that it is not the case that blind people always gesture less than sighted people. Blind people produced a comparable amount of gesture to sighted people particularly when expressing concepts related to motor and haptic experience (Chapter 4). This suggests that when the same underlying imagery is tapped into for both blind and sighted individuals (i.e., motor imagery), then gesture production can also be comparable. Furthermore, blind people more frequently used pointing gestures (such as whole-hand or index-like finger pointing) than sighted people when they mentioned landmarks of events in speech (Chapter 3). This difference may arise due to blind people's higher sensitivity to auditory cues during localization (e.g., Battal et al., 2020; Lessard et al., 1998; Röder et al., 1999; Voss et al., 2004) and navigation tasks, and also their sensitivity to echolocation (e.g., Dufour et al., 2005; Kolarik et al., 2014; Schenkman & Nilsson, 2010). So, the more frequent use of pointing gestures in the blind group might be related to their ability to build the relevant spatial representations: for example, blind speakers used pointing gestures to localize imaginary landmarks in space. It can be speculated that blind people benefit more from these gestures' self-oriented cognitive functions (e.g., helping speakers manipulate their spatial-motoric representations for speaking and thinking) than their communicative function (Kita et al., 2017; Kita & Emmorey, 2023). The findings also suggest that deictic gestures can help speakers encode and package spatial information into a suitable verbal unit—see also Kita and Emmorey (2023) for a discussion of the use of deictic co-sign gestures serving a similar purpose to deictic co-speech gestures in the current thesis.

Taken together, the results of this thesis underscore the importance of distinguishing between types of gestures such as iconic or deictic (Chapter 3), as well as recognizing the different aspects of similar types of gestures (Chapter 4). This distinction is crucial for drawing accurate conclusions regarding gesture use in the blind population. I did not find a difference when comparing the overall spontaneous gesture rate of blind and sighted speakers (Chapter 3). Yet, the gestural behavior of blind speakers was different than sighted speakers when I looked into the use of iconic and deictic gestures separately. This is not surprising considering that iconic and deictic gestures—although both representational gestures—serve different functions during communication (McNeill, 1992), and they may rely on sensorimotor experience to different extents. Still, researchers tend to lump representational gestures together when they examine gesture use. This work suggests that gesture type needs to be considered separately in future investigations of multimodal language production, particularly in relation to their use by blind people.

Methodological contributions

The current thesis also offers methodological contributions in the investigation of multimodal language production of both blind and sighted individuals. Chapter 2 investigated the influence of the perceptual modality of input (i.e., visual, auditory, or audiovisual) on speech and gesture used to express motion events. Although this is not the first study creating sounds of biological motion (e.g., Cottrell & Campbell, 2014; Geangu et al., 2015; Lewis et al., 2011; Quadrelli et al., 2019), this thesis used this approach in an original manner to explore multimodal language production. I combined sounds of human locomotion (i.e., walking, running, and limping) along with environmental sounds (e.g., opening elevator door) to generate auditory motion events (e.g., someone running into an elevator). This method enabled me to present non-visual spatial events through an ecologically relevant but experimentally controlled technique.

Furthermore, in Chapter 3, using these auditory motion events I tried to mimic 3D real-life auditory experience as much as possible. For this, I created surround sound using 5 sound speakers and manipulated the sound direction to increase the diversity and richness in events—e.g., participants could follow whether a figure was moving closer to or further away from themselves. Using this methodology, I could hold input modality constant for blind and sighted participants, which was not the case in the previous studies (Iverson, 1999; Özçalışkan et al., 2016b, 2018) where blind people were presented with tactile information while sighted people relied on vision. This method, therefore, eliminates a critical limitation of some earlier work.

Methodological limitations and future directions

How generalizable are the findings from this thesis, derived from a task in a controlled laboratory setting, to naturalistic language use situations? I created ecologically relevant stimuli based on human locomotion events to present to blind and sighted speakers, and then examined speech and gesture production (Chapters 2 and 3). This was not an everyday communication task as speakers were alone during their event descriptions without an addressee interacting with them. To boost the communicative intent of speakers, participants were instructed that other participants would watch their descriptions and watch/listen to the same events in order to match descriptions with events. However, this imaginary addressee might not substitute for an actual, physically present interactive partner. Having an addressee may influence speech as well as the frequency and form of gestures, for instance, depending on addressee knowledge (i.e., common ground between interlocutors) or whether individuals are in a monologue or dialogue (see Holler & Bavelas, 2017 for a review). Future work should examine relatively more naturalistic multimodal productions in a communication context.

Related to the degree of communication context, an important question remains regarding the extent to which blind speakers use gesture as means of communication. Blind speakers used fewer spontaneous iconic gestures, consistent with findings from earlier studies (Chapter 3). This suggests gesture is a less preferred modality of communication for blind speakers. But, blind speakers did use more pointing gestures (e.g., whole-hand or index-like finger pointing) than non-blind speakers when they talked about landmarks (Chapter 3). Pointing gestures are usually produced with communicative intent to direct listeners' attention to a physical or imaginary referent (e.g., Enfield et al., 2007; Raghavan et al., 2023). Yet, as speakers in the current thesis described the events to an imaginary addressee, the higher use of pointing gestures in the blind group might not be fully attributed to communicative intent. Although the current thesis provides some arguments for the functions of gesture in blind people, several questions remain: for example, under what circumstances do blind people benefit from gestures' self-oriented cognitive functions? Or, what are the factors (e.g., individual differences) leading to the variation in the frequency of gesture in blind people? Some of these questions can be addressed in future research, for instance, by manipulating the presence or visual status (i.e., blind or sighted) of an addressee in a more communication-focused paradigm.

It is also important to keep in mind that seeing the gestures of others may be essential for acquiring specific gesture forms to convey particular concepts effectively. Influenced by others' gestures, people may shape their own gestures to ensure they are decodable by others, which eventually leads to systematicity in the forms of gestures sighted people produce for certain concepts (e.g., van Nispen et al., 2017). Blind people relied on different gesture forms than sighted people to depict concepts (Chapter 4). There were also cases where they used the same gestural strategy as sighted participants, but where the content of gestures differed—e.g., when depicting the concept *bread* with an acting strategy, a sighted gesturer mimicked the action of slicing bread, while a blind gesturer mimicked the action of smelling bread. The current thesis did not measure the systematicity and comprehensibility of gestures, but observationally it seems that blind people's gestures may be less comprehensible than sighted people's gestures. Indeed, there is some supporting evidence for this proposal from a recent study showing that people are more successful in interpreting sighted people's gestures compared to blind people's gestures (Fay et al., 2022). Having no feedback from sighted interlocutors may also be one of the reasons that blind people rely less on iconic gestures as a means of communication. This could be a promising area of investigation for future studies.

Conclusion

This thesis contributes to a larger body of knowledge regarding theories of multimodal language production—including the nature of imagery underlying gesture production and the relation between speech and gesture, blind individuals' gesture productions, and discussions surrounding the relation between perceptual experience and conceptual representations.

The aim of the thesis was to examine the influence of perceptual experience on the encoding of spatial events in multimodal language, in speech and gesture, and conceptual representations that manifest in these observable behaviors. I showed that spatial affordances of perceptual modalities (vision and audition) influence how sighted speakers encode motion events in their verbal but not gestural expressions, with the visual modality emphasizing manner over path of motion in speech. I also found that visual input leads to more detailed linguistic descriptions of motion events than auditory input, highlighting the dominance of visual modality in spatial perception. Furthermore, I showed that the absence of visual experience influences both verbal and gestural encoding of motion event components, in line with differences underlying spatial cognition in blind individuals. Blind people describe spatial information in motion events from an egocentric perspective (i.e., relying on their own position in space) and tend to have segmented path representations (i.e., breaking the path into smaller chunks) with fewer iconic gestures spontaneously accompanying their speech. Lastly, I showed that visual experience shapes how object concepts are depicted through gestures. Blind people adjust their gestures based on their sensorimotor experience, and such adaptations in gesture forms may arise from differences in conceptual knowledge, aligning with an embodied cognition framework.

Overall, the findings demonstrate that perceptual experience shapes how blind and sighted people express object and event concepts in multimodal language. These changes are in line with the affordances of perceptual inputs and posited differences underlying spatial cognition in blind and sighted individuals. Looking at multimodal language production, particularly gesture, can provide different insights into how perceptual experience shapes conceptual knowledge than what language-based measures alone can offer.

R

References

Akhavan, N., Nozari, N., & Göksun, T. (2017). Expression of motion events in Farsi. *Language, Cognition and Neuroscience*, *32*(6), 792–804. https://doi.org/10.1080/23273798.2016.1276607

Aktan-Erciyes, A., Akbuğa, E., Kızıldere, E., & Göksun, T. (2022). Motion event representation in L1-Turkish versus L2-English speech and gesture: Relations to eye movements for event components. *International Journal of Bilingualism*, 13670069221076838. https://doi.org/10.1177/13670069221076838

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition & Computation*, *5*(4), 307–331. https://doi.org/10.1207/s15427633scc0504_2

Allen, S., Özyürek, A., Kita, S., Brown, A., Furman, R., Ishizuka, T., & Fujii, M. (2007). Language-specific and universal influences in children's syntactic packaging of Manner and Path: A comparison of English, Japanese, and Turkish. *Cognition*, *102*(1), 16–48. https://doi.org/10.1016/j.cognition.2005.12.006

Azar, Z., Backus, A., & Özyürek, A. (2020). Language contact does not drive gesture transfer: Heritage speakers maintain language specific gesture patterns in each language. *Bilingualism: Language and Cognition*, *23*(2), 414–428. https://doi.org/10.1017/S136672891900018X

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barsalou, L. W. (2016). On staying grounded and avoiding quixotic dead ends. *Psychonomic Bulletin & Review*, *23*(4), 1122–1142. https://doi.org/10.3758/s13423-016-1028-3

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), Article 1. https://doi.org/10.18637/jss.v067.i01

Battal, C., Occelli, V., Bertonati, G., Falagiarda, F., & Collignon, O. (2020). General enhancement of spatial hearing in congenitally blind people. *Psychological Science*, *31*(9), 1129–1139. https://doi.org/10.1177/0956797620935584

Bedny, M., Caramazza, A., Grossman, E., Pascual-Leone, A., & Saxe, R. (2008). Concepts are more than percepts: The case of action verbs. *Journal of Neuroscience*, *28*(44), 11347–11353. https://doi.org/10.1523/JNEUROSCI.3039-08.2008

Bedny, M., Caramazza, A., Pascual-Leone, A., & Saxe, R. (2012). Typical neural representations of action verbs develop without vision. *Cerebral Cortex*, *22*(2), 286–293. https://doi.org/10.1093/cercor/bhr081

Bedny, M., Koster-Hale, J., Elli, G., Yazzolino, L., & Saxe, R. (2019). There's more to "sparkle" than meets the eye: Knowledge of vision and light verbs among congenitally blind and sighted individuals. *Cognition*, *189*, 105–115. https://doi.org/10.1016/j.cognition.2019.03.017

Bedny, M., & Saxe, R. (2012). Insights into the origins of knowledge from the cognitive neuroscience of blindness. *Cognitive Neuropsychology*, *29*(1–2), 56–84. https://doi.org/10.1080/02643294.2012.713342

Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture Changes Thought by Grounding It in Action. *Psychological Science*, *21*(11), 1605–1610. https://doi.org/10.1177/0956797610385353

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, *26*(12), 1153–1170. https://doi.org/10.1016/j.tics.2022.09.015

Brentari, D., Renzo, A. D., Keane, J., & Volterra, V. (2015). Cognitive, Cultural, and Linguistic Sources of a Handshape Distinction Expressing Agentivity. *Topics in Cognitive Science*, *7*(1), 95–123. https://doi.org/10.1111/tops.12123

Brown, A., & Chen, J. (2013). Construal of Manner in speech and gesture in Mandarin, English, and Japanese. *Cognitive Linguistics*, *24*(4), 605–631. https://doi.org/10.1515/cog-2013-0021

Casasanto, D., & Lupyan, G. (2015). All Concepts Are Ad Hoc Concepts. In E. Margolis & S. Laurence (Eds.), *The Conceptual Mind* (pp. 543–566). The MIT Press. https://doi.org/10.7551/mitpress/9383.003.0031

Cattaneo, Z., & Vecchi, T. (2011). *Blind vision: The neuroscience of visual impairment*. MIT Press. https://doi.org/10.7551/mitpress/9780262015035.001.0001

Choi, S., & Lantolf, J. P. (2008). REPRESENTATION AND EMBODIMENT OF MEANING IN L2 COMMUNICATION: Motion Events in the Speech and Gesture of Advanced L2 Korean and L2 English Speakers. *Studies in Second Language Acquisition*, *30*(2), 191–224. https://doi.org/10.1017/S0272263108080315

Chui, K. (2009). Linguistic and imagistic representations of motion events. *Journal of Pragmatics*, *41*(9), 1767–1777. https://doi.org/10.1016/j.pragma.2009.04.006

Church, Alibali, & Kelly. (2017). *Why Gesture? How the hands function in speaking, thinking and communicating*. John Benjamins Publishing Company.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*, 213–220. https://doi.org/10.1037/h0026256

Connell, L. (2019). What have labels ever done for us? The linguistic shortcut in conceptual processing. *Language, Cognition and Neuroscience*, *34*(10), 1308–1318. https://doi.org/10.1080/23273798.2018.1471512

Connell, L., & Lynott, D. (2014). Principles of Representation: Why You Can't Represent the Same Concept Twice. *Topics in Cognitive Science*, *6*(3), 390–406. https://doi.org/10.1111/tops.12097

R

Connolly, A. C., Gleitman, L. R., & Thompson-Schill, S. L. (2007). Effect of congenital blindness on the semantic representation of some everyday concepts. *Proceedings of the National Academy of Sciences*, *104*(20), 8241–8246. https://doi.org/10.1073/pnas.0702812104

Cook, S. W., & Tanenhaus, M. K. (2009). Embodied communication: Speakers' gestures affect listeners' actions. *Cognition*, *113*(1), 98–104. https://doi.org/10.1016/j.cognition.2009.06.006

Cottrell, D., & Campbell, M. E. J. (2014). Auditory Perception of a Human Walker. *Perception*, *43*(11), 1225–1238. https://doi.org/10.1068/p7669

Cravotta, A., Bus, à M. G., & Prieto, P. (2019). Effects of encouraging the use of gestures on speech. *Journal of Speech, Language, and Hearing Research*, *62*(9), 3204–3219. https://doi.org/10.1044/2019_JSLHR-S-18-0493

Crollen, V., & Collignon, O. (2020). How visual is the « number sense »? Insights from the blind. *Neuroscience & Biobehavioral Reviews*, *118*, 290–297. https://doi.org/10.1016/j.neubiorev.2020.07.022

Davis, C. P., & Yee, E. (2021). Building semantic memory from embodied and distributional language experience. *WIREs Cognitive Science*, *12*(5), e1555. https://doi.org/10.1002/wcs.1555

de Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and Gesture* (pp. 284–311). Cambridge University Press.

de Ruiter, J. P. (2007). Postcards from the mind: The relationship between speech, imagistic gesture, and thought. *Gesture*, *7*(1), 21–38. https://doi.org/10.1075/gest.7.1.03rui

Dove, G. (2022). *Abstract Concepts and the Embodied Mind: Rethinking Grounded Cognition*. Oxford University Press.

Dove, G. (2023). Rethinking the role of language in embodied cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *378*(1870), 20210375. https://doi.org/10.1098/rstb.2021.0375

Dufour, A., Després, O., & Candas, V. (2005). Enhanced sensitivity to echo cues in blind subjects. *Experimental Brain Research*, *165*(4), 515–519. https://doi.org/10.1007/s00221-005-2329-3

Eimer, M. (2004). Multisensory integration: How visual experience shapes spatial perception. *Current Biology*, *14*(3), R115–R117. https://doi.org/10.1016/j.cub.2004.01.018

Elli, G. V., Bedny, M., & Landau, B. (2021). How does a blind person see? Developmental change in applying visual verbs to agents with disabilities. *Cognition*, *212*, 104683. https://doi.org/10.1016/j.cognition.2021.104683

Emmorey, K. (2014). Iconicity as structure mapping. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130301. https://doi.org/10.1098/rstb.2013.0301

Enfield, N. J., Kita, S., & de Ruiter, J. P. (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics*, *39*(10), 1722–1741. https://doi.org/10.1016/j.pragma.2007.03.001

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

Fay, N., Walker, B., Ellison, T. M., Blundell, Z., De Kleine, N., Garde, M., Lister, C. J., & Goldin-Meadow, S. (2022). Gesture is the primary modality for language creation. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1970), 20220066. https://doi.org/10.1098/rspb.2022.0066

Floyd, S., Roque, L. S., & Majid, A. (2018). Smell is coded in grammar and frequent in discourse: Cha'palaa olfactory language in cross-linguistic perspective. *Journal of Linguistic Anthropology*, *28*(2), 175–196. https://doi.org/10.1111/jola.12190

Geangu, E., Quadrelli, E., Lewis, J. W., Macchi Cassia, V., & Turati, C. (2015). By the sound of it. An ERP investigation of human action sound processing in 7-month-old infants. *Developmental Cognitive Neuroscience*, *12*, 134–144. https://doi.org/10.1016/j.dcn.2015.01.005

Geangu, E., Roberti, E., & Turati, C. (2021). Do infants represent human actions cross-modally? An ERP visual-auditory priming study. *Biological Psychology*, *160*, 108047. https://doi.org/10.1016/j.biopsycho.2021.108047

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, *83*(1), 49–79. https://doi.org/10.1016/S0010-0277(01)00166-4

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's Influence on Thought: The Case of Gesture. *Perspectives on Psychological Science*, *5*(6), 664–674. https://doi.org/10.1177/1745691610388764

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, *105*(27), 9163–9168. https://doi.org/10.1073/pnas.0710060105

Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R. J., & Belin, P. (2004). Neuropsychology: Pitch discrimination in the early blind. *Nature*, *430*(6997), 309. https://doi.org/10.1038/430309a

Göz, İ. (2003). *Yazılı Türkçenin Kelime Sıklığı Sözlüğü*. TDK yayınları.

Gullberg, M., Hendriks, H., & Hickmann, M. (2008). Learning to talk and gesture about motion in French. *First Language*, *28*(2), 200–236. https://doi.org/10.1177/0142723707088074

He, C., Peelen, M. V., Han, Z., Lin, N., Caramazza, A., & Bi, Y. (2013). Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. *NeuroImage*, *79*, 1–9. https://doi.org/10.1016/j.neuroimage.2013.04.051

R

Holler, J., & Bavelas, J. (2017). Multi-modal communication of common ground: A review of social functions. In *Why gesture? How the hands function in speaking, thinking and communicating* (pp. 213–240). Benjamins. https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2352807

Holler, J., Drijvers, L., Rafiee, A., & Majid, A. (2022). Embodied space-pitch associations are shaped by language. *Cognitive Science*, *46*(2), e13083. https://doi.org/10.1111/cogs.13083

Holler, J., & Levinson, S. C. (2019). Multimodal Language Processing in Human Communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. https://doi.org/10.1016/j.tics.2019.05.006

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495–514. https://doi.org/10.3758/PBR.15.3.495

Hostetter, A. B., & Alibali, M. W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review*, *26*(3), 721–752. https://doi.org/10.3758/s13423-018-1548-0

Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation* (p. 533). John Wiley & Sons.

Iachini, T., Ruggiero, G., & Ruotolo, F. (2014). Does blindness affect egocentric and allocentric frames of reference in small and large scale spaces? *Behavioural Brain Research*, *273*, 73–81. https://doi.org/10.1016/j.bbr.2014.07.032

Iverson, J. M. (1999). How to get to the cafeteria: Gesture and speech in blind and sighted children's spatial descriptions. *Developmental Psychology*, *35*(4), 1132–1142. https://doi.org/10.1037/0012-1649.35.4.1132

Iverson, J. M., & Goldin-Meadow, S. (1997). What's communication got to do with it? Gesture in children blind from birth. *Developmental Psychology*, *33*(3), 453–467. https://doi.org/10.1037/0012-1649.33.3.453

Kemmerer, D. (2023). Grounded Cognition Entails Linguistic Relativity: A Neglected Implication of a Major Semantic Theory. *Topics in Cognitive Science*, *n/a*(n/a). https://doi.org/10.1111/tops.12628

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge University Press.

Kim, J. S., Aheimer, B., Montané Manrara, V., & Bedny, M. (2021). Shared understanding of color among sighted and blind adults. *Proceedings of the National Academy of Sciences*, *118*(33), e2020192118. https://doi.org/10.1073/pnas.2020192118

Kim, J. S., Elli, G. V., & Bedny, M. (2019a). Knowledge of animal appearance among sighted and blind adults. *Proceedings of the National Academy of Sciences*, *116*(23), 11213–11222. https://doi.org/10.1073/pnas.1900952116

Kim, J. S., Elli, G. V., & Bedny, M. (2019b). Reply to Lewis et al.: Inference is key to learning appearance from language, for humans and distributional semantic models alike. *Proceedings of the National Academy of Sciences*, *116*(39), 19239–19240. https://doi.org/10.1073/pnas.1910410116

Kim, J. S., Elli, G. V., & Bedny, M. (2019c). Reply to Ostarek et al.: Language, but not co-occurrence statistics, is useful for learning animal appearance. *Proceedings of the National Academy of Sciences*, *116*(44), 21974–21975. https://doi.org/10.1073/pnas.1912854116

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and Gesture* (pp. 162–185). Cambridge University Press.

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, *124*(3), 245–266. https://doi.org/10.1037/rev0000059

Kita, S., & Emmorey, K. (2023). Gesture links language and cognition for spoken and signed languages. *Nature Reviews Psychology*, *2*(7), Article 7. https://doi.org/10.1038/s44159-023-00186-9

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16–32. https://doi.org/10.1016/S0749-596X(02)00505-3

Kolarik, A. J., Cirstea, S., Pardhan, S., & Moore, B. C. J. (2014). A summary of research investigating echolocation abilities of blind and sighted humans. *Hearing Research*, *310*, 60–68. https://doi.org/10.1016/j.heares.2014.01.010

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and Gesture* (pp. 261–283). Cambridge University Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(1), Article 1. https://doi.org/10.18637/jss.v082.i13

Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child*. Harvard University Press.

Lenci, A., Baroni, M., Cazzolli, G., & Marotta, G. (2013). BLIND: A set of semantic feature norms from the congenitally blind. *Behavior Research Methods*, *45*(4), 1218–1233. https://doi.org/10.3758/s13428-013-0323-4

Lenth, R. (2022). *Emmeans: Estimated marginal means, aka least-squares means. R package version 1.7.3*. https://CRAN.R-project.org/package=emmeans

Lessard, N., Paré, M., Lepore, F., & Lassonde, M. (1998). Early-blind human subjects localize sound sources better than sighted subjects. *Nature*, *395*(6699), 278–280. https://doi.org/10.1038/26228

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.

R

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130302. https://doi.org/10.1098/rstb.2013.0302

Levinson, S. C., & Majid, A. (2014). Differential ineffability and the senses. *Mind & Language*, *29*(4), 407–427. https://doi.org/10.1111/mila.12057

Lewis, J. W., Talkington, W. J., Puce, A., Engel, L. R., & Frum, C. (2011). Cortical Networks Representing Object Categories and High-level Attributes of Familiar Real-world Action Sounds. *Journal of Cognitive Neuroscience*, *23*(8), 2079–2101. https://doi.org/10.1162/jocn.2010.21570

Lewis, M., Zettersten, M., & Lupyan, G. (2019). Distributional semantics as a source of visual knowledge. *Proceedings of the National Academy of Sciences*, *116*(39), 19237–19238. https://doi.org/10.1073/pnas.1910148116

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster sensorimotor norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, *52*(3), 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

Macuch Silva, V., Holler, J., Ozyurek, A., & Roberts, S. G. (2020). Multimodality and the origin of a novel communication system in face-to-face interaction. *Royal Society Open Science*, *7*(1), 182056. https://doi.org/10.1098/rsos.182056

Mahon, B. Z. (2015). What is embodied about cognition? *Language, Cognition and Neuroscience*, *30*(4), 420–429. https://doi.org/10.1080/23273798.2014.987791

Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. *Neuron*, *63*(3), 397–405. https://doi.org/10.1016/j.neuron.2009.07.012

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris*, *102*(1), 59–70. https://doi.org/10.1016/j.jphysparis.2008.03.004

Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'Grady, L., Woll, B., LeLan, B., Sousa, H. de, Cansler, B. L., Shayan, S., Vos, C. de, Senft, G., Enfield, N. J., Razak, R. A., Fedden, S., Tufvesson, S., Dingemanse, M., Ozturk, O., … Levinson, S. C. (2018). Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences*, *115*(45), 11369–11376. https://doi.org/10.1073/pnas.1720419115

Malt, B. C., Ameel, E., Imai, M., Gennari, S. P., Saji, N., & Majid, A. (2014). Human locomotion in languages: Constraints on moving and meaning. *Journal of Memory and Language*, *74*, 107–123. https://doi.org/10.1016/j.jml.2013.08.003

Mamus, E., Rissman, L., Majid, A., & Özyürek, A. (2019). Effects of blindfolding on verbal and gestural expression of path in auditory motion events. In A. K. Goel, C. M. Seifert, & C. C. Freksa (Eds.), *Proceedings of the 41st Annual Meeting of the Cognitive Science Society (CogSci 2019)* (pp. 2275–2281). Cognitive Science Society.

Mamus, E., Speed, L. J., Özyürek, A., & Majid, A. (2023). The effect of input sensory modality on the multimodal encoding of motion events. *Language, Cognition and Neuroscience*, *38*(5), 711–723. https://doi.org/10.1080/23273798.2022.2141282

Mamus, E., Speed, L. J., Rissman, L., Majid, A., & Özyürek, A. (2023). Lack of Visual Experience Affects Multimodal Language Production: Evidence From Congenitally Blind and Sighted People. *Cognitive Science*, *47*(1), e13228. https://doi.org/10.1111/cogs.13228

Marmor, G. S. (1978). Age at onset of blindness and the development of the semantics of color names. *Journal of Experimental Child Psychology*, *25*(2), 267–278. https://doi.org/10.1016/0022-0965(78)90082-6

Marques, J. F. (2010). The effect of visual deprivation on the organization of conceptual knowledge: Testing the grounded cognition hypothesis. *Experimental Psychology*, *57*, 83–88. https://doi.org/10.1027/1618-3169/a000011

Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2016). Can you handle this? The impact of object affordances on how co-speech gestures are produced. *Language, Cognition and Neuroscience*, *31*(3), 430–440. https://doi.org/10.1080/23273798.2015.1108448

Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2017). How What We See and What We Know Influence Iconic Gesture Production. *Journal of Nonverbal Behavior*, *41*(4), 367–394. https://doi.org/10.1007/s10919-017-0261-4

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

McNeill, D. (2000). *Language and Gesture*. Cambridge University Press.

McNeill, D., & Duncan, S. D. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and Gesture* (pp. 141–161). Cambridge University Press.

McRae, K., Cree, G. S., Seidenberg, M. S., & Mcnorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559. https://doi.org/10.3758/BF03192726

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804. https://doi.org/10.1016/j.cortex.2010.11.002

Minervino, R. A., Martín, A., Tavernini, L. M., & Trench, M. (2018). The Understanding of Visual Metaphors by the Congenitally Blind. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01242

Nielsen, A. K., & Dingemanse, M. (2021). Iconicity in Word Learning and Beyond: A Critical Review. *Language and Speech*, *64*(1), 52–72. https://doi.org/10.1177/0023830920914339

Noordzij, M. L., Zuidhoek, S., & Postma, A. (2006). The influence of visual experience on the ability to form spatial mental models based on route and survey descriptions. *Cognition*, *100*(2), 321–342. https://doi.org/10.1016/j.cognition.2005.05.006

R

Ortega, G., & Özyürek, A. (2020a). Systematic mappings between semantic categories and types of iconic representations in the manual modality: A normed database of silent gesture. *Behavior Research Methods*, *52*(1), 51–67. https://doi.org/10.3758/s13428-019-01204-6

Ortega, G., & Özyürek, A. (2020b). Types of iconicity and combinatorial strategies distinguish semantic categories in silent gesture across cultures. *Language and Cognition*, *12*(1), 84–113. https://doi.org/10.1017/langcog.2019.28

Ostarek, M., Paridon, J. van, & Montero-Melis, G. (2019). Sighted people's language is not helpful for blind individuals' acquisition of typical animal colors. *Proceedings of the National Academy of Sciences*, *116*(44), 21972–21973. https://doi.org/10.1073/pnas.1912302116

Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016a). Does language shape silent gesture? *Cognition*, *148*, 10–18. https://doi.org/10.1016/j.cognition.2015.12.001

Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016b). Is seeing gesture necessary to gesture like a native speaker? *Psychological Science*, *27*(5), 737–747. https://doi.org/10.1177/0956797616629931

Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2018). Blind speakers show language-specific patterns in co-speech gesture but not silent gesture. *Cognitive Science*, *42*(3), 1001–1014. https://doi.org/10.1111/cogs.12502

Özyürek, A., Kita, S., Allen, S., Furman, R., & Brown, A. (2005). How does linguistic framing of events influence co-speech gestures?: Insights from crosslinguistic variations and similarities. *Gesture*, *5*(1–2), 219–240. https://doi.org/10.1075/gest.5.1.15ozy

Padden, C., Hwang, S.-O., Lepic, R., & Seegers, S. (2015). Tools for Language: Patterned Iconicity in Sign Language Nouns and Verbs. *Topics in Cognitive Science*, *7*(1), 81–94. https://doi.org/10.1111/tops.12121

Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition*, *84*(2), 189–219. https://doi.org/10.1016/S0010-0277(02)00046-X

Papies, E. K., Tatar, B., Keesman, M., Best, M., Lindner, K., Barsalou, L., Rusz, D., & Dutriaux, L. (2020). *Measuring and interpreting cognitive representations of foods and drinks: A procedure for collecting and coding feature listing data*. OSF Preprints. https://doi.org/10.31219/osf.io/ufpx8

Pasqualotto, A., & Proulx, M. J. (2012). The role of visual experience for the neural basis of spatial cognition. *Neuroscience & Biobehavioral Reviews*, *36*(4), 1179–1187. https://doi.org/10.1016/j.neubiorev.2012.01.008

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), Article 12. https://doi.org/10.1038/nrn2277

Peelen, M. V., Bracci, S., Lu, X., He, C., Caramazza, A., & Bi, Y. (2013). Tool Selectivity in Left Occipitotemporal Cortex Develops without Vision. *Journal of Cognitive Neuroscience*, *25*(8), 1225–1234. https://doi.org/10.1162/jocn_a_00411

Perlman, M., & Cain, A. A. (2014). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture*, *14*(3), 320–350. https://doi.org/10.1075/gest.14.3.03per

Perlman, M., & Lupyan, G. (2018). People Can Create Iconic Vocalizations to Communicate Various Meanings to Naïve Listeners. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-20961-6

Perniss, P. (2018). Why We Should Study Multimodal Language. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01109

Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a General Property of Language: Evidence from Spoken and Signed Languages. *Frontiers in Psychology*, *1*. https://www.frontiersin.org/articles/10.3389/fpsyg.2010.00227

Pouw, W. T. J. L., de Nooijer, J. A., van Gog, T., Zwaan, R. A., & Paas, F. (2014). Toward a more embedded/extended perspective on the cognitive function of gestures. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/article/10.3389/fpsyg.2014.00359

Pouw, W., Wassenburg, S. I., Hostetter, A. B., de Koning, B. B., & Paas, F. (2020). Does gesture strengthen sensorimotor knowledge of objects? The case of the size-weight illusion. *Psychological Research*, *84*(4), 966–980. https://doi.org/10.1007/s00426-018-1128-y

Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, *17*(9), 458–470. https://doi.org/10.1016/j.tics.2013.06.004

Quadrelli, E., Geangu, E., & Turati, C. (2019). Human action sounds elicit sensorimotor activation early in life. *Cortex*, *117*, 323–335. https://doi.org/10.1016/j.cortex.2019.05.009

R Core Team. (2022). *R: A language and environment for statistical computing*.

Raghavan, R., Raviv, L., & Peeters, D. (2023). What's your point? Insights from virtual reality on the relation between intention and action in the production of pointing gestures. *Cognition*, *240*, 105581. https://doi.org/10.1016/j.cognition.2023.105581

Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *Journal of Neurophysiology*, *89*(2), 1078–1093. https://doi.org/10.1152/jn.00706.2002

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), 1085–1099. https://doi.org/10.1037/0096-1523.28.5.1085

Röder, B., Teder-Sälejärvi, W., Sterr, A., Rösler, F., Hillyard, S. A., & Neville, H. J. (1999). Improved auditory spatial tuning in blind humans. *Nature*, *400*(6740), 162–166. https://doi.org/10.1038/22106

Ruggiero, G., Ruotolo, F., & Iachini, T. (2021). How ageing and blindness affect egocentric and allocentric spatial memory. *Quarterly Journal of Experimental Psychology*, 17470218211056772. https://doi.org/10.1177/17470218211056772

San Roque, L., Kendrick, K. H., Norcliffe, E., Brown, P., Defina, R., Dingemanse, M., Dirksmeyer, T., Enfield, N. J., Floyd, S., Hammond, J., Rossi, G., Tufvesson, S., Van Putten, S., & Majid, A. (2015). Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. *Cognitive Linguistics*, *26*, 31–60. https://doi.org/10.1515/cog-2014-0089

Saysani, A., Corballis, M. C., & Corballis, P. M. (2018). Colour envisioned: Concepts of colour in the blind and sighted. *Visual Cognition*, *26*(5), 382–392. https://doi.org/10.1080/13506285.2018.1465148

Saysani, A., Corballis, M. C., & Corballis, P. M. (2021). Seeing colour through language: Colour knowledge in the blind and sighted. *Visual Cognition*, *29*(1), 63–71. https://doi.org/10.1080/13506285.2020.1866726

Schenkman, B. N., & Nilsson, M. E. (2010). Human Echolocation: Blind and Sighted Persons' Ability to Detect Sounds Recorded in the Presence of a Reflecting Object. *Perception*, *39*(4), 483–501. https://doi.org/10.1068/p6473

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, *408*(6814), Article 6814. https://doi.org/10.1038/35048669

Shepard, R. N., & Cooper, L. A. (1992). Representation of colors in the blind, color-blind, and normally sighted. *Psychological Science*, *3*(2), 97–104. https://doi.org/10.1111/j.1467-9280.1992.tb00006.x

Slobin, D. (1996). From "thought" and "language" to "thinking for speaking." In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge University Press.

Slobin, D. I., Ibarretxe-Antuñano, I., Kopecka, A., & Majid, A. (2014). Manners of human gait: A crosslinguistic event-naming study. *Cognitive Linguistics*, *25*(4), 701–741. https://doi.org/10.1515/cog-2014-0061

Slonimska, A., Özyürek, A., & Capirci, O. (2020). The role of iconicity and simultaneity for efficient communication: The case of Italian Sign Language (LIS). *Cognition*, *200*, 104246. https://doi.org/10.1016/j.cognition.2020.104246

Slonimska, A., Özyürek, A., & Capirci, O. (2022). Simultaneity as an Emergent Property of Efficient Communication in Language: A Comparison of Silent Gesture and Sign Language. *Cognitive Science*, *46*(5), e13133. https://doi.org/10.1111/cogs.13133

Speed, L. J., & Majid, A. (2017). Dutch modality exclusivity norms: Simulating perceptual modality in space. *Behavior Research Methods*, *49*(6), 2204–2218. https://doi.org/10.3758/s13428-017-0852-3

Spence, C., & Squire, S. (2003). Multisensory integration: Maintaining the perception of synchrony. *Current Biology*, *13*(13), R519–R521. https://doi.org/10.1016/S0960-9822(03)00445-7

Striem-Amit, E., Wang, X., Bi, Y., & Caramazza, A. (2018). Neural representation of visual concepts in people born blind. *Nature Communications*, *9*(1), 5250. https://doi.org/10.1038/s41467-018-07574-3

Sümer, B., & Özyürek, A. (2022). Cross-modal investigation of event component omissions in language development: A comparison of signing and speaking children. *Language, Cognition and Neuroscience*, 1–17. https://doi.org/10.1080/23273798.2022.2042336

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and semantic description* (pp. 36–149). Cambridge University Press.

Talmy, L. (2000). *Toward a cognitive semantics* (Vol. 1). MIT Press.

Taub, S. F. (2001). *Language from the Body: Iconicity and Metaphor in American Sign Language*. Cambridge University Press.

Ter Bekke, M., Özyürek, A., & Ünal, E. (2022). Speaking but not gesturing predicts event memory: A cross-linguistic comparison. *Language and Cognition*, *14*(3), 362–384. https://doi.org/10.1017/langcog.2022.3

Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. (2021). The neglected 95% revisited: Is American psychology becoming less American? *The American Psychologist*, *76*(1), 116–129. https://doi.org/10.1037/amp0000622

Thinus-Blanc, C., & Gaunet, F. (1997). Representation of space in blind persons: Vision as a spatial sense? *Psychological Bulletin*, *121*(1), 20–42. https://doi.org/10.1037/0033-2909.121.1.20

Ünal, E., Manhardt, F., & Özyürek, A. (2022). Speaking and gesturing guide event perception during message conceptualization: Evidence from eye movements. *Cognition*, *225*, 105127. https://doi.org/10.1016/j.cognition.2022.105127

van Nispen, K., van de Sandt-Koenderman, W. Mieke. E., & Krahmer, E. (2017). Production and Comprehension of Pantomimes Used to Depict Objects. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01095

van Paridon, J., Liu, Q., & Lupyan, G. (2021). How do blind people know that blue is cold? Distributional semantics encode color-adjective associations. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*. https://escholarship.org/uc/item/6sq7h506

Vercillo, T., Tonelli, A., & Gori, M. (2018). Early visual deprivation prompts the use of body-centered frames of reference for auditory localization. *Cognition*, *170*, 263–269. https://doi.org/10.1016/j.cognition.2017.10.013

Viberg, Å. (1983). The verbs of perception: A typological study. *Linguistics*, *21*, 123–162.

Vigliocco, G., & Vinson, D. P. (2007). Semantic representation. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198568971.013.0012

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, *40*(1), 183–190. https://doi.org/10.3758/BRM.40.1.183

Voss, P., Lassonde, M., Gougoux, F., Fortin, M., Guillemot, J.-P., & Lepore, F. (2004). Early- and late-onset blind individuals show supra-normal auditory abilities in far-space. *Current Biology*, *14*(19), 1734–1738. https://doi.org/10.1016/j.cub.2004.09.051

Wan, C. Y., Wood, A. G., Reutens, D. C., & Wilson, S. J. (2010). Early but not late-blindness leads to enhanced auditory perception. *Neuropsychologia*, *48*(1), 344–348. https://doi.org/10.1016/j.neuropsychologia.2009.08.016

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

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636.

Winter, B., Perlman, M., & Majid, A. (2018). Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition*, *179*, 213–220. https://doi.org/10.1016/j.cognition.2018.05.008

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). *ELAN: A professional framework for multimodality research*. 1556–1559.

Wöstmann, M., Schmitt, L.-M., & Obleser, J. (2020). Does closing the eyes enhance auditory attention? Eye closure increases attentional alpha-power modulation but not listening performance. *Journal of Cognitive Neuroscience*, *32*(2), 212–225. https://doi.org/10.1162/jocn_a_01403

Appendices

A. Research data management

B. Supplementary materials for Chapter 4

English Summary

Nederlandse samenvatting

Curriculum Vitae

Publications

Acknowledgements

A. Research data management

Data

For Chapters 2 and 3 of this thesis, data has been collected in 2019 at Bogazici University in Istanbul, Turkey. For Chapter 4, data has been collected online in 2021. No other datasets have been used for the research reported in this thesis. For Chapters 2 and 3, participants were recruited via the Bogazici participant pool system and Bogazici GETEM (a tech and education center for blind people). For Chapter 4, participants were recruited via online advertisements as well as Bogazici GETEM.

Ethical approval and informed consent

The studies in Chapters 2-4 met the criteria of the Radboud Ethics Assessment Committee with the approval for the VICI-project (Giving Cognition a Hand: Linking Spatial Cognition to Linguistic Expression in Native and Late Learners of Sign Language and Bimodal Bilinguals, application 2036). In addition, the studies in Chapters 2 and 3 met the criteria of Ethics Assessment Committee of Bogazici University in Istanbul, Turkey, where the data was collected.

Participants received written information about the study when they signed up, and prior to the testing day they received an email with more detailed information about the study. At the start of the testing session, key information was repeated verbally. Notably, participants were informed that audio-video recordings would be made. Written informed consent for the studies of Chapters 2-3 and online informed consent for the study of Chapter 4 were obtained before data collection started. On the consent form, participants could optionally agree to the sharing of audio/video data for educational purposes and/or to promote the research, through a) presentations/lectures (not publicly available) and b) journals or other (online) news outlets (publicly available through internet).

Data storage

This thesis project is archived in the Max Planck for Psycholinguistics Repository (https://archive.mpi.nl/). This includes the video and audio data in their original form as well as processed data, ELAN files, and files related to pre-processing and coding. Here, original means without any manipulations that limit future analyses of these data.

In order for researchers to get access, they need to sign a custom-made Data Use Agreement, specifying restrictions on data storage and further sharing. Furthermore, for each empirical chapter in this thesis (chapter 2-4), the anonymized raw data (including annotations as well as the R scripts that were used for processing and analyzing the data) is publicly available via the Open science framework and can be found via unique links provided in each empirical chapter, ensuring research transparency and reproducibility. Below are the citations for the published and unpublished studies of the thesis, along with their content and OSF links.

- Chapter 2:
  Mamus, E., Speed, L., Özyürek, A., & Majid, A. (2023). The effect of input sensory modality on the multimodal encoding of motion events. Language, Cognition and Neuroscience, 38(5), 711-723. doi:10.1080/23273798.2022.2141282.
  Content: 90 sighted participants (30 per condition), speech and gesture annotations, supplementary materials, analysis code, accesible via https://osf.io/qe7dz/.
- Chapter 3:
  Mamus, E., Speed, L. J., Rissman, L., Majid, A., & Özyürek, A. (2023). Lack of visual experience affects multimodal language production: Evidence from congenitally blind and sighted people. Cognitive Science, 47(1): e13228. doi:10.1111/cogs.13228.
  Content: 63 sighted, blindfolded, and blind participants (21 per condition), speech and gesture annotations, supplementary materials, analysis code, accesible via https://osf.io/qsr7j/.
- Chapter 4:
  Mamus, E., Speed, L. J., Ortega, G., Majid, A., & Ozyurek, A. (under revision). Gestures reveal how visual experience shapes concepts in blind and sighted individuals.
  Content: 60 sighted and blind participants (30 per condition), gesture annotations, feature listing questionnaire, supplementary materials, analysis code, accesible via https://osf.io/6j7xr/?view_only=6417551c5125408fabcac75985e590c4.

Appendix B

Table A. Top 10 perceptual features per semantic category. The shared features across the groups are indicated in bold. The frequency of each feature was given in brackets.

Perceptual features								
Manipulable objects			Non-manipulable objects			Animals		
	Blind	Sighted		Blind	Sighted		Blind	Sighted
Rank	Feature	Feature	Rank	Feature	Feature	Rank	Feature	Feature
1	round [58]	colored/colorful [51]	1	tall/long [45]	tall/long [46]	1	two/four-footed [112]	colored/colorful [100]
2	tall/long [56]	round [45]	2	small [40]	small [31]	2	furry [72]	neck [96]
					propeller [31]			
					light [31]			
					white [31]			
3	colored/colorful [44]	small [35]	3	propeller [29]	big [30]	3	tall/long [64]	tall/long [91]
4	(with) shell [40]	tall/long [30]	4	(with) sound [25]	red [28]	4	big [60]	big [80]
		yellow [30]		(with) wing [25]				
5	small [34]	soft [27]	5	big [24]	high [24]	5	neck [59]	(with) horn [73]
					triangle [24]			
6	thin [29]	(with) shell [26]	6	speedy [21]	speedy [23]	6	(with) wing [47]	furry [72]
				high [21]	siren [23]			
				light [21]	stair [23]			
7	shape/shaped [27]	(with) smell [24]	7	siren [20]	(with) sound [21]	7	(with) sound [46]	two/four-footed [57]
								(with) wing [57]
8	handle [26]	orange [23]	8	railroad car [18]	railroad car [16]	8	(with) horn [34]	(with) tail [42]
9	soft [22]	thin [19]	9	flat [16]	shape/shaped [15]	9	speedy [32]	speedy [39]
10	(with) wheels [21]	light (weight) [18]	10	white [15]	(with) wing [13]	10	(with) tail [29]	small [35]

Table B. Top 10 non-perceptual features per semantic category. The shared features across the groups are indicated in bold. The frequency of each feature was given in brackets.

Non-perceptual features								
Manipulable objects			Non-manipulable objects			Animals		
	Blind	Sighted		Blind	Sighted		Blind	Sighted
Rank	Feature	Feature	Rank	Feature	Feature	Rank	Feature	Feature
1	plastic [75]	clean(ing) [69]	1	vehicle [120]	vehicle [96]	1	animal [318]	animal [339]
2	tool [54]	plastic [59]	2	place [69]	(on the) sea [65]	2	carnivore / herbivore [93]	wild [73]
3	clean(ing) [53]	wool [53]	3	(on the) sea [68]	passenger [44]	3	wild [57]	forest [68]
4	glass [49]	metal [52]	4	freightage [66]	place [41]	4	forest [56]	carnivore / herbivore [64]
					patient (noun) [41]			
5	metal [45]	fruit [50]	5	passenger [38]	metal [40]	5	domestic [37]	domestic [51]
					(in the) air [40]			
6	fruit [41]	portable [43]	6	transportation [36]	freightage [39]	6	animate [34]	mammal [46]
7	food [39]	food [34]	7	(in the) air [34]	(distribute) sound [36]	7	mammal [29]	reptile [39]
	wool [39]			(distribute) sound [34]				
8	fabric [35]	clothes [30]	8	emergency [31]	sport [32]	8	reptile [28]	desert [27]
							cute [28]	strong [27]
9	object/thing [29]	game [29]	9	iron [29]	emergency [30]	9	predatory [27]	dangerous [26]
	portable [29]							
10	fire/flame [28]	electric [28]	10	patient (noun) [24]	iron [29]	10	banana [23]	cute [25]
		tool [28]			Egypt [29]		cat/feline [23]	banana [25]
								predatory [25]

English Summary

We experience the world through different senses: we see, hear, smell, touch, and taste things. Each of these senses offers unique information but also certain limitations. Together, these determine how we understand objects and events, and thus concepts. For example, while a car passes by, we observe that it is a fast-moving sports car and also hear the whooshing noise the car makes. The visual and auditory cues together inform us about the speed of the car. But, what happens when one of these cues is absent, as in the experience of individuals who are blind from birth? Does this affect the way individuals describe this event? When we communicate about our experiences, we use different communicational means, such as words, hand gestures, and facial expressions. As with simple sensory experiences, each communication means has its own benefits and restrictions. For example, gesture can provide precise information of how an object moves whereas speech may not have the correct word in its vocabulary. In the current thesis, I investigate to what extent perceptual experience influences multimodal language use in speech and gesture as well as the underlying conceptual knowledge that gives rise to these visible behaviors. Therefore, I have focused on two key questions: First, does the perceptual input (e.g., visual or auditory) influence the way sighted individuals describe events in speech and gesture? Second, does having access to visual experience (i.e., being sighted or congenitally blind) influence how people use speech and gesture to express objects and events?

In Chapter 2, I investigated whether the perceptual input (vision, audition) influences event descriptions in speech and gesture. I conducted an experiment including a comparison of Turkish speakers' speech and gesture use to express different components of motion events (e.g., a woman runs away from the elevator.) These motion events were presented to speakers as audio-only, visual-only, or multimodal (visual + audio) clips. I found that when speakers watched events (in the visual-only or multimodal condition), they produced more motion event descriptions compared to when they only listened to the events (audio-only condition). Speakers in the visual conditions mentioned manner of motion (i.e., how a motion is performed) more often and path of motion (i.e., trajectory of motion) less often than speakers in the auditory condition. However, gesture use for path and manner did not differ between the audio-only and visual conditions. Together, the findings suggest that the perceptual input (visual and auditory) influences how speakers describe events in their speech but not in their gestures.

In Chapter 3, I investigated whether visual experience affects how speakers describe different components of motion events in speech and gesture by comparing individuals with blindness from birth to sighted counterparts. I presented motion events auditorily

to all speakers (e.g., hearing footsteps of a person walking into a room). I compared speech and gesture use of Turkish blind, blindfolded (sighted with covered eyes), and sighted speakers while describing motion events. I found that blind speakers used landmarks in events (e.g., a room) more in their speech compared to sighted speakers, and particularly they used self-anchored landmarks in their descriptions (i.e., referring to locations relative to their own body, such as "to/from my left" instead of saying "a room or an elevator"). Blind speakers divided events into smaller chunks using more path verbs (about trajectory of movement) but mentioned manner of motion (e.g., running vs. walking) less often than sighted speakers. Blind speakers also produced more pointing gestures to landmarks than sighted speakers but had fewer iconic gestures for path and manner of motion than sighted speakers. Together, the findings suggest that the lifetime experience of being blind influences how people describe motion events in both speech and co-speech gesture.

In Chapter 4, I investigated how single objects are depicted in gestures by blind and sighted individuals. I presented pre-recorded spoken words to participants and asked them to produce gestures to express concepts without speaking. I compared gesture strategies used by blind and sighted individuals for concepts that rely on visual (i.e., non-manipulable objects and animals) vs. motor (i.e., manipulable objects like tools) information to different extents. As a secondary task, I asked participants to list the core features of the same concepts. These features were coded as perceptual (i.e., information gained through a primary sensory channel such as size, shape, appearance, sound, body parts, and kinematic information) or non-perceptual (e.g., functional, taxonomic, or encyclopedic information). I found that blind participants produced fewer gestures compared to sighted participants for non-manipulable objects (e.g., a bridge, a house) and animal concepts. There was no group difference for manipulable objects (e.g., a banana, a spoon). Therefore, blind participants skipped a higher number of trials when concepts were predominantly dependent on visual rather than motor information. The gesture results overall showed that blind participants adjusted their gesture strategies according to their sensorimotor experience with objects. For the feature listing, blind participants reported fewer perceptual features for animals and fewer non-perceptual features for non-manipulable objects than sighted participants. Together, the findings suggest that visual experience influences how concepts are expressed in gestures and, to some degree, through written features.

Lastly, in Chapter 5, I bring together my research findings from Chapters 2 to 4 and formulate a conclusion. I conclude that overall, perceptual experience shapes how blind and sighted people express object and event concepts in multimodal language. Looking at multimodal language use, particularly gesture, can provide new insights into how perceptual experience affects our knowledge about concepts.

Nederlandse samenvatting

We ervaren de wereld via verschillende zintuigen: we zien, horen, ruiken, voelen en proeven dingen. Elk van deze zintuigen biedt unieke informatie, maar heeft ook bepaalde beperkingen. Samen bepalen deze hoe we objecten en gebeurtenissen, en dus concepten, begrijpen. Wanneer er bijvoorbeeld een auto voorbijrijdt, zien we dat het een snel rijdende sportwagen is en horen we ook het suizende geluid dat de auto maakt. De visuele en auditieve signalen informeren ons samen over de snelheid van de auto. Maar wat gebeurt er als een van deze signalen ontbreekt, zoals in de ervaring van individuen die vanaf hun geboorte blind zijn? Heeft dit invloed op de manier waarop individuen deze gebeurtenis beschrijven? Wanneer we over onze ervaringen communiceren, gebruiken we verschillende communicatiemiddelen zoals woorden, handgebaren en gezichtsuitdrukkingen. Net als bij eenvoudige zintuiglijke ervaringen heeft elk communicatiemiddel zijn eigen voordelen en beperkingen. Gebaren kunnen bijvoorbeeld nauwkeurige informatie verschaffen over hoe een object beweegt, terwijl dit in spraak niet makkelijk te omschrijven is. In dit proefschrift onderzoek ik in welke mate perceptuele ervaring het multimodale taalgebruik in spraak en gebaren beïnvloedt, evenals de onderliggende conceptuele kennis die aanleiding geeft tot dit zichtbare gedrag. Daarom heb ik mij geconcentreerd op twee belangrijke vragen: Ten eerste, beïnvloedt de perceptuele input (bijvoorbeeld visueel of auditief) de manier waarop ziende personen gebeurtenissen beschrijven in spraak en gebaren? Ten tweede, beïnvloedt het hebben van toegang tot visuele ervaringen (oftewel ziend of aangeboren blind zijn) de manier waarop mensen spraak en gebaren gebruiken om objecten en gebeurtenissen uit te drukken?

In Hoofdstuk 2 onderzocht ik of de perceptuele input (visueel, auditief) de beschrijving van gebeurtenissen in spraak en gebaren beïnvloedt. Ik heb een experiment uitgevoerd waarin ik een vergelijking maak tussen het spraak- en gebarengebruik van Turkse sprekers om verschillende componenten van motion events uit te drukken (bijvoorbeeld: een vrouw rent weg uit de lift). Deze motion events werden aan sprekers gepresenteerd als alleen audio, alleen visuele, of multimodale (visueel + audio) clips. Ik ontdekte dat sprekers meer beschrijvingen van motion events produceerden wanneer sprekers naar de gebeurtenissen keken (in de visuele of multimodale conditie) vergeleken met wanneer ze alleen naar de gebeurtenissen luisterden (alleen audio-conditie). Sprekers in de visuele conditie noemden de manner of motion (oftewel hoe een beweging wordt uitgevoerd) vaker en path of motion (oftewel het traject van de beweging) minder vaak dan sprekers in de alleen audio-conditie. Het gebruik van gebaren voor path of motion en manner of motion verschilde echter niet tussen de alleen audio- en visuele condities. Samen suggereren de bevindingen dat de perceptuele input (visueel en auditief) invloed heeft op de manier waarop sprekers gebeurtenissen beschrijven in hun spraak, maar niet in hun gebaren.

In Hoofdstuk 3 onderzocht ik of visuele ervaring invloed heeft op de manier waarop sprekers verschillende componenten van motion events in spraak en gebaren beschrijven door individuen met aangeboren blindheid te vergelijken met ziende individuen. Ik presenteerde bewegingsgebeurtenissen auditief aan alle sprekers (bijvoorbeeld het horen van voetstappen van een persoon die een kamer binnenkomt). Ik vergeleek het spraak- en gebarengebruik van Turkse blinde, geblinddoekte (ziend met bedekte ogen) en ziende sprekers tijdens het beschrijven van bewegingsgebeurtenissen. Ik ontdekte dat blinde sprekers meer oriëntatiepunten in gebeurtenissen (bijvoorbeeld een kamer) gebruikten in hun spraak dan ziende sprekers, en vooral dat ze zelfverankerde oriëntatiepunten gebruikten in hun beschrijvingen (d.w.z. verwijzen naar locaties ten opzichte van hun eigen lichaam, zoals "van/naar mijn linkerkant" in plaats van te zeggen "een kamer of een lift"). Blinde sprekers verdeelden gebeurtenissen in kleinere stukken met meer werkwoorden gerelateerd aan path of motion (over het traject van de beweging), maar noemden de manner of motion (bijvoorbeeld rennen versus lopen) minder vaak dan ziende sprekers. Blinde sprekers produceerden ook meer wijzende gebaren naar oriëntatiepunten dan ziende sprekers, maar hadden minder iconische gebaren voor path of motion en manner of motion dan ziende sprekers. Samen suggereren de bevindingen dat de levenslange ervaring van blind zijn van invloed is op de manier waarop mensen bewegingsgebeurtenissen beschrijven in zowel spraak als co-spraakgebaren.

In hoofdstuk 4 heb ik onderzocht hoe afzonderlijke objecten in gebaren worden weergegeven door blinde en ziende individuen. Ik presenteerde vooraf opgenomen gesproken woorden aan de deelnemers en vroeg hen gebaren te maken om concepten uit te drukken zonder te spreken. Ik heb gebaarstrategieën vergeleken die door blinde en ziende individuen worden gebruikt voor concepten die in verschillende mate afhankelijk zijn van visuele (oftwel niet-manipuleerbare objecten en dieren) versus motorische (oftewel manipuleerbare objecten zoals gereedschappen) informatie. Als secundaire taak vroeg ik de deelnemers om de essentiele kenmerken van dezelfde concepten op te sommen. Deze kenmerken werden gecodeerd als perceptueel (oftewel informatie verkregen via een primair sensorisch kanaal zoals grootte, vorm, uiterlijk, geluid, lichaamsdelen en kinematische informatie) of niet-perceptueel (bijvoorbeeld functionele, taxonomische of encyclopedische informatie). Ik ontdekte dat blinde deelnemers minder gebaren produceerden vergeleken met ziende deelnemers voor niet-manipuleerbare objecten (bijvoorbeeld een brug, een huis) en dieren. Er was geen verschil tussen de groepen voor manipuleerbare voorwerpen (bijvoorbeeld een banaan, een lepel). Daarom sloegen blinde deelnemers een groter aantal rondes over wanneer concepten voornamelijk afhankelijk waren van visuele in plaats van motorische informatie. De resultaten met betrekking tot gebaren lieten over het algemeen zien dat blinde deelnemers hun gebaarstrategieën aanpasten op basis van hun sensomotorische ervaring met objecten. Voor de lijst met kenmerken rapporteerden blinde deelnemers minder perceptuele kenmerken voor dieren en minder niet-perceptuele kenmerken voor niet-manipuleerbare objecten dan ziende deelnemers. Samen suggereren de bevindingen dat visuele ervaring invloed heeft op de manier waarop concepten worden uitgedrukt in gebaren en, tot op zekere hoogte, via geschreven kenmerken.

Ten slotte breng ik in hoofdstuk 5 mijn onderzoeksresultaten uit de hoofdstukken 2 tot en met 4 samen en formuleer ik een conclusie. Ik concludeer dat de perceptuele ervaring bepaalt hoe blinde en ziende mensen object- en gebeurtenisconcepten in multimodale taal uitdrukken. Kijken naar multimodaal taalgebruik, in het bijzonder gebaren, kan nieuwe inzichten opleveren in hoe perceptuele ervaring onze kennis over concepten beïnvloedt.

Curriculum Vitae

Ezgi Mamus was born in 1990 in Istanbul, Turkey. She obtained her bachelor's degree in Pyschology (2013) and her master's degree in Psychological Sciences (2016) from Boğaziçi Üniversitesi in Istanbul, Turkey. During her bachelor's and master's studies, she worked as a research assistant at the Autobiographical Memory Lab (under the supervision of Dr. Ali Tekcan) and Cognitive Processes Lab (under the supervision of Dr. Aysecan Boduroglu). During this period, she conducted a study on how visual input influences access to autobiographical involuntary memories and characteristics of such memories by examining this process in early blind and sighted individuals. In her master thesis, she investigated the necessity of context and semantic consistency in a visual memory phenomenon, known as Boundary Extension. Following her master's degree, she worked as a research assistant in an EU-Horizon 2020 project, *the L2TOR Project*, which explored the use of social robots in second language acquisition.

In 2017, she began her PhD research in the Multimodal Language and Cognition group of the Radboud University with Prof. Aslı Özyürek and Prof. Asifa Majid. During her PhD, she actively participated in the organization committees of various workshops and conferences. Additionally, she supervised internship students and worked part-time as a lecturer at the Department of Language and Communication of the Radboud University. Currently, she works as a postdoctoral researcher in the Multimodal Language Department at the Max Planck Institute for Psycholinguistics in Nijmegen.

Publications

Thesis-related publications

Mamus, E., Speed, L. J., Ortega, G., Majid, A., & Ozyurek, A. (revised & resubmitted). Gestures reveal how visual experience shapes concepts in blind and sighted individuals.

Mamus, E., Speed, L. J., Rissman, L., Majid, A., & Özyürek, A. (2023). Lack of visual experience affects multimodal language production: Evidence from congenitally blind and sighted people. *Cognitive Science*, *47*(1), e13228. https://doi.org/10.1111/cogs.13228

Mamus, E., Speed, L. J., Özyürek, A., & Majid, A. (2023). The effect of input sensory modality on the multimodal encoding of motion events. *Language, Cognition and Neuroscience, 38*(5), 711-723. https://doi.org/10.1080/23273798.2022.2141282

Mamus, E., Speed, L. J., Ozyurek, A., & Majid, A. (2021). *Sensory modality of input influences encoding of motion events in speech but not co-speech gestures*. In T. Fitch, C. Lamm, H. Leder, & K. Teßmar-Raible (Eds.), Proceedings of the 43rd Annual Conference of the Cognitive Science Society (CogSci 2021) (pp. 376-382). Vienna: Cognitive Science Society.

Mamus, E., Rissman, L., Majid, A., & Ozyurek, A. (2019). *Effects of blindfolding on verbal and gestural expression of path in auditory motion events*. In A. K. Goel, C. M. Seifert, & C. C. Freksa (Eds.), Proceedings of the 41st Annual Meeting of the Cognitive Science Society (CogSci 2019) (pp. 2275-2281). Montreal, QB: Cognitive Science Society.

Other publications

Ünal, E., **Mamus, E.**, & Özyürek, A. (2023). Multimodal encoding of motion events in speech, gesture, and cognition. *Language and Cognition*. Advance online publication. doi:10.1017/langcog.2023.61.

Belpaeme, T., Vogt, P., Van den Berghe, R., Bergmann, K., Göksun, T., De Haas, M., Kanero, J., Kennedy, J., Küntay, A. C., Oudgenoeg-Paz, O., Papadopoulos, F., Schodde, T., Verhagen, J., Wallbridge, C. D., Willemsen, B., De Wit, J., Geçkin, V., Hoffmann, L., Kopp, S., Krahmer, E., **Mamus, E.**, Montanier, J.-M., Oranç, C., & Pandey, A. K. (2018). Guidelines for designing social robots as second language tutors. *International Journal of Social Robotics, 10*(3), 325-341. doi:10.1007/s12369-018-0467-6.

Kanero, J., Geçkin, V., Oranç, C., **Mamus, E.**, Küntay, A. C., & Göksun, T. (2018). Social robots for early language learning: Current evidence and future directions. *Child Development Perspectives, 12*(3), 146-151. doi:10.1111/cdep.12277.

Mamus, E., & Karadöller, D. Z. (2018). Anıları Zihinde Canlandırma [Imagery in autobiographical memories]. In S. Gülgöz, B. Ece, & S. Öner (Eds.), *Hayatı Hatırlamak: Otobiyografik Belleğe Bilimsel Yaklaşımlar [Remembering Life: Scientific Approaches to Autobiographical Memory]* (pp. 185-200). Istanbul, Turkey: Koç University Press.

Mamus, E., & Boduroglu, A. (2018). The role of context on boundary extension. *Visual Cognition, 26*(2), 115-130. doi:10.1080/13506285.2017.1399947

A

Acknowledgements

Reflecting on my PhD journey, I resonate with the words of Douglas Adams: “I love deadlines. I like the whooshing sound they make as they fly by”. I haven’t rushed to complete my PhD; the duration speaks for itself. Instead, I have tried to enjoy the ride. Just like many fellow PhDs, I have faced several common challenges, such as the pandemic, and personal ones like a long-term knee injury. Several individuals played a part in shaping this journey, directly or indirectly contributing to the development and completion of this thesis. Therefore, I want to express my gratitude to each one of them, as well as to the journey itself for all the valuable lessons it has taught me.

First and foremost, it was a privilege to learn from two exceptionally intelligent and inspiring scholars—my co-promoters, Aslı and Asifa.

Aslı, thank you for having faith in me, someone aspiring to enter the field of psycholinguistics despite limited prior knowledge and expertise. Your support has been invaluable and has made a significant impact on my confidence and success. Thank you for always keeping your door open and taking the time to talk whenever I needed it. It was a delight to be a member of such a motivating and encouraging research atmosphere, and I always felt warmly welcomed.

Asifa, I want to express my heartfelt gratitude to you for many reasons. You have always been generous in sharing your knowledge and providing insightful and thought-provoking criticism. You have encouraged me to make my own decisions. Even when you were physically distant, you consistently looked out for my well-being. I am indebted to you for everything. I hope our paths cross again in the future.

I would also like to express my gratitude to the other members of my supervision team. **Laura**, I am thankful to have you on the supervision team. Despite joining the team later, your contributions have greatly improved my work. Thank you for always being ready to lend a hand and doing so with a smile on your face. **Lilia Rissman**, although we only worked together during the first year of my PhD, it was great to learn from your knowledge and enthusiasm for research. Thank you also for initiating those cool social gatherings.

I would also like to extend my sincere thanks to the members of the thesis committee and examination board, **Antje Meyer**, **Aylin Küntay**, **Judith Holler**, **Olivier Collignon**, **Sotaro Kita**, and **Susanne Brouwer**, for finding the time to read and evaluate this thesis.

I would like to thank the members of Meaning, Culture, and Cognition Lab: **Alex C.**, **Claudia**, **Ewelina**, **Hasan E.**, **Ilja**, **John**, **Lila**, **Mariann**, and **Saskia**, thank you for your feedback and support during the first year of my PhD, and for the fun celebrations whenever there is a new publication.

I would like to thank all past and present members of the Multimodal Language Lab and Department: **Alessia**, **Berna**, **Beyza**, **Brigit**, **Chinmaya**, **Dota**, **Gerardo**, **Hasan D.**, **James**, **Jessica**, **Junfei**, **Kazuki**, **Kimberley**, **Kristel**, **Lanmiao**, **Linda**, **Loïs**, **Louise**, **Maddy**, **Marlijn**, **Maryam**, **Mounika**, **Paula**, **Pelin**, **Renske**, **Sharice**, **Sjoerd**, **Tom**, **Vicky**, **Wim**, **Yaren**, **YaSuei**, and **Zeynep**, thank you for your feedback and support, as well as for the long lunches, fun lab outings, and delicious foods from all around the world. **Erce**, thank you for sharing the passion of picking wild blueberries from the bushes around MPI to make gin and tonics. Also, thank you for providing feedback whenever I need it. I wish we had more time together at MPI. **Francie**, thank you for being such an amazing office buddy in the greenest office at MPI, for all the help with R, and for the fantastic DIY activities we had. **Marlou**, thank you for sharing the challenges of the last months of our PhDs, I miss having you in the office.

I would like to express my gratitude to the blind community in Turkey and my participants for dedicating their time and sharing their insights with me during the experiments. In particular, I would like to thank **GETEM** at Boğaziçi University, the center for technology and education for visually impaired individuals, and **Engin Yılmaz**, the director of GETEM, for their invaluable support. A huge thank you to the students who worked with me as assistants or interns: **Sevilay**, thank you for coming over to my place for training during those tough days when I could not walk and come to the office. Also, thanks for coding gestures for long hours while still maintaining such a positive vibe. **Uğur** and **Hüma**, thank you for massive amount of coding and consistently challenging me with your questions. **Ayşe Serra**, **Esma**, and Şevval, thank you for all your help with transcription and annotation of event descriptions.

I would like to express my gratitude to the **MPI community**. While CLS provided financial support for my doctoral research, I had the advantage of having an office at MPI and benefiting from their technical and scientific support. A big shout-out to **Technical Group** for consistently offering solutions to all of my requests during my studies. **Jeroen**, thank you for your help with the audio setup, working together with me to figure out the direction of each sound, and handling all the video data processing. **Maarten**, thank you for your help with the Kinect script and its synchronization with my setup. **Alex D.**, thank you for your patience in helping me with many software and hardware issues. **Kevin**, our lovely PhD coordinator, thank you for several fun pizza

nights and interesting IMPRS activities, and for offering support and encouragement whenever it was needed. Last but not least, dear MPI volleyball team members—**Ad**, **Ine**, **Lara**, **Marijn**, **Marvin**, **Maurice**, and **Reiner**—thank you for all the fun on Thursday evenings and for making me a better player. I could set aside all my concerns and simply cherish our time playing together.

A big thank you to my lovely paranymphs for the defense preparations. **Anita**, thank you for all the funny moments we shared in the office and on the balcony of our house. I wasn't expecting to meet such a good friend at the end of my PhD journey. **Sho**, thank you for always being eager to help whenever I stop by your office, and for being cheerful all the time. **Yeşim**, I am grateful to that vierdaagse when our friendship truly began. You've been the best partner in crime when it comes to fun. Thank you for being a lively little sister to me.

To my friends whose paths crossed with mine in Nijmegen: **Demet**, my ultimate quarantine buddy, it was wonderful having you in Nijmegen. Thank you for the fun times in Istanbul and Nijmegen, and all the roads in between. **Melis** and **Figen**, thank you for the coffee breaks, lunches, nights spent watching series, and for bearing with me through the challenges of finishing up. **Nazlı**, thank you for being there for me whenever I needed a hug.

My non-academic friends, **Deniz** and **Didem**, I am thankful for your presence in my life, the emotional support, the Pomodoro days during the pandemic, and for sharing the dream of having a glass of uzo with some tarama during the darkest times.

Many thanks to my dear friends (aka Dallas) who have been part of my life since those lovely years at Boğaziçi: **Aslı Aktan Erciyes**, always ready to make a voodoo doll when I need one, thank you for all the laughter we shared. **Burcu Kaya Kızılöz**, thank you for being on the other end of the phone when I don't know where to go. **Dilay Karadöller Astarlıoğlu**, thank you for sharing your sisterhood with me during our time together at the MPI. **Müge Özbek**, thank you for all the beautiful memories we made in Paris, Berlin, and Aarhus. **Müge Özvarol**, thank you for being the ultimate partner in crime and for always finding a way to brighten my day. **Melisa Akan** and **Pınar Kurdoğlu Ersoy**, thank you for the support even though we were miles apart.

To my parents, **Semin** and **Yaşar**, my brother, **Burak**, and my beloved grandmother, **Gönül**: İstediğim şeylerin peşinden gitmemi destekleyen, benim iyiliğimi önceliklendiren ve binbir huysuzluğumu çeken canım annem, babam ve abim iyi ki varsınız. Hem doktorada hem de bugüne kadar yanımda olduğunuz için teşekkürler. Bana okuma sevgisi aşılayan canım Ananecim, hep kalbimdesin.

A special thank you goes to an amazing woman, **Gamze Sofuoğlu**. I am lucky to have you as a friend who bears with the long hours of phone chats. I cherish all the discussions and disagreements we have. Thank you for being a constant source of inspiration.

And finally, **Özcan**, this thesis owes its existence not only to your creative ideas and technical skills, which played a huge role in making the sounds for my experiments, but also to your endless love and patience. You always had more faith in me than I had in moments of despair. Thank you for being my constant home wherever we are.

MAX PLANCK INSTITUTE
FOR PSYCHOLINGUISTICS

VISITING ADDRESS
Wundtlaan 1
6525 XD Nijmegen
The Netherlands

POSTAL ADDRESS
P.O. Box 310
6500 AH Nijmegen
The Netherlands

CONTACT
T +31(0)24 3521 911
F +31(0)24 3521 213
E info@mpi.nl
Twitter @MPI_NL
www.mpi.nl

CLS | Centre for Language Studies
Radboud University