BRIDGING *WHEN* AND *FOR WHOM*: GESTURE PROCESSING DURING LANGUAGE COMPREHENSION

DEMET ÖZER

KOÇ UNIVERSITY
2020

BRIDGING *WHEN* AND *FOR WHOM*: GESTURE PROCESSING DURING LANGUAGE COMPREHENSION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES AND HUMANITIES
OF
KOÇ UNIVERSITY

BY

DEMET ÖZER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF PSYCHOLOGY

DECEMBER 2020

Koç University
Graduate School of Social Sciences and Humanities

This is to certify that I have examined this copy of a doctoral thesis by
DEMET ÖZER
and have found that it is complete and satisfactory in all respects, as a thesis for the degree of Doctor of Philosophy.

Committee Members:

Assoc. Prof. Tilbe Göksun *advisor*	Koç University
Prof. Aylin C. Küntay	Koç University
Prof. Aslı Özyürek	Max Planck Institute for Psycholinguistics
Prof. Sami Gülgöz	Koç University
Prof. Spencer Kelly	Colgate University

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Demet Özer

December 20, 2020

Dissertation Abstract

Demet Özer, "Bridging *When* and *For Whom*: Gesture Processing During Language Comprehension"

Speakers use iconic co-speech hand gestures, which are spontaneous hand movements that co-occur with speech and represent concrete object attributes, events, and actions. Gestures help speakers and listeners during communication and thinking. However, the beneficial effects of gestures exhibit variation across individuals and situations. This dissertation investigates how and to what extent listeners' attention to and processing of gestures show variation as a function of (1) individuals' cognitive dispositions, and (2) the semantic overlap between speech and gesture. In Chapter 2, I reviewed the current evidence on how individual differences in cognitive abilities relate to (1) *gesture use*: how frequently speakers use gestures and how and to what extent speakers benefit from using gestures during speaking and thinking and (2) *gesture processing*: how and to what extent listeners attend to, process, and benefit from observing gestures during comprehension and learning across different ages. This review provided a detailed assessment of the field's current state by taking a life-span perspective (starting from childhood through adulthood to healthy aging) and bridging fields of gesture production and gesture processing. I suggested that speakers might use gestures as a compensatory tool during speaking and thinking that interact with their cognitive dispositions. Chapter 2 revealed little research on the comprehension domain, and it is unknown how individual differences in different cognitive abilities relate to listeners' attention to and processing of gestures during language comprehension. Based on this, in Chapter 3, I investigated how individual differences in visual-spatial and verbal abilities related to the processing of concurrent spoken vs. gestural expressions in a mismatch paradigm (Kelly, Özyürek, & Maris, 2010). The results replicated earlier findings showing that speech and gesture bidirectionally affect each other during comprehension in a different language, Turkish. Moreover, I demonstrated that this is not a monolithic process and show variation across individuals. The results showed that listeners with higher visual-spatial abilities were better at processing gestures, whereas people with higher verbal abilities were better at processing speech. These findings suggest that visual-spatial cognitive resources are critical to process visual semantic information conveyed by gestures. Chapter 4 investigated how and to what extent the relative information value of gestures indexed by the semantic overlap between speech and gesture affected listeners' attention to gestures by employing the behavioral and eye-tracking methodology. In particular, I asked whether and how (1) listeners allocated overt visual attention to gestures that express complementary (e.g., saying here and gesturing to the right side) vs. redundant (e.g., saying right and gesturing to the right side) information to the accompanying speech and (2) individual differences in gazing at gestures related to how much listeners benefitted from seeing gestures during the comprehension of spatial relations between objects. I found that listeners fixated to gestures more when they complemented spoken deictic constructions (here) than when they expressed redundant information. Overt gaze attention to gestures did not predict how much information listeners gleaned from gestures. These results suggest that the perceived relative communicative value of gestures modulates listeners'

attention to gestures; however, direct visual attention to gestures does not relate to information uptake from gestures as they can be processed peripherally. Overall, in this dissertation, I demonstrated that individual differences in visual-spatial cognitive resources and the relative communicative value of gestures across different contexts modulated listeners' attention to co-speech gestures. I contributed to the current literature by demonstrating that the so-called beneficial effects of gestures are not monolithic and show variation across individuals and situations. This dissertation demonstrated that gesture accounts and future research should consider different internal and external factors to have a full picture of *when*, *why*, and *for whom* gestures exhibit their supposed roles.

Tez Özeti

Demet Özer, "*Ne Zaman* ve *Kimin İçin*? El Jestlerini Anlamada Farklılıklar"

Konuşmacılar, iletişim esnasında sözlü dillerine eşlik eden ve somut nesne niteliklerini, olayları, hareketleri temsil eden ikonik el jestleri kullanırlar. Hem konuşmacılar hem de dinleyiciler iletişim ve düşünme esnasında bu ikonik el jestlerinden yararlanırlar. Bununla birlikte, jestlerin yararlı etkileri, bireyler ve durumlar arasında farklılıklar gösterir. Bu tezde, (1) bireylerin bilişsel eğilimleri ile (2) sözlü dil ve sözlü dile eşlik eden jest arasındaki anlamsal örtüşmenin, dinleyicilerin el jestlerini işlemlemesinde nasıl ve ne ölçüde farklılıklara yol açtığı araştırılmıştır. Bölüm 2'de, bilişsel yeteneklerdeki bireysel farklılıkların jest kullanma ve jest işlemleme ile nasıl ilişkili olduğuna dair mevcut kanıtları gözden geçirdim. Bu bölümde, (1) *jest üretimi:* konuşmacıların jestleri ne sıklıkla kullandığı ve iletişim ile düşünme sırasında jestleri kullanmaktan ne derece istifade ettiği ile (2*) jest işleme-anlama:* dinleyicilerin, konuşmacıların kullandığı jestleri nasıl ve ne derece işlemlediği, anladığı, dikkat ettiği ile iletişim ve öğrenme esnasında bu jestleri görmekten ne derece istifade ettiği ile ilişkili kişisel farklılıklar derlenmiştir. Bu inceleme, yaşam boyu perspektifi (çocukluktan yetişkinliğe ve sağlıklı yaşlanmaya kadar) alarak ve jest üretimi ile jest işleme alanlarını birleştirerek alanın mevcut durumuna ilişkin ayrıntılı bir değerlendirme sağlamıştır. Bu bölümde, konuşmacıların hem iletişim hem de düşünme esnasında el jestlerini, düşük bilişsel yetilerini telafi edici bir araç olarak kullanabilecekleri önerilmiştir. Bölüm 2'deki derleme, özellikle jest işleme alanındaki kişisel farklılıklar üzerine çok az araştırma olduğunu ve farklı bilişsel yeteneklerdeki bireysel farklılıkların, dinleyicilerin jestlere olan dikkat ve işlemlemesini nasıl etkilediğinin bilinmediğini ortaya koymuştur. Buna dayanarak Bölüm 3'de, görsel-uzamsal ve sözel yeteneklerdeki bireysel farklılıkların eş zamanlı iletilen ve birbirinden uyumsuz anlamsal bilgiler içeren sözlü dil ve jesti anlamadaki etkisini araştırdım (Kelly, Özyürek ve Maris, 2010). Sonuçlar, alan yazındaki öncel bulguları farklı bir dilde (Türkçe'de) tekrarlanmış ve sözlü dil ile jestin dil anlama esnasında birbirlerini karşılıklı olarak etkilediğini göstermiştir. Daha da önemlisi, bu araştırma, bunun monolitik bir süreç olmadığını ve bireyler arasında farklılık olabileceğini göstermiştir. Sonuçlar, daha yüksek görsel-uzamsal yeteneklere sahip dinleyicilerin jestleri işlemede, daha yüksek sözel yeteneklere sahip kişilerin ise sözlü dili işlemede daha iyi olduğunu gösterdi. Bu bulgular, görsel-uzamsal bilişsel kaynakların, jestler yoluyla iletilen görsel anlamsal bilgiyi işlemek için kritik öneme sahip olduğunu göstermektedir. Bölüm 4'de, davranışsal metodolojiye ek olarak göz izleme metodolojisi de kullanılmış ve sözlü dil ile sözlü dile eşlik eden jestler arasındaki anlamsal örtüşmenin, dinleyicilerin el jestlerine olan görsel dikkatini nasıl ve ne ölçüde etkilediği araştırılmıştır. Daha detaylı olarak, bu bölümde, (1) dinleyicilerin, sözlü dili tamamlayıcı bilgiler ifade eden jestlere (örneğin, "burada" deyip sağ tarafı gösteren bir jest yapmak), sözlü dil ile aynı anlama gelen bilgiler ifade eden jestlere (örneğin, "sağ" deyip sağ tarafı gösteren bir jest yapmak) oranla daha fazla görsel dikkat gösterip göstermediğini ve (2) jestlere gösterilen direkt görsel dikkatteki bireysel farklılıkların, dil anlama esnasında jestlerin pekiştirici rolü ile nasıl ve ne ölçüde ilgili olduğu araştırılmıştır. Sonuçlar, dinleyicilerin, sözlü dili tamamlayan el jestlerine daha fazla direkt

görsel dikkat gösterdiğini; ancak, jestlere yönelik doğrudan görsel dikkatin jest anlama ile ilgili olmadığını göstermiştir. Genel olarak, bu tezde, (1) görsel-uzamsal bilişsel kaynaklardaki bireysel farklılıkların ve (2) el jestlerinin sözlü dile göreli iletişimsel değerinin, dinleyicilerin jestlere olan dikkat ve işlemlemesinde farklılıklara yol açtığını gösterdim. Bu tez, alan yazında sıkça vurgulanan jestlerin faydalı etkilerinin monolitik olmadığını ve bireyler ile durumlar arasında değişkenlik gösterebileceğini göstererek güncel literatüre katkıda bulunmaktadır. İleriki çalışmalar, jestlerin pekiştirici etkisinin ne zaman, neden ve kimler için olduğunu anlamak için, bu süreçte farklılık yaratabilecek pek çok faktörü dikkate almalıdır.

ACKNOWLEDGMENTS

After a very long journey with all the ups and downs, I still can't believe that I finally get to close this chapter of my life. None of the work written in this thesis would be possible without the help, support, encouragement, and love of many amazing people I had the luck to walk with on the way.

First of all, I would like to thank my advisor, *Tilbe Göksun.* Throughout the years, one feeling always accompanied me on the way working with her: deep gratitude. She always believed in me every step that I have taken – sometimes more than I believed in myself. I have learned many things from her over the years, but the most important one is that one can be very successful and hardworking, but also kind, supportive, and fun at the same time! Her enthusiasm was and always will be an inspiration for me. *Hocam*, thanks for being such an amazing supervisor! This thesis would not have been possible without your tremendous support and encouragement in every means possible.

I also would like to thank *Aslı Özyürek* for her guidance, generous support, and critical evaluation of my work. Aslı Hocam, every conversation we had was an academic inspiration! I also would like to thank the other committee members, *Aylin Küntay*, *Sami Gülgöz,* and *Spencer Kelly* for reading this thesis and sharing their encouraging and valuable comments. Thank you.

I feel so lucky and grateful for having had the chance to work among brilliant members of *Language and Cognition* and *Language and Communication Development* Labs. Our lab, famously known as SOS B20, was almost a second home to me in which I always found a friend to cheer me up in times of despair. I would like to thank all previous and current members with whom I had the pleasure to work with: *Aslı, Cansu, Hilal, Niloofar, Hazal, Zeynep, Sümeyye, Berna, Elif, Ezgi, Seda, Burcu, Ahmet, Erim, Emir, Mert, Işıl, Gyulten, Feyza, Esma, Deniz, Süleyman.* Also, my beautiful friends: *Gamze, Tutku, Esin,* and *Begüm*. These wonderful colleagues and friends always provided me an intellectually enriching environment, but more importantly, also an enjoyable one! They were the ones who made graduate life bearable for me. Thank you.

Finally, my family! Their endless love, support, and encouragement mean everything to me. It is rather easy to follow your dreams when you know that on the way, there will be people who will encourage you unconditionally. *My mom* and *my dad*; my sisters, *Deniz* and *Derya*; my brothers, *Serhan* and *Emir*; and my little ones, *Ege*, *Elif*, and *Zeynep*: thanks for being there no matter what.

FUNDING

This work was supported by International Research Fellowship Programme for Ph.D. Students (2214-A) awarded by The Scientific and Technological Research Council of Turkey (TÜBİTAK) to Demet Özer and TÜBA-GEBİP 2018 award given (Turkish Academy of Sciences Outstanding Young Scientist Award) and a James S. McDonnell Foundation Scholar Award (Grant no: 220020510) to Tilbe Göksun.

To my family…

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

GENERAL INTRODUCTION

Natural face-to-face communication is a multimodal activity in which two or more interlocutors interact by exchanging several communicative cues. This multimodal human communication includes extra-linguistic information such as visible speech (i.e., lip movements), body posture, eye-gaze, pitch accent, and hand gestures. This dissertation focuses on one of these multimodal cues: co-speech iconic hand gestures (henceforth; gestures), which are spontaneous hand movements that co-occur with relevant speech segments (e.g., words, phrases) and represent concrete object attributes, events, and actions.

Gesture and speech have different representational properties. Gestures are realized in visual-manual modalities and convey information holistically, whereas speech is realized in auditory-vocal modalities and expresses information linearly in a segmented and hierarchical fashion (Kendon, 1994; McNeill, 1992; Özyürek, 2014). Despite their differences in representational format, gestures are semantically, pragmatically, and temporally related to the accompanying speech. Gesture and speech together 'constitute language' (McNeill, 1992) and form a 'composite signal' (Clark, 1996) during utterance generation and comprehension.

Speakers spontaneously produce gestures as they speak or think. Listeners, in turn, attend to the speaker's gestures that facilitate comprehension and learning (Dargue, Sweller, & Jones, 2019; Hostetter, 2011; Özyürek, 2014). However, gesturing is not mandatory for communication. How speakers and listeners employ gestures exhibit variation across individuals and situations. The variation can emerge from two factors. First, internal factors, such as individuals' cognitive dispositions, lead to variations in gesture use and processing. People exhibit individual differences in how frequently they use spontaneous gestures when they speak or think, to what extent they attend to the speaker's gestures, and how much they benefit from using and observing gestures depending on their cognitive abilities (e.g., Chu, Meyer, Foulkes, & Kita, 2014; Gillespie, James, Federmeier, & Watson, 2014; Göksun, Goldin-Meadow, Newcombe, & Shipley, 2013; Hostetter & Alibali, 2007, 2011). Second, external factors such as the content of information (e.g., spatial vs. non-spatial, Alibali, 2005) or the semantic overlap between speech and gestures (e.g., redundant vs. complementary; Hostetter 2011) affect how and to what extent people employ gestures. Previous studies have examined variations of gesture *use* based on internal and external factors (e.g., Chu et al., 2014; Gillespie et al., 2014; Hostetter & Alibali, 2007, 2011; Hostetter & Pothoff, 2012), but less is known how gesture *processing* can be affected by these factors (Aldugom, Fenn, & Cook, 2020; Wu & Coulson, 2014a, 2014b). To fill this gap in the literature and provide evidence on variation in multimodal comprehension, in this dissertation, I aim to investigate internal and external factors that affect listeners' attention to and processing of gestures during comprehension. Based on the gaps in the current literature, we focused on two possible factors that can lead to variations in gesture processing: individuals' own cognitive dispositions (i.e., internal factor) and the semantic overlap between speech and gesture (i.e., external factor). In particular, I aim to study whether and how (1) listeners' visual-spatial and verbal abilities relate to the processing of concurrent gestural vs. spoken expressions and (2) listeners attend to gestures that express non-redundant (complementary) vs. redundant information in relation to the accompanying speech.

1.1. The role of gestures in speech comprehension

Gestures convey visual semantic information picked up by listeners (Kendon, 1994; Hostetter, 2011). Gestures are analyzed for meaning and subject to semantic processing analogous to other content laden stimuli like pictures or words (Bernardis & Caramelli, 2009; Gunter & Bach, 2004; So, Yi-Feng, Yap, Kheng, & Melvin Yap, 2013; Wu & Coulson, 2005, 2007; Yap, So, Melvin yap, Tan, & Teoh, 2011). Listeners integrate semantic information gleaned from speech and gesture to have a unified representation of the message (Alibali, Flevares, & Goldin-Meadow, 1997; Cassell, McNeill, & McCullough, 1999; Kelly, Barr, Church, & Lynch, 1999). The Integrated Systems Hypothesis (Kelly, Özyürek, & Maris, 2010) states that speech and gesture mutually and obligatorily affect each other during comprehension. These two types of information create meaning synergistically (i.e., in an interactive way), each serving as a context for the interpretation of the other. Electrophysiological and neuroimaging evidence shows that observing gestures affects listeners' processing of the concurrent speech at various levels spanning from early sensory/phonological to late semantic and syntactic processing (e.g., Biau & Soto-Faraco, 2013; Hubbard, Wilson, Callan, & Dapretto, 2009; Holle & Gunter, 2007; Holle et al., 2012; Özyürek, Willems, Kita, & Hagoort, 2007; Kelly, Kravitz, & Hopkins, 2004; Wang & Chu, 2013; Willems, Özyürek, & Hagoort, 2007, 2009). Additionally, the neural instantiation of speech vs. gesture processing exhibits similar properties, suggesting that they are inextricably linked during language comprehension (Straube, Green, Weis, & Kircher, 2012; Wu & Coulson, 2005).

Gestures provide additional visual semantic information to the accompanying speech and benefit listeners' comprehension (Beattie & Shovelton, 1999a, 1999b, 2001; Graham & Argyle, 1975; Holler, Shovelton, & Beattie, 2009; see Dargue et al., 2019 and Hostetter, 2011 for a meta-analysis). Beattie and Shovelton (1999a, 1999b, 2001) showed that listeners comprehended the size, shape, and location of objects better when speakers used gestures along with speech compared to when they did not. Gestures facilitate comprehension because most of the time, they provide additional visual information to the accompanying speech, particularly for visual-spatial information. Thus, the semantic overlap between speech and gesture is an important factor for gestural enhancement outcomes. Gestures that expressed incongruent information in relation to the concurrent speech hindered listeners' comprehension (e.g., Kelly, Kravitz, & Hopkins, 2004;

Kelly et al., 2010; Özyürek et al., 2007). On the other hand, gestures that disambiguate concurrent speech facilitated comprehension (Holle & Gunter, 2007; Holle et al., 2012; Kelly et al., 1999). For example, listeners understood indirect requests better when they were complemented with pointing gestures (e.g., "It is getting hot in here" complemented with a pointing gesture towards the window) compared to when there was no gesture (Kelly et al., 1999). Similarly, listeners benefitted from seeing gestures when understanding homonyms' intended meaning (Holle & Gunter, 2007). Indeed, meta-analyses showed that gestures that express non-redundant information to the accompanying speech benefit listeners' comprehension more than gestures that convey redundant information (Dargue et al., 2019; Hostetter, 2011).

Although the Integrated Systems Hypothesis suggested that listeners attend to and integrate gestures with spoken information in a mandatory and automatic manner (Kelly et al., 2010; Kelly, Creigh, & Bartolotti, 2010), recent evidence suggests that these processes are not monolithic and exhibit significant variation across individuals and situations. Several internal and external factors affect how and to what extent listeners attend to, process, and benefit from observing gestures during comprehension and learning. One important internal factor is individuals' own cognitive dispositions. People exhibit individual differences in gesture processing, depending on their propensities to use different cognitive resources. Despite being addressed extensively in the gesture production literature (e.g., Baxter, Winters, & Hammer, 1968; Chu, Meyer, Foulkes, & Kita, 2014; Gillespie, James, Federmeier, & Watson, 2014; Göksun, Goldin-Meadow, Shipley, & Newcombe, 2013; Hostetter & Alibali, 2007, 2011; Hostetter & Potthoff, 2012; Trafton et al., 2006), there is little research in the comprehension domain (but see Aldugom et al., 2020; Wu & Coulson, 2014a, 2014b). It is mostly unknown how individual differences in different cognitive abilities relate to listeners' gesture processing. Visual-spatial cognitive resources are critical for processing gestures, and individual differences in this capacity might lead to variation across individuals on how they process gestures (Kelly & Goldsmith, 2004).

How people employ gestures also exhibit variation across situations. Individuals use and process gestures differently under different circumstances. Evidence showed that many external factors affect listeners' processing of gestures. For example, gestures enhance listeners' comprehension of spatial information more compared to non-spatial information, especially when they expressed core semantic features such as size, shape, and the location of objects (Beattie & Shovelton, 1999a, 1999b, 2001; Holler et al., 2009; Hostetter, 2011). Listeners benefit more from seeing gestures when speech is relatively unintelligible such as in degraded vs. clear speech comprehension (e.g., Drijvers et al., 2017, 2019; Holle et al., 2010). Listeners processed gestures differently when gestures were perceived as communicatively intended (Holle & Gunter, 2007; Holler et al., 2014; Kelly, Ward, Creigh, & Bartolotti, 2007). For example, Holler and colleagues observed differential behavioral and neural outcomes of gesture processing for addressed (i.e., speakers' eye gaze directed towards them) and unaddressed (i.e., eye-gaze averted to another listener) participants in a triadic communication setting. These suggest that the relative information or communicative value of gestures modulates listeners' sensitivity to and processing gestures.

One important factor that affects the relative information value of gestures is the semantic overlap between speech and gesture across different speech contexts. Speakers use gestures to complement their spoken utterances, particularly for visual-spatial information such as when a speaker says, "The cake was this big" and makes a two-hand gesture to represent the size of the cake (Beattie & Shovelton, 2006; Emmorey & Casey, 2001; Holler & Stevens, 2007). Listeners might attend to gestures more as their relative information value in relation to the accompanying context increases. To date, there is no work directly comparing how listeners attend to, process, and benefit from observing redundant vs. non-redundant (i.e., complementary) gestures.

1.2. The Scope and outline of the dissertation

This dissertation investigates whether and how listeners' attention to and processing gestures exhibit variation across individuals with different cognitive skills, and situations in which gestures express complementary vs. redundant information in relation to the concurrent speech. To this end, **Chapter 2** reviews current evidence on how individual differences in different cognitive resources relate with (1) *gesture use*: how frequently speakers use gestures and how and to what extent speakers benefit from using gestures during speaking and thinking and (2) *gesture processing*: how and to what extent listeners attend to, process, and benefit from observing gestures during comprehension and learning across different ages. This review provides a detailed assessment of the field's current state by taking a life-span perspective (starting from childhood through adulthood to healthy aging) and bridging fields of gesture production and gesture processing. As evident in this review, there is little research on individual differences in gesture processing. Based on this, **Chapter 3** investigates the role of individual differences in visual-spatial and verbal abilities in the online processing of concurrent gestural vs. spoken information. Building on the work by Kelly et al. (2010), this chapter brings an individual-differences approach to multimodal communication and asks whether listeners differentially attend to speech vs. gestures depending on their propensities to use visual-spatial and verbal resources. **Chapter 4** focuses on an external factor that can modulate listeners' attention to gestures. This chapter investigates how the relative information value of gestures in relation to the accompanying speech affects listeners' attention to gestures through their overt gaze allocation. Eye-tracking provides a useful method to study how listeners' attention allocation across different channels of information is reflected in gaze behavior.

By employing eye-tracking methodology, this study directly compares how complementary vs. redundant gestures affect listeners' direct visual attention to gestures.

CHAPTER 2

GESTURE USE AND PROCESSING: A REVIEW ON INDIVIDUAL DIFFERENCES IN COGNITIVE RESOURCES

This chapter is based on Özer, D. & Göksun, T. (2020). Gesture use and processing: a review on individual differences in cognitive resources. *Frontiers in Psychology*, 11:573555. doi: 10.3389/fpsyg.2020.57355

Abstract

Speakers use spontaneous hand gestures as they speak and think. These gestures serve many functions for speakers who produce them as well as for listeners who observe them. To date, studies in the gesture literature mostly focused on group-comparisons or the external sources of variation to examine when people use, process, and benefit from using and observing gestures. However, there are also internal sources of variation in gesture use and processing. People differ in how frequently they use gestures, how salient their gestures are, for what purposes they produce gestures, and how much they benefit from using and seeing gestures during comprehension and learning depending on their cognitive dispositions. This review addresses how individual differences in different cognitive skills relate to how people employ gestures in production and comprehension across different ages (from infancy through adulthood to healthy aging) from a functionalist perspective. We conclude that speakers and listeners can use gestures as a compensation tool during communication and thinking that interacts with individuals' cognitive dispositions.

2.1. Introduction

Human language occurs in a face-to-face interactional setting with the exchange of multiple multimodal cues such as eye-gaze, lip movements, body posture, and hand gestures. In this review paper, we focus on one of these multimodal cues: iconic hand gestures (henceforth, gestures) that represent objects, events, and actions. Speakers use an abundant number of gestures as they speak or think. These gestures serve many functions for speakers who produce them and for listeners who observe them (Dargue, Sweller, & Jones, 2019; Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Kita, Alibali, & Chu, 2017; McNeill, 2005; Novack & Goldin-Meadow, 2017; Özyürek, 2014). Although gesture and speech express meaning in a coordinated and integrated manner, gesturing is not mandatory for communication and hence, shows variation across situations and individuals (Kendon, 2004; Kita & Özyürek, 2003; McNeill, 2005; Streeck, 2009). Speakers differ in how frequently they use gestures, how salient their gestures are, and how much they benefit from *using* gestures during encoding and learning. On the other hand, listeners also differ in how much they attend to the speaker's gestures and benefit from *observing* gestures during comprehension and learning. The current paper will discuss individual differences in gesture use and processing.

There is individual variation in all human traits. People exhibit individual differences in cognitive abilities such as working memory capacity, attention, speech production, and processing as well as language acquisition (e.g., Bates, Dale, & Thal, 1995; Broadway & Engle, 2011; Daneman & Green, 1986; Huettig & Janse, 2015; Just & Carpenter, 1992; Kane & Engle, 2002; Kidd, Donnelly, & Christiansen, 2018). Current theories in cognitive science have not fully accounted for the existence as well as the causes of these individual differences for scientific gain (Underwood, 1975; Vogel & Awh, 2008). Most of the earlier studies in the gesture literature disregarded the variation among individuals and focused on group comparisons based on age (e.g., Austin & Sweller, 2014; Colletta, Pellenq & Guidetti, 2010; Feyereisen & Havard, 1999; Özer et al., 2017), sex (e.g., Özçalışkan & Goldin-Meadow, 2010), neuropsychological impairments (e.g., Akhavan, Göksun, & Nozari, 2018; Akbıyık, Karaduman, Göksun, & Chatterjee, 2018; Cleary, Poliakoff, Galpin, Dick, & Holler, 2011; Göksun, Malykhina, Lehet, & Chatterjee, 2013, 2015; Hilverman, Cook, Duff, 2018; Özer, Göksun, & Chatterjee, 2019; see Clough & Duff, 2020 for a review), culture, and the native status of the speakers and the listeners (i.e., bilinguals vs. monolinguals, e.g., Azar, Backus, & Özyürek, 2019, 2020; Goldin-Meadow & Saltzman, 2000; Gullberg, 2010; Kim & Lausberg, 2018; Kita, 2009; Mayberry & Nicoladis, 2000; Nicoladis, Pika, Marentette, 2009; Pika, Nicoladis, & Marentette, 2006; Smithson, Nicoladis, & Marentette, 2011) to understand how human multimodal language faculty operates at a general level. The gesture theories and current experimental practices in the gesture literature mostly downplayed the significance of individual differences and treated them as error variance. These studies create an illusionary and incorrect assumption that gesturing and the cognitive and communicative benefits of using and seeing gestures are invariant across people. However, using and observing gestures show not only across-group but also within-group variation (e.g., Chu, Meyer, Foulkes, & Kita, 2014; Dargue et al., 2019; Hostetter & Alibali, 2007, 2011; Özer et al., 2019; Özer & Göksun, 2020; Wu & Coulson, 2014a, 2014b). What drives this variation?

There are external and internal sources of variation in gesture use and processing. The external sources of variation could be speech content (e.g., spatial vs. non-spatial topics; Alibali, 2005; Feyereisen & Havard, 1999; Hostetter, 2011; Lausberg & Kita, 2003; Rauscher, Krauss, & Chen, 1996), communicative context (e.g., the visibility between interlocutors, communicative intention, and audience design; Alibali, Heath, & Myers, 2001; Schubotz, Holler, & Özyürek, 2019; Trujillo, Simanova, Bekkering, & Özyürek, 2018), task difficulty and cognitive load levels (e.g., complex spatial tasks such as mental rotation; Kita & Davies, 2009; Wesp, Hesse, Keutmann, & Wheaton, 2001). There are also internal sources variation; even under the same external circumstances, people can behave differently. Insights into which mechanisms contribute to these individual differences just started to emerge (e.g., Aldugom, Fenn, & Wagner-Cook, 2020; Chu et al., 2014; Dargue et al., 2019; Hostetter & Alibali, 2007, 2011; Kartalkanat & Göksun, 2020; Özer & Göksun, 2020; Wu & Coulson, 2014a, 2014b).

Individual differences in personality characteristics, age, cognitive, and perceptual skills contribute to variation among individuals in terms of gesture use and processing (e.g., Austin & Sweller, 2017; 2018; Chu et al., 2014; Cohen & Borsoi, 1996; Dargue & Sweller, 2020; Eielts et al., 2018; Galati, Weisberg, Newcombe, & Avraamides, 2018; Gillespie, James, Federmeier, & Watson, 2014; Göksun, Goldin-Meadow, Newcombe, & Shipley, 2013; Hostetter & Alibali, 2007, 2011; Hostetter & Potthoff, 2012; Kartalkanat & Göksun, 2020; Marstaller & Burianova, 2013; Özer & Göksun, 2020; Pouw, Mavilidi, van Gog, & Paas, 2016; Vanetti & Allen, 1988; Wartenburger et al., 2010; Wu & Coulson, 2014a, 2014b). However, most of the research on individual differences focused on gesture production, particularly on the cognitive correlates of variation in spontaneous gesture use and how much people benefit from using gestures during problem-solving and encoding of information. Research on individual variation in how listeners attend to speakers' gestures and benefit from observing gestures for comprehension and learning is limited (Aldugom et al., 2020; Özer & Göksun, 2020; Wu & Coulson, 2014a, 2014b).

In the current review paper, we discuss individual differences in (1) *gesture use:* how frequently speakers use gestures during spontaneous speech and how much they benefit from using gestures during task solving and learning, and (2) *gesture processing:* how listeners attend to and process speakers' gestures and how much they benefit from observing speakers' gestures for online comprehension or subsequent learning. We specifically focus on individual differences in cognitive and perceptual abilities (see Hostetter & Potthoff, 2012 for personality characteristics). This review has three highlights: (1) We attempt to bring a complete picture of individual differences in gesture by bridging production (i.e., using gestures) and comprehension (e.g., seeing gestures) fields, (2) We adopt a *functionalist* approach to discuss possible cognitive correlates of gesture use and processing. *Functionalist gesture theories* (as opposed to *mechanistic approaches* such as Hostetter & Alibali, 2008; McNeill, 1992, 2005) discuss *why* speakers use gestures and what *functions* gestures serve for speakers and listeners during communication and thinking (e.g., Cook & Fenn, 2017; Kita & Özyürek, 2013; Kita et al., 2017; Novack & Goldin-Meadow, 2017; Pouw, de Nooijer, van Gog, Zwaan, & Paas, 2014). Theories asserting for what purposes speakers and listeners employ gestures might inform us about the possible cognitive correlates of individual differences in gesture use and processing. (3) We also take on a life-span developmental perspective, which covers how gesture use and processing differ in changing cognitive skills throughout the developmental trajectory (from childhood through adulthood to healthy aging).

The literature on how different populations across ages use and process gestures during communication and learning is quite rich. It is noteworthy that the current paper is not a comprehensive review of the general literature. Instead, we specifically focus on studies investigating individual differences in these processes. We first review the functions of gestures during communication and learning (Section 2.2). Then, we address evidence on individual differences in gesture use (Section 2.3) and gesture processing (Section 2.4) for children, young adults, and elderly adults. Last, we conclude the current state of the field and discuss areas that are open to further investigation (Section 2.5).

2.2. Functions of gestures during communication and learning

Several theories suggest how and why gestures occur during communication and thinking. Mechanistic theories mostly propose *how* gestures arise during communication and thinking (e.g., Hostetter & Alibali, 2008, 2018; McNeill, 1992; 2005). Functionalist theories, on the other hand, try to explain *why* we use gestures and the functions that gestures serve during communication and thinking, both for the speaker and the listener (e.g., Cook & Fenn, 2017; Goldin-Meadow et al., 2001; Kita & Özyürek, 2013; Kita et al., 2017; Novack & Goldin-Meadow, 2017; Pouw et al., 2014). The approach in this review will be from functionalist perspectives as they can give insight into which mechanisms might contribute to individual differences in gesture use and processing.

Gestures have several functions during communication and thinking. First, gestures affect communication between interlocutors. Speakers and listeners employ gestures for communicative purposes. Speakers produce gestures to communicate information, and listeners, in turn, benefit from these gestures to comprehend the to-be-communicated message (e.g., Alibali et al., 2001; Beattie & Shovelton, 1999; Goldin-Meadow & Alibali, 2012; Holler & Stevens, 2007; Hostetter, 2011). Speakers use gestures as an alternative channel of expression. Hence, both speakers and listeners employ gestures more in communicative challenges stemming from cognitive dispositions such as when a speaker is linguistically non-competent (e.g., bilinguals talking in their non-native language, Gullberg, 2010; Smithson et al., 2001) or has hearing impairments (Obermeier, Dolk, & Gunter, 2012). The communicative function of gestures suggests that speakers and listeners with low communicative capacity (e.g., low linguistic proficiency, low semantic fluency, or the non-native status of the speaker and the listener) might employ and benefit from gestures more.

Second, gestures affect speakers' and listeners' cognitive processes. Gestures help activate, maintain, manipulate, and package visual, spatial, and motoric information for speaking and thinking (Kita et al., 2017). Gestures reduce cognitive load by keeping spatial-motoric information active in working memory (Cook, Yip, & Goldin-Meadow, 2012; Goldin-Meadow et al., 2001; Marstaller & Burianova, 2013; Morsella & Krauss, 2004; Ping & Goldin-Meadow, 2010; Wesp et al., 2001) and by projecting internal representations to an external space (e.g., Pouw et al., 2014). Producing gestures provides an external visual feedback that can be used to maintain or retrieve task-related visual-spatial information and hence, reduces the cognitive load. Considering this, we expect that people might use gestures as a compensatory tool to manage their cognitive load. For example, people with lower visual-spatial cognitive capacity (e.g., lower visual-spatial WM capacity, lower general spatial skills assessed by mental rotation, lower fluid intelligence assessed by Raven's Matrices) might use gestures more frequently to compensate for their limited resources when talking and thinking, especially about spatial information (e.g., Chu et al., 2014; Galati et al., 2018; Göksun et al., 2013; Trafton et al., 2006). In a similar vein, speakers' gestures provide a stable visual representation for observers (i.e., listeners) and help listeners during comprehension and learning. People with lower cognitive resources might be in a greater need for external aids, and thus benefit more from seeing gestures (e.g., de Nooijer, van Gog, Paas, Zwaan, 2013; Özer & Göksun, 2020; Wu & Coulson, 2014a).

Functional gesture theories assert that gestures help convey information during communication and manage cognitive load during speaking, thinking, and learning (e.g., Kita et al., 2017; Novack & Goldin-Meadow, 2017). This suggests that gesture use and processing are sensitive to the cognitive dispositions of the speakers and the listeners. People might convey gestures to manage and compensate for their limited cognitive resources.

Mechanistic gesture theories, on the other hand, emphasize how people employ gestures. As opposed to functionalist theories, one of the first and most influential mechanistic accounts of gesture production (*The Growth Point Theory*, McNeill, 1992, 2005; McNeill & Duncan, 2000) posits that gestures do not compensate for thinking and speaking. According to this account, gesture and speech originate from a single representational system where an utterance contains both linguistic and imagistic structures that cannot be separated. Speech stems from propositional linguistic representations and gestures stem from non-propositional imagistic representations and reflect visual, spatial, and motoric thinking (McNeill, 1992, 2005; Krauss, Chen, & Gottesmann, 2000). This account suggests that gestures are the manifestations of the imagistic component of the thought. Although mechanistic accounts would not be against the role gestures play for people to manage cognitive processes, they emphasize *how* people employ gestures rather than *why* they gesture.

In the following sections, we review evidence regarding how individual differences in cognitive domains relate with gesture use and processing from a functionalist account, mainly considering the *gesture-as-a-compensation-tool* view. That is, following the functionalist approach, we will illustrate the functions of gestures for speakers and listeners who use their cognitive resources differently. Gestures might not be used as a compensatory tool for every situation across different groups (e.g., Chui, 2011; de Ruiter, Bangerter, & Dings, 2012; So, Kita, & Goldin-Meadow, 2010); yet the current state of the field supports the beneficial part of gestures for communication, thinking, and learning (e.g., Goldin-Meadow et al., 2001; Kita et al., 2017; Novack & Goldin-Meadow, 2017).

2.3. Individual differences in gesture production

People from all ages show variation in terms of how frequently they use gestures, how salient their gestures are, and what types of gestures they use during spontaneous speech (e.g., Arslan & Göksun, in press; Chu et al., 2014; Feyereisen & Havard, 1999; Nagels, Kircher, Steines, Grosvald, & Straube, 2015; Priesters & Mittelberg, 2013; Schmalenbach, Billino, Kircher, van Kemenade, & Straube, 2017; Richmond, McCroskey, & Johnson, 2003). People also differ in how much they benefit from using gestures during speaking, encoding, and subsequent learning (e.g., Galati et al., 2018; Goldin-Meadow et al., 2001; Ping & Goldin-Meadow, 2010). To date, studies mostly focused on two possible cognitive correlates: visual-spatial vs. verbal cognitive resources. We discuss how individual differences in visual-spatial and verbal cognitive capacities relate to gesture production in children, adults, and elderly adults.

2.3.1. Individual differences in gesture production in children

Babies start to use pointing gestures at around 12 months of age and iconic gestures at around three years of age (Özçalışkan & Goldin-Meadow, 2005, 2010; Iverson, Capirci, & Caselli, 1994). Gestures open the way for the transition from prelinguistic to linguistic period, and gestures become increasingly intertwined with speech as children become older (e.g., Capirci, Contaldo, Caselli, & Volterra, 2005; Capirci & Volterra, 2008; Liszkowski, Carpenter, & Tomasello, 2008). Özçalışkan and Goldin-Meadow (2005) analyzed children’s gestures at 14, 18, and 22 months of ages when children are interacting spontaneously with their mothers. They showed that children used more gestures as they got older. Moreover, there was a developmental shift towards the use of more supplementary gestures (e.g., saying “ride” and pointing at the bike) as opposed to reinforcing gestures (e.g., saying “bike” and pointing at the bike) by older children. Yet, there was no difference in the quality or the quantity of the maternal input across development, suggesting that changes in children's gestural behavior might reflect developmental changes in children's own cognitive processes. Then, individual differences in several cognitive processes might lead to variations in how and to what extent children use gestures in spontaneous speech. Children, even as early as 14 months of age, show individual variation in whether they use iconic gestures and how frequently they use them (e.g., Iverson, Capirci, Longobardi, & Caselli, 1999; Özçalışkan & Goldin-Meadow, 2005). What drives these very early individual differences in gesture use? To date, the gesturing behavior of young children mostly focused on how individual differences in early gesture use predicted later language development (e.g., Demir, Levine, & Goldin-Meadow, 2015; Rowe & Goldin-Meadow, 2009). Studies examining the precursors of these variations, on the other hand, primarily focused on how parental language input (speech and gesture) relates with children's spontaneous gesture production (e.g., Iverson, Capirci, Volterra, & Goldin-Meadow, 2008; Rowe, Özçalışkan, & Goldin-Meadow, 2008; Tamis-LeMonda, Song, Leavell, Kahana-Kalman, & Yoshikawa, 2012). It is unknown which cognitive and perceptual abilities of these young children drive early individual differences in gesture production. Early socio-cognitive precursors of gesture use in infancy is an open area for further investigation.

How do children use gestures at later ages, such as during preschool and school-age? Children have not yet fully developed verbal skills as compared to young adults; thus, they might use gestures more during speaking as gestures provide an alternative channel of expression (e.g., Melinger & Levelt, 2004) and help facilitate speaking (Krauss et al., 2000). Indeed, studies report that preschool-aged children benefit more from gestures than older children and adults, especially when using complex language (e.g., Austin & Sweller, 2014; Church, Kelly, & Lynch, 2000). Moreover, children in transitional stages (i.e., children who have the conceptual knowledge but not yet the skills to verbalize that knowledge) used more gestures to convey ideas compared to children who had necessary verbal resources to convey the same idea linguistically (e.g., Church & Goldin-Meadow, 1986; Perry, Church, & Goldin-Meadow, 1992). These gestures (so-called *gesture mismatches*) expressed non-redundant information that was not found in the accompanying speech. Children (ages 5-10) used more non-redundant speech-gesture combinations both at the clause and word levels compared to adults (Alibali, Evans, Hostetter, Ryan, & Mainela-Arnold, 2009). This is also evident in the expression of other linguistically challenging categories such as causal or spatial relations (e.g., Austin & Sweller, 2018; Calero, Shalom, Spelke, & Sigman, 2019; Göksun, Hirsh-Pasek, & Golinkoff, 2010; Karadöller et al., 2019). Children used more gestures to convey additional information when they could not verbalize instruments of causal events (Göksun et al., 2010) or spatial relations such as left-right (Karadöller et al., 2019). For example, ambiguous spatial terms such as "*here*" can be complemented by gestures to specify the spatial relation (Karadöller et al., 2019). The multimodal discourse continues to develop during the school-age years. There is a developmental shift towards the use of a higher number of gestures per clause by 10-year-old children and adults than 6-year-olds in narrative production tasks (e.g., Alamillo, Colletta, & Guidetti, 2013; Colletta et al., 2010).

Developmental studies suggest that children might use gestures as an alternative channel of expression to compensate for their limited linguistic proficiency (e.g., younger vs. older children or children vs. adults; Alibali et al., 2009; Church et al., 2000; Colletta et al., 2010). This is in line with bilingualism research showing that bilingual children speaking in their L2 used more gestures than monolinguals (e.g., Smithson et al., 2001; Wermelinger, Gampe, Helbling, & Daum, 2020). Moreover, research on clinical populations with communication and language delays suggests that although there are delays in gesture production in the first 2 years, gesture might be used to compensate for communication and language difficulties at preschool and school ages by some children (LeBarton & Iverson, 2017; Özçalışkan, Levine, & Goldin-Meadow, 2013). Children with language impairments (LI) used gestures at a higher rate and produced greater proportions of gestures that added unique information to the accompanying speech compared to typically developing (TD) peers, suggesting that children with LI employ gestures as an alternative channel of expression in the face of language difficulties (Blake, Myszczysyn, Jokel, & Bebiroglu, 2008; Evans, Alibali, & McNeill, 2001; Iverson & Braddock, 2011; Mainela-Arnold, Alibali, Ryan, & Evans, 2011; Mainela-Arnold, Alibali, Hostetter, & Evans, 2014).

Similar to children with LI, children with Down Syndrome (DS) used more gesture-only expressions and expressed information uniquely in their gestures compared to TD children to compensate for spoken language delays (Dimitrova, Özçalışkan, & Adamson, 2016; Özçalışkan, Adamson, Dimitrova, & Baumann, 2017; Stefanini, Caselli, & Volterra, 2007). Children with Williams Syndrome (WS) also used more iconic gestures in a picture naming task compared to TD children to alleviate their word-finding difficulties (Bello, Capirci, & Volterra, 2003). Yet, not all children with language delays benefit from gestures as a compensatory tool. Children with Autism Spectrum Disorder (ASD) exhibit delays in gesture production that are apparent both in frequency and complexity (Colgan et al., 2006; Dimitrova et al., 2016; Özçalışkan, Adamson, & Dimitrova, 2016; Özçalışkan et al., 2017; Rozga et al., 2011; Watson, Crais, Baranek, Dykstra, & Wilson, 2013). Research shows that children with ASD used gestures to initiate and sustain joint attention and to compensate for speech limitations by supplementing speech to a lesser degree compared to TD peers, leading to negative consequences for learning and social interaction opportunities (Sowden, Clegg, & Perkins, 2013; Mastrogiuseppe, Capirci, Cuva, & Venuti, 2015; Watson et al., 2013). Impairments in gesture production are more pronounced in ASD compared to other developmental delays such as Down Syndrome (DS; Mastrogiuseppe et al., 2015), LI (Stone, Ousley, Yoder, Hogan, & Hepburn, 1997), and general intellectual delay (Mundy, Sigman, & Kasari, 1990) and thus, are a central component of problems in social interactions and delays in social development in ASD. Moreover, language delays not only affect children's gesture productions but also gestural input they receive from their caregivers, resulting in cascading consequences for language development.

Research suggests that children's language level affects caregivers' gestures to a greater extent when a child's language skills are limited (Dimitrova et al., 2016; Iverson, Longobardi, Spampinato, & Caselli, 2015; Özçalışkan et al., 2017; Özçalışkan, Adamson, Dimitrova, & Baumann, 2018; Talbott, Nelson, & Tager-Flusberg, 2015). For example, mothers of non-diagnosed high-risk ASD infants gestured more frequently compared to mothers of low-risk ASD infants (Talbott et al., 2015). The evidence on the compensatory use of gestures by children with language delays indicate that gesture is a tool that should be harnessed to support learning, especially for child clinical populations (LeBarton & Iverson, 2017). Gesture is also an early diagnostic tool to foresee persistent language delay, especially for children with unilateral brain lesions (Özçalışkan et al., 2013; Sauer, Levine, & Goldin-Meadow, 2010). Although these studies suggest a link between early spoken language abilities and gesture production in children, the direct evidence on how individual differences in early receptive and expressive language skills relate with spontaneous gesture use within children with and without language delays is quite limited (Kartalkanat & Göksun, 2020; Wermelinger et al., 2020).

A growing body of literature shows that using gestures benefit children's subsequent memory and learning (e.g., Alibali & di Russo, 1999; Wakefield, Hall, James, & Goldin-Meadow, 2018). Do all children benefit similarly from using gestures? Post and colleagues (2013) showed that children who simultaneously produced and observed gestures when learning grammatical rules performed worse than children who only observed gestures. However, the adverse effects of gesturing on learning were only visible for children with lower verbal skills, suggesting that producing and observing gestures simultaneously might be too cognitively demanding, especially for children with lower verbal resources (Kalyuga, 2007). Nevertheless, it should be noted that this study tested the effects of using gestures on learning under high cognitive load. There is no direct evidence on how verbal skills relate to how much children benefit from using gestures when they are under average cognitive load (e.g., without observing gestures simultaneously).

Developmental studies mostly compared different age groups (e.g., children vs. adults or younger vs. older children, e.g., Colletta et al., 2010), bilinguals vs. monolinguals (e.g., Mayberry & Nicoladis, 2000; Smithson et al., 2001) and clinical vs. non-clinical groups (e.g., Bello, Capirci, & Volterra, 2003; Dimitrova, Özçalışkan, & Adamson, 2016; LeBarton & Iverson, 2017). These studies suggest that children use gestures as a compensatory tool, and individual differences in verbal skills play a role in how much children use and benefit from using gestures during learning. Moreover, visual-spatial skills follow a protracted development, and children show individual variation in visual-spatial abilities (Newcombe, Uttal, & Sauter, 2013). Given that gestures are visual-spatial entities and help activate, maintain, and manipulate visual-spatial information (Kita et al., 2017), individual differences in visual-spatial abilities during childhood might affect how much children use gestures and benefit from using gestures for learning. However, there is no direct evidence on how individual differences in verbal and visual-spatial skills relate to children's gesture use, which begs for future research.

2.3.2. Individual differences in gesture production in young adults

Most of the research on individual differences in gesture production focused on young adults. Studies showed that young adults with lower cognitive capacities used more spontaneous gestures and benefited more from using gestures (e.g., Chu et al., 2014; Gillespie et al., 2014; but see Hostetter & Alibali, 2007), supporting the functionalist accounts (Goldin-Meadow et al., 2011; Kita et al., 2017; Marstaller & Burianova, 2013) and gesture's role as a compensation tool. The visual-spatial cognitive capacity is related to how much speakers employ gestures during speaking and thinking. People with lower visual and spatial working memory (WM) capacities, mental rotation skills, and spatial conceptualization abilities (Kita & Davies, 2009) used more gestures compared to high-spatial ability individuals when explaining abstract phrases or social dilemmas (Chu et al., 2014). In a spatial gesture elicitation task, Göksun and colleagues (2013) asked young adults to describe how they solved mental rotation problems and found that people with lower spatial abilities (lower mental rotation scores) used more gestures compared to people who had higher scores. However, low-and high-spatial ability individuals not only differed in the frequency of gestures but also in the type of gestures they used. People with low-spatial ability used more static gestures depicting objects (i.e., cubes or whole objects) whereas high-spatial ability individuals used more dynamic gestures to express motion, such as rotation or direction or static gestures referring to object pieces (e.g., the bottom part of the L shape). This finding is in line with a previous study showing that although lower- and higher-fluid intelligence individuals (as measured by Raven's matrices) used an equal number of gestures when describing how to solve geometric analogies, people with higher fluid intelligence used more gestures to express motion than people with lower fluid intelligence (Sassenberg, Foth, Wartenburger, & van der Meer, 2011; Wartenburger et al., 2010).

Verbal cognitive capacity is another predictor for how and to what extent speakers use gestures (e.g., Baxter, Winters, & Hammer, 1968; Gillespie et al., 2014; Hostetter, Alibali, 2007, 2011; Napgal, Nicoladis, & Marentette, 2011; Smithson & Nicoladis, 2013; for c.f. see Chu et al., 2014 and Frick-Horbury, 2002). Young adults with lower verbal abilities such as lower verbal WM capacity, vocabulary size, and semantic fluency (i.e., phonological and lexical retrieval ability) used more gestures during spontaneous speech than individuals with higher verbal abilities (e.g., Gillespie et al., 2014; Hosttetter & Alibali, 2007, 2011; Smithson & Nicoladis, 2013; but see Chu et al., 2014). These findings corroborate with bilingual research, showing that bilinguals used more gestures when talking in their L2 compared to L1 or monolinguals (e.g., Gullberg, 1998; Nagpal et al., 2011). Verbal WM also predicted gesture frequency similarly in bilinguals and monolinguals (Smithson & Nicoladis, 2013).

Is there an interaction between verbal and spatial skills in gesture use? Hostetter and Alibali (2007) showed a quadratic relationship between verbal resources and spontaneous gesture use. People with the lowest and highest verbal skills (i.e., phonemic fluency) gestured more than people with average verbal skills when they were retelling a cartoon story and describing how to wrap a package. Moreover, low verbal/high visual-spatial individuals produced the largest number of gestures and used more non-redundant gestures (Hostetter & Alibali, 2007, 2011; Vanetti & Allen, 1998). This might suggest that gestures are more helpful when speakers have spatial information in the non-propositional format in mind but are unable to lexicalize or to encode verbally (e.g., Graham & Heywood, 1975; Krauss & Hadar, 1999).

Young adults also show individual variation in how much they benefit from using gestures during task solving or subsequent memory and learning (e.g., Marstaller & Burianova, 2013). Young adults use many gestures when encoding information that facilitates their subsequent memory and learning, especially for visual and spatial information (e.g., Chu & Kita, 2011; So, Shum, & Wong, 2015). However, using gestures is especially beneficial for people with lower cognitive capacity (e.g., Galati et al., 2018; Marstaller & Burianova, 2013; Pouw et al., 2016). People who used gestures when trying to learn new routes had better memory in a subsequent navigation task; however, this was only evident for people with a lower spatial perspective-taking ability (Galati et al., 2018). Moreover, gesturing benefited problem solving under higher cognitive load (e.g., dual-task paradigm, Marstaller & Burianova, 2013) and when internal cognitive resources are taxed or limited (e.g., Pouw et al., 2016).

Individual differences in verbal and visual-spatial skills affect how much young adult speakers use and benefit from producing gestures during speaking and problem-solving. Conforming the *gesture-as-a-compensation-tool* account, speakers employ gestures to compensate for lower verbal and spatial cognitive resources. However, we should be cautious about the generalizability of these findings as to the use of different cognitive measures, and gesture elicitation tasks (e.g., spatial vs. non-spatial, abstract) might yield different results. Further research is needed to replicate these conclusions across contexts.

2.3.3. Individual differences in gesture production in healthy aging

Evidence on spontaneous gesture use in healthy aging is minimal. Most of the research compared young and elderly adults and showed that spontaneous gesture production and gesture imitation is impaired in aged populations (e.g., Cohen & Borsoi, 1996; Dimeck, Roy, & Hall, 1998; Feyereisen & Havard, 1999). Elderly adults used less representational gestures compared to young adults whereas overall gesture frequency or the use of non-representational gestures (e.g., beat or conduit gestures) was comparable across two groups (Arslan & Göksun, in press; Cohen & Borsoi, 1996; Feyereisen & Havard, 1999; Glosser, Wiley, & Barnoskir, 1998; for c.f. see Özer et al., 2017;

Schubotz, Holler, & Özyürek, 2019). This might be due to declining visual-spatial cognitive resources in aging. For example, mental imagery declines with aging (e.g., Andersen & Ni, 2008; Copeland & Radvansky, 2007; Dror & Kosslyn, 1994) and indeed, individual differences in mental imagery, but not spatial WM capacity was associated with how frequently young and elderly individuals used spontaneous gestures, particularly for a spatial address description task (Arslan & Göksun, in press). Elderly individuals were also impaired in designing their multimodal utterances for their addressees (i.e., audience design, Schubotz et al., 2019). When narrating comic cartoons, young adults used fewer gestures when they knew that their addressee also watched the comic cartoon compared to when their addressee did not see the cartoon. However, elderly adults used an equal number of gestures in both cases.

We might expect that declining visual-spatial skills in aging would lead to higher use of gestures by older adults than in younger ones. However, gestures might be used as a compensatory tool only to manage cognitive load when the person has the necessary/intact resources. Most of the studies comparing younger vs. older adults tested individuals who are older than 60 years of age (e.g., Cohen & Borsoi, 1996), and it is unknown whether visual-spatial skills are severely impaired in this age group. Less is also known on the decline in which cognitive abilities in healthy aging leads to age-related impairments in gesture production (but see Arslan & Göksun, in press). It is important to note that this area is open to investigation and future research should study the decline in which cognitive resources lead to impaired gesturing in aging, whether the effects of aging on gesturing is similar for everyone and, and which cognitive resources might play a protective role for the decline of gesture production. More research is also needed to examine whether elderly individuals benefit from using gestures as young adults and children do or producing gestures impose an extra cognitive burden to their already limited cognitive resources.

Moreover, we mainly focused on gesture use in healthy aging, yet, the line of research on how people with neurodegenerative disorders use gestures is informative as well (e.g., Akhavan et al., 2018; Cleary et al., 2011; Klooster, Cook, Uc, & Duff, 2015; Özer et al., 2019; Rousseaux, Renier, Anicet, Pasquier, Mackowiak-Cordoliani, 2012). People with different types of neurodegenerative disorders such as Alzheimer's disease, primary progressive aphasia, and Parkinson's disease are natural targets to study gesture in aged populations because the prevalence rates of these diseases consistently increase with age (e.g., Brayne et al., 2006; Jorm, Korten, & Henderson, 1987). For example, Klooster and colleagues (2015) showed that the beneficial effects of using and observing gestures on new learning in a Tower of Hanoi paradigm were absent in elderly patients with intact declarative memory, but impaired procedural memory as a consequence of Parkinson's disease. This suggests that the procedural memory system supports the ability of gestures to drive new learning. Thus, the decline in different memory systems in different neurodegenerative disorders that increase with age might lead to variation in how elderly adults benefit from gestures during learning. Future studies should test the cognitive correlates of impaired gesture use in different neuropsychological groups.

2.4. Individual differences in gesture processing

Listeners are sensitive to speakers' gestures and benefit from observing these gestures during online language comprehension, encoding, and subsequent memory and learning (Dargue et al., 2019; Hostetter, 2011; Kelly, Özyürek, & Maris, 2010; Holler, Shovelton, & Beattie, 2009). The facilitative effects of observing gestures are evidenced across children (e.g., Aussems & Kita, 2019; Austin & Sweller, 2014; 2017; Cook, Mitchell, & Goldin-Meadow, 2008; Dargue & Sweller, 2020; Holler, Kendrick, Levinson, 2018; Kartalkanat & Göksun, 2020; Macoun & Sweller, 2016; Vogt & Kauschke, 2017) and young adults (e.g., Beattie & Shovelton, 1999;

Dargue & Sweller, 2019; Holle & Gunter, 2007; Hostetter, 2011; Kelly, Manning, & Rodak, 2008; Roth, 2001; Rueckert, Church, Avila, & Trejo, 2017). Research regarding individual differences in how listeners attend to and process speakers' gestures and how much they benefit from observing gestures during comprehension and learning is quite limited, especially when compared to the literature on individual differences in gesture production (e.g., Özer & Göksun, 2020; Post, van Gog, Paas, & Zwaan, 2013; Yeo & Tzeng, 2019; Wu & Coulson, 2014a, 2014b). In the next subsections, we review evidence on individual differences in gesture processing and its effects on comprehension and learning in children, young adults, and elderly adults. Then, we discuss several possible cognitive mechanisms that might yield individual differences in gesture processing, suggesting new venues for future research.

2.4.1. Individual differences in gesture processing in children

Electrophysiological studies showed that children start to process iconic gestures as semantic entities like words at around 18 months of age (Sheehan, Namy, & Mills, 2007). Behaviorally, they start to comprehend iconic gestures representing entities at around three years of age (Stanfield, Williamson, & Özçalışkan, 2014) and iconic gestures representing events at around four years of age (Glasser, Williamson, & Özçalışkan, 2018). Studies showed that 3-year-olds could not integrate speech and gesture, whereas 5-year-olds and adults did (e.g., Sekine & Kita, 2015; Sekine, Sowden, & Kita, 2018). Moreover, children starting from 6 years of age integrate speech and gesture in an online fashion comparable to adults (Dick, Goldin-Meadow, Solodkin, & Small, 2011; Sekine et al., in press). Demir-Lira and colleagues (2018) showed that gesture-speech integration recruits the same neural network as in adults. Yet, this was true only for children who were able to successfully integrate speech and gesture behaviorally. Then, what drives these individual differences in early gesture-speech integration ability?

Gesture-speech integration requires a global developmental shift. The precursors of gesture comprehension and gesture-speech integration are unknown. Gestures are visual-spatial entities and the processing, and the interpretation of gestures requires visual-spatial cognitive resources (e.g., Kelly & Goldsmith, 2004). Children with lower visual-spatial skills might have difficulty in processing and comprehending gestures compared to children with higher visual-spatial skills. The global development of executive attention and general WM capacity, on the other hand, might play a role in gesture-speech integration. For example, children with lower overall WM capacity might have difficulty in maintaining and integrating two different kinds of information simultaneously, especially in offline integration tasks (e.g., Demir-Lira et al., 2018). Cognitive predictors of individual differences in children's gesture comprehension and gesture-speech integration abilities require further attention.

What about individual differences in the beneficial effects of observing gestures for subsequent learning? Not all children benefit from visual aids such as diagrammatical illustrations when learning math (e.g., Cooper, Sidney, & Alibali, 2007). Indeed, observing gestures does not assist all children's comprehension of narratives or learning new skills (e.g., Bohn, Kordt, Braun, Call, & Tomasello, in press; Church, Ayman-Nolley, & Mahootian, 2004; Kartalkanat & Göksun, 2020; van Vermeskerken, Fijan, Eielts, & Pouw, 2016; Yeo & Tzeng, 2019). Kartalkanat and Göksun (2020) found a positive relationship with verbal skills and the beneficial effects of observing gestures, preschoolers with higher expressive language ability benefited more from observing iconic gestures in the encoding of spatial events. Bohn and colleagues (in press) found that children benefited from observing gestures when learning novel skills (e.g., how to open a novel apparatus) as they became older. On the other hand, Demir and colleagues (2014) showed that children with pre- and perinatal unilateral brain injury (BI) and had difficulty in narrative production benefited more from observing gestures when retelling narratives compared to TD children (Demir, Fisher, Goldin-Meadow, & Levine, 2014). Moreover, children with Specific Language Impairment (SLI) benefited more from observing gestures compared to TD and used the same gestures they observed when retelling the inferred meaning of the spoken messages (Kirk, Pine, & Ryder, 2011). The contradictory findings regarding the relation between verbal abilities and the beneficial effects of observing gestures between children with language impairments and children with intact language abilities pose a challenge. We might expect TD children with lower verbal abilities to benefit more from observing gestures as predicted by *gesture-as-a-compensation-tool* account; however, young children's limited verbal resources might be already consumed with processing speech, leaving few resources to process and benefit from external visual cues (i.e., gestures, Kalyuga, 2007). Again, gestures might help children manage cognitive load when they have fully developed verbal abilities. However, children with language impairments might employ gestures to compensate for their already-impaired spoken language abilities.

Individual differences in verbal (e.g., digit span task, Kartalkanat & Göksun, 2020), visual (e.g., visual patterns task, van Vermeskerken et al., 2016) or general WM capacity (e.g., operation span task, Yeo & Tzeng, 2019) did not predict how much children benefited from observing gestures for learning. There was hardly any variance in WM capacity in most of these studies (e.g., van Vermeskerken et al., 2016). This might obscure the otherwise possible effects of different WM capacities on the values of observing gestures in children. Additionally, how general spatial skills (e.g., mental rotation, mental imagery) relate to how much children benefit from observing gestures needs to be investigated in future research.

2.4.2. Individual differences in gesture processing in young adults

Young adults also differ in how they process spontaneous co-speech gestures. Processing gestures require visual, spatial, and motoric cognitive resources (e.g., Kelly & Goldsmith, 2004; Wu & Coulson, 2014a). We expect people with higher visual-spatial abilities to process and comprehend gestures better. Indeed, Wu and Coulson (2014a) found that people with higher spatial WM (but not verbal WM) were better at processing co-speech gestures as they were more sensitive to speech-gesture mismatches (i.e., high-spatial individuals were affected more negatively when gesture and speech expressed incongruent information). Moreover, people who have larger spans for retaining and manipulating bodily configurations (i.e., motor movement span task assessing individuals' ability to retain body-centric motor information) comprehended gestures better (Wu & Coulson, 2014b). In a recent study, we asked how visual-spatial vs. verbal WM capacity relates to processing concurrent visual (i.e., gesture) and verbal (i.e., speech) information in a mismatch paradigm used initially by Kelly and colleagues in 2011 (Özer & Göksun, 2020). We demonstrated that listeners showed differential sensitivity in processing concurrent gestural vs. spoken information. Although gesture-speech mismatches hindered overall comprehension, how listeners got affected by mismatches in different modalities (gesture vs. speech mismatches) was dependent on the listeners' cognitive dispositions on visual-spatial vs. verbal resources. Observing mismatching visual information (i.e., gesture) imposes an additional visual-spatial cognitive load and people with higher spatial abilities were better at maintaining and processing two different and mismatching visual information due to their higher capacity. As a result, these individuals perform better when gestures expressed mismatching information compared to people with lower spatial abilities. People with higher verbal abilities, on the other hand, had better performance when speech expressed mismatching information compared to people with lower verbal abilities. These findings suggest that visual-spatial cognitive resources are critical for gesture processing and observing mismatching gestures increase visual-spatial cognitive load (e.g., Hostetter, Murch, Rothschild, & Gillard, 2018; Kelly & Goldsmith, 2014). Processing mismatching information in visual modality would be less demanding for people with larger visual-spatial cognitive resources.

What about individual differences in how much listeners benefit from observing gestures? Earlier studies are limited in suggesting how listeners integrate visual information with speech and use gestures to encode information either for online language comprehension or for subsequent learning. Research on how learners benefit from different multimedia materials (visual vs. verbal representations) might give us insight in this matter (Ausburn & Ausburn, 1978; Kiat & Belli, 2018; Kirby, Moore, Shofield, 1988; Koć-Januchta, Höffler, Thoma, Prechtl, & Leutner, 2017; but see Kirchner, 2017). Individuals show variation in how they benefit from visual vs. verbal information (Alfred & Kraemer, 2017; Kirby et al., 1988; Kozhevnikov, Hegarty, & Mayer, 2002; Meneghetti, Labata, Grassano, Ronconi, & Pazzaglia, 2014; Mendelson & Thorson, 2004; Riding, Burton, Rees, & Sharratt, 1995). For example, learners show variation in how they fixated to text vs. pictures when learning from multimedia resources (Koć-Januchta et al., 2017) and students with higher spatial abilities benefited more from the presence of 3D models when learning cell biology compared to students with lower spatial abilities (Huk, 2006). This suggests that listeners' cognitive dispositions might be related to how much they benefit from observing gestures vs. hearing speech. A very recent study directly tested how different WM capacities related to how much young adults benefited from observing gestures (Aldugom, Fenn, & Wagner-Cook, 2020). Undergraduate students with higher visual WM capacity (i.e., visual patterns task) benefited more from observing gestures during math learning whereas verbal (i.e., sentence span task) and motoric (i.e., movement span task, Wu & Coulson, 2014b) WM capacities did not predict the beneficial effects of observing gestures (Aldugom et al., 2020). Although it is well established in the literature that gestures facilitate listeners' comprehension and learning (see Özyürek, 2014 for review), evidence suggests that this is not a monolithic process. It is also possible that observing gestures do not always facilitate comprehension and learning. For example, observing gestures hurt learning phonetic distinctions at the syllable level within a word for English-speaking adults learning vowel length contrasts in Japanese (Kelly, Hirata, Manansala, & Huang, 2014). However, as in the case with children (Kartalkanat & Göksun, 2020), learners' level of second-language proficiency might play a role for benefitting from gestures, which is another cognitive resource to be examined. Future studies should investigate the cognitive precursors of individual differences in the beneficial effects of observing gestures across different learning contexts (e.g., spatial vs. non-spatial) and different stages of language processing (e.g., phonological vs. semantic, Kelly et al., 2014).

2.4.3. Individual differences in gesture processing in healthy aging

Few studies examined how elderly individuals process gestures and benefit from observing gestures (e.g., Cocks, Morgan, & Kita, 2011, Montepare, Koff, Zaitchik, & Albert, 1999; Ska & Croisile, 1998; Thompson, 1995; Thompson & Guzman, 1999). Elderly individuals are impaired in their comprehension of pantomimes and emotional gestures compared to young individuals (Montepare et al., 1999; Ska & Croisile, 1998). Moreover, elderly adults are impaired in integrating speech and gesture compared to young adults (Cocks et al., 2011). However, they performed equally when two cues are presented in isolation, suggesting that they might be impaired in gesture-speech integration with a preserved ability to process gestures. Indeed, elderly adults mostly relied on visible speech and did not benefit from observing gestures when recalling sentences (Thompson, 1995).

Although young adults benefited from visual aids (i.e., visible speech and gestures) under challenging listening conditions (i.e., dichotic shadowing task), older adults did not (Thompson & Guzman, 1999). The differences in the effects of observing gestures between younger and older adults might be related to the declining cognitive abilities associated with aging, mainly due to WM capacity as WM is required to maintain and manipulate different information. However, this has not been addressed directly.

Previous research suggests that elderly adults have difficulty in integrating visual (i.e., gesture) and verbal (i.e., speech) information compared to younger adults. It might be due to a decline in global cognitive skills such as executive attention and general WM capacity. Yet, it has not been directly tested. Future studies should compare younger vs. older adults with several cognitive measures to understand the cognitive architecture behind impaired gesture processing and gesture-speech integration in healthy aging.

2.5. Conclusion and future directions

Speakers use gestures as they speak and think, and listeners, in turn, are sensitive to speakers' gestures. Gestures (both using by speakers and observing by listeners) have beneficial effects on language comprehension, problem-solving, encoding, and subsequent learning. Studies, to date, mostly focused on the role of different external factors (e.g., speech content, communicative context) on gestural behavior to answer *when* we use and benefit from gestures. However, it is also essential to ask *who* uses and benefits from gestures for *which* purposes. Research on the cognitive precursors of these individual differences in gesture use and processing has just started to emerge. Examining individual differences in gesture use and processing will help us uncover the cognitive architecture behind these processes and inform gesture research that is based on group data. The accounts that explain how and why gestures are employed should integrate individual differences research to have a full picture of when, why, and for whom gestures exhibit their supposed roles. This line of research is also informative for the development of educational programs incorporating the use of gestures by learners or teachers. The instructional programs should be tailored according to the cognitive dispositions and needs of the learners for optimal learning outcomes.

Most of the research on individual differences in the gesture literature examined gesture production in young adults. Studies on gesture use in children and elderly adults focused on group comparisons (i.e., comparing children at different ages, children vs. adults, younger vs. older adults, clinical vs. non-clinical groups). Moreover, individual differences in gesture processing are limited compared to the production literature. In the current review paper, we (1) combined two lines of research: using gestures and observing gestures, and (2) discussed the possible cognitive precursors of gesture use and processing in different age groups. We also highlighted the functions of producing and seeing gestures regarding their compensatory roles in speaking and thinking.

Gestures provide an alternative expression channel and assist speakers and listeners communicate (e.g., Alibali et al., 2001; Hostetter, 2011). Gestures also decrease speakers' and listeners' cognitive load by aiding them to activate, maintain, and manipulate visual-spatial information (e.g., Kita et al., 2007; Novack & Goldin-Meadow, 2017). Gestures help people manage cognitive load and are used as a compensatory tool. Listeners' and speakers' cognitive dispositions interact with this compensatory role of gestures, leading to individual differences in how much people benefit from using and seeing gestures for speaking, comprehension, task solving, and learning. As the *gesture-as-a-compensation-tool* account would argue, children and adults with lower cognitive resources use and benefit from using gestures more to manage cognitive load compared to people with higher cognitive resources (e.g., Austin & Sweller, 2014; Chu et al., 2014; Church et al., 2000; Galati et al., 2018; Gillespie et al., 2014; Göksun et al., 2010;

Marstaller & Burianova, 2013). However, we suggest that gestures do not replace the impaired cognitive abilities; instead, gestures help manage cognitive load when cognitive resources are intact. Gestures might not compensate for already-impaired cognitive abilities. For example, people with aphasia use more gestures to compensate for impaired speech, but only when they have the intact conceptual knowledge of what they express (e.g., Göksun et al., 2013, 2015). In a similar vein, there is a decrease in gesture production in healthy aging that might be due to impaired visual-spatial abilities such as mental imagery (e.g., Arslan & Göksun, in press; Cohen & Borsoi, 1996). There is also evidence of individual differences in gesture processing. Processing and comprehending gestures require visual-spatial cognitive resources (e.g., Kelly & Goldsmith, 2004; Hostetter et al., 2018). People with higher visual-spatial skills (or older children compared to younger children) process gestures better compared to people with lower visual-spatial skills (e.g., Özer & Göksun, 2020; Wu & Coulson, 2014a, 2014b). In line with the *gesture-as-a-compensation-tool* account, we expect that people with lower cognitive resources (especially visual-spatial) would benefit more from observing external visual cues (i.e., gestures, but see Aldugom et al., 2020). However, research on how visual-spatial abilities relate to how much listeners benefit from observing gestures is inconclusive and begs for further investigation.

Although group-comparison studies are informative, future work should address within-group variation more, especially in children and in elderly adults. How different cognitive skills are associated with gesture production and processing should be tested directly across different conditions. The employment of different cognitive measures, gesture elicitation tasks, and learning contexts might yield different results, and these should be incorporated to have a full picture of whom for and when gestures are helpful. For example, the relationship between visual-spatial abilities and how frequently speakers use gestures and how much they benefit from using and observing gestures depend on the content of the information to be communicated or learned (Arslan & Göksun, in press; Chu et al., 2008; Hostetter & Alibali, 2007; Lausberg & Kita, 2003). The role of visual-spatial abilities in gesture use and processing might be more pronounced in spatial speech compared to non-spatial speech (e.g., Alibali, 2005; Arslan & Göksun, in press). Future work should also investigate how internal sources of variation (e.g., individual differences in several abilities) interact with external sources of variation (e.g., speech content, task difficulty).

One area open for future investigation is the cognitive predictors of gesture processing; that is, how listeners attend, process, and benefit from observing gestures. Studies on how visual, verbal, and motoric WM capacities are linked to individuals' processing of concurrent gesture vs. speech employed mismatch paradigms (Özer & Göksun, 2020; Wu & Coulson, 2014a, 2014b). However, gesture mismatches are rare in natural communication, and we should investigate how different cognitive abilities relate to gesture processing in more ecologically valid paradigms. It is also unknown whether there are any individual differences in visual attention to gestures. Gestures are visual articulators and subject to visual processing. Although earlier research found that gestures can be processed peripherally and do not require direct visual attention (e.g., Gullberg & Holmqvitz, 1999, 2006; Gullberg & Kita, 2009), recent evidence suggests that several factors might modulate how listeners allocate overt visual attention to gestures such as the comprehensibility of speech, and the native/non-native status of the listener (e.g., Drijvers, Vaitonyte, & Özyürek, 2019). Future studies should address whether people with different visual-spatial vs. verbal abilities show differential overt visual attention to gestures and how this relates to individual differences in gesture processing (Wakefield, Novack, Congdon, Franconeri, & Goldin-Meadow, 2018). Above attending to and processing gestures, very little is known on whether and how individuals benefit from observing gestures during online language comprehension and learning across different learning contexts (Aldugom et al., 2020). We currently investigate how visual-spatial skills relate to how much listeners benefit from observing gestures when comprehending spatial relations between objects.

All studies reviewed above tested individual differences behaviorally. Electrophysiological and neuroimaging studies investigate the neural architecture of gesture use and processing (e.g., Kelly, Kravitz, & Hopkins, 2004; Willems, Özyürek, & Hagoort, 2009; Wu & Coulson, 2005). We might observe individual differences in neural data, that is otherwise non-observable behaviorally (e.g., Demir-Lira et al., 2018). Future work should examine individual differences in the recruitment of different neural networks when using and observing gestures and how these differences in neural data relate to behavioral performance after considering individuals' cognitive skills.

The current review only focused on how individual differences in cognitive skills (mostly verbal and visual-spatial skills) relate to gesture use and processing. However, individual differences in other domains might also affect how people employ gestures. Individual differences in other domains such as personality (Hostetter & Potthoff, 2012) and other aspects of cognitive and perceptual skills such as selective attention, auditory processing, and the speed of multisensory processing should be tested (e.g., Schmalenbach et al., 2017). Moreover, it is also important to study the relation between gesture production and processing. How individual differences in spontaneous gesture use predict how people attend to and benefit from observing gestures or vice versa are unknown (Wakefield, James, & James, 2013). Gesture processing might be affected by to what extent people themselves use gestures, and future studies should address the production-perception cycle and the mechanisms behind it.

In sum, gesture use and processing are not monolithic processes and show individual variation. Speakers and listeners can use gestures as a compensation tool during communication and thinking that interact with individuals' cognitive dispositions.

CHAPTER 3

VISUAL-SPATIAL AND VERBAL ABILITIES DIFFERENTIALLY AFFECT PROCESSING OF GESTURAL VS. SPOKEN EXPRESSIONS

This chapter is based on Özer, D. & Göksun, T. (2020). Visual-spatial and verbal abilities differentially affect processing of gestural vs. spoken expressions. *Language, Cognition and Neuroscience, 35(7),* 896-914. https://doi.org/10.1080/23273798.2019.1703016

Abstract

Listeners are sensitive to speakers' co-speech iconic gestures. Concurrent visual and verbal information compete for attentional resources during multimodal comprehension. The current study examined the role of individual differences in visual-spatial vs. verbal abilities on individuals' differential sensitivity to gestural vs. spoken expressions. Turkish-speaking adults (*N*=83) were tested on their sensitivity to concurrent gesture vs. speech in an online task (Kelly et al., 2010) and were administered spatial and verbal working memory measures. Participants were slower and less accurate when gesture and speech were incongruent to one another compared to the baseline condition, in which they expressed congruent information. People with higher spatial working memory capacity were more efficient in processing gestures whereas people with higher verbal working memory capacity were more sensitive to spoken expressions. These suggest that not all people are equally sensitive to co-speech gestures and some people may benefit more from gestures during comprehension.

3.1. Introduction

Behavioral and neurophysiological studies have repeatedly demonstrated that inputs from different sensory modalities are not processed in isolation. Human perceptual systems integrate information from different modalities to create a coherent and a unified representation of the external world (Beauchamp, Lee, Argall, & Martin, 2004; Calvert, 2001; Calvert & Thesen, 2004; Calvert, Spence, & Stein, 2004; Spence & Driver, 2004). Natural human face-to-face communication is multimodal by nature, consisting of many visual articulators such as visible speech (i.e., lip and tongue movements), body posture, eye-gaze, and hand gestures (Vigliocco, Perniss, & Vinson, 2014). During language comprehension, listeners continuously integrate these different sources of information into an unfolding representation of speech context (Hagoort, 2005, 2013; Hagoort, Baggio, & Willems, 2009; Willems, Özyürek, & Hagoort, 2007, 2009). The present study will focus on the processing of speech with concurrent iconic hand gestures, which represent meaning that is related to the semantic content of the speech segment they accompany (McNeill, 2005).

Although speech and gesture are semantically, temporally, and pragmatically related, they have different mediums of expression (McNeill, 1992, 2005). Gestures express information visually in a holistic manner, whereas speech conveys information auditorily in a linear fashion. Though gesture and speech constitute two different sources of information stemming from different sensory modalities (visual vs. auditory), they form an integrated system during both language production and language comprehension (Kelly, Özyürek, & Maris, 2010a; Kita, 2000; Kita & Özyürek, 2003; McNeill, 2005). Gestures are part of the language system, and indeed gesture and speech together constitute language (Clark, 1996; Enfield, 2009; McNeill, 2005). Many studies showed that the processing of speech and gesture recruit the same neural resources and listeners rapidly integrate these two sources of information during comprehension (Beattie &

Shovelton, 1999; Cassell, McNeill, & McCullough, 1999; Kelly, Kravitz, & Hopkins, 2004; Kelly et al., 2010a; Özyürek, Willems, Kita, & Hagoort, 2007; Skipper, Goldin-Meadow, Nusbaum, & Small, 2007, 2009; Wu & Coulson, 2005, 2007). Yet, little is known about how people allocate attention to these two different sources of information during comprehension given that concurrent visual and verbal communicative signals compete for attentional resources. Processing concurrent gestural and spoken information seems automatic; however, there might be individual differences in how listeners show differential attentional sensitivity and processing ease to gestural vs. verbal information. The current study addressed how individual differences in the propensity to use visual-spatial and verbal resources relate with listeners' sensitivity in processing gestural vs. verbal information during multimodal comprehension.

Gestures enhance listeners' comprehension and facilitate encoding, especially when the content includes visual, spatial, and motoric information (Alibali, 2005; Beattie & Shovelton, 1999; Holle & Gunter, 2007; So, Shum, & Wong, 2015). Yet, gestures can also hinder comprehension when they express incongruent information to that is expressed in speech (Kelly et al., 2010a). The paradigms in which listeners are presented with bimodal utterances containing incongruent speech and gestures are highly used across gesture-speech integration studies (e.g., Cassell et al., 1999; Kelly et al., 2004, 2010a; Kelly, Creigh, & Bartolotti, 2010b; Özyürek et al., 2007; Willems et al., 2007; Wu & Coulson, 2014a). In one of these studies, Kelly, Özyürek, and Maris (2010a) presented participants with short video clips in which different actions were carried out (i.e., action primes, e.g., someone chopping vegetables). Following the action prime, participants were presented with a target video in which an actor makes a gesture and utters a word and were asked to respond if any part of the target video (i.e., speech and/or gesture) was related with the preceding action prime. In the baseline condition, both speech (e.g., saying "chop") and gesture (e.g., gesturing "chop") were related to the preceding action prime, conveying congruent information in relation to one another. In other experimental conditions, only one source of information (i.e., either speech or gesture) was related to the action prime, and thus, gesture and speech were incongruent to one another. In the gesture-unrelated condition (GU), speech was related (e.g., saying "chop") while gesture was unrelated (e.g., gesturing "twist") to the preceding action prime. In the speech-unrelated condition (SU), gesture was related (e.g., gesturing "chop") while speech was unrelated (e.g., saying "twist") to the action prime. Participants made more errors and response latencies were longer when speech and gesture were incongruent to one another compared to the baseline condition. This suggests that speech and gesture interact with each other in a bidirectional way during comprehension. This effect was even visible when participants were instructed to focus only on speech and disregard the gestural information, suggesting that listeners are sensitive to gestural information that is semantically analyzed in an automatic fashion (Kelly et al., 2010b). In the current study, we employed the adapted version of Kelly et al.'s (2010a) paradigm. Although Kelly et al. (2010a) were interested in speech-gesture congruence (i.e., the semantic (in)congruence between the speech and the gesture within the target video), we were mainly interested in the semantic relationship between the action prime and the target speech/gesture. In all unrelated conditions, there was a local speech-gesture incongruence. The unrelated conditions also posed extra processing load compared to the baseline condition. Because in all unrelated conditions, participants were required to maintain and process a different (i.e., unrelated target speech or gesture) information along with the one expressed in the action prime. However, the modality in which the cognitive load was introduced differed between conditions. In the GU condition, participants need to keep and process two different visual information in working memory: the visual representation of the action prime and an unrelated gesture. This leads to an increased *visual processing load* in the GU condition. In the SU condition, on the other hand, participants needed to keep and process two different verbal information in working memory: the action verb they presumably labeled as they watched the action prime and an unrelated speech target. This leads to an increased *verbal processing load* in the SU condition. In the current study, we asked whether listeners show differential sensitivity to process information expressed in the visual (i.e., gesture) vs. auditory (i.e., speech) modalities.

Which cognitive resources are important for listeners to process gestural and verbal information? Gestures are visual articulators; they represent meaning in the visual modality. Many theories assert that visual-spatial resources are essential for gesture processing. Gestures stem from imagistic (i.e., non-propositional) mental representations and reflect spatial-motoric thinking (Hadar & Butterworth, 1997; Hostetter & Alibali, 2008, 2018; Kita, 2000; Kita & Özyürek, 2003; Krauss, Chen, & Gottesman, 2000; McNeill, 2005). Gesture production studies have suggested that visual, spatial, and motoric (i.e., imagistic) contents generate the use of more representational gestures such as iconic gestures that indicate actions, objects, locations compared to non-imagistic content (i.e., abstract topics; Alibali, Heath, & Myers, 2001; Beattie & Shovelton, 2002; Feyereisen & Havard, 1999; Hostetter & Hopkins, 2002; Hostetter & Sullivan, 2011; Lavergne & Kimura, 1987; Rauscher, Krauss, & Chen, 1996; Trafton et al., 2006). People with higher spatial and imagery skills produce more spontaneous co-speech iconic gestures than people with lower spatial and imagery skills (Hostetter & Alibali, 2007, 2011), and elderly adults use fewer gestures than younger adults due to declined mental imagery skills with normal aging (Cohen & Borsoi, 1996). Although the comprehension of gestures received less attention, studies suggest that the beneficial effects of gestures on comprehension and learning might differ depending on listeners' own spatial skills (Eielts et al., 2018; Galati, Weisberg, Newcombe, & Avraamides, 2015; Goldin-

Meadow, Nusbaum, Kelly, & Wagner, 2001; Marstaller & Burianova, 2013; Pouw, Mavidili, van Gog, & Paas, 2016).

Verbal resources, on the other hand, might play role in the processing of speech that represents meaning linearly in auditory modality. People with higher verbal skills (i.e., higher verbal fluency or verbal working memory capacity) might process verbal information (i.e., speech) better compared to visual ones (i.e., gestures). Verbal resources might also be important for processing gestures. Some accounts suggest that gestures depend on propositional (i.e., verbal) rather than visual-spatial resources (e.g., Wagner, Nusbaum, & Goldin-Meadow, 2004). Moreover, most iconic gestures occur in the context of concurrent speech, and unlike other representations such as pictures or goal-directed actions, gestures are sketchy and hard to interpret without the concurrent speech (Krauss, Morrel-Samuels, & Colosante, 1991). Yet, we find this quite unlikely for the target gestures in the current stimuli. The target gestures were not hard to interpret without the concurrent speech; rather, they were isomorphic to the actions depicted in the preceding action prime video except for the use of the instrument.

Research on individual differences in gesture comprehension and gesture-speech integration ability is limited compared to studies addressing individual differences in gesture production (e.g., Chu, Meyer, Foulkes, & Kita, 2014; Gillespie, James, Federmeier, & Watson, 2014; Göksun, Goldin-Meadow, Shipley, & Newcombe, 2013; Hostetter & Alibali, 2007, 2011). Neuropsychological evidence suggests that gesture-speech integration depends more on non-declarative memory systems since this ability is spared in patients with hippocampal amnesia (Hilverman, Clough, Duff, & Wagner-Cook, 2018). Additionally, studies on aging evidenced that elderly adults could process gesture and speech in isolation, but they exhibit deficits in integrating these two compared to younger adults (Cocks, Morgan, & Kita, 2011; Huyse, Leybaert, &

Berthommier, 2014; Thompson, 1995; Thompson & Guzman, 1999). Though not tested directly, these studies argue that elderly individuals' poor performance in multimodal integration might stem from declined working memory (WM) capacity as with normal aging (Andersen & Ni, 2008; Copeland & Radvansky, 2007). However, these studies do not provide direct evidence on cognitive resources mediating concurrent gesture and speech processing. They also did not distinguish visual-spatial and verbal working memory capacities, which can show differential decline patterns during aging. Wu and Coulson (2014a, 2014b) addressed the role of visual-spatial and verbal WM capacity on gesture perception more closely. In Wu and Coulson (2014a), participants were shown video clips of spontaneous discourse involving iconic gestures, and then were presented with two picture probes among which they chose the one that was relevant to the preceding discourse. For target picture probes, speech was always related with the following picture probe whereas, in some trials, gestures were not related with the picture probe, creating a local incongruence between speech and gesture (e.g., making a gesture of swinging a golf club paired with the utterance of "*manual adjustment lens*"). As found in the earlier studies (e.g., Kelly et al., 2004; Kelly et al., 2010a), participants were faster and more accurate when gesture and speech in the spontaneous discourse were congruent with one another compared to when they were incongruent. However, this congruency effect (i.e., the cost of incongruency, calculated as incongruent minus congruent, collapsed over target pictures) was positively correlated with visual-spatial WM (i.e., Corsi block span task), but was not associated with verbal WM (i.e., sentence span task). That is, the adverse effects of incongruent gestures were greater for people with higher visual-spatial WM capacity.

Even though Wu and Coulson (2014a) provided support for individual differences in processing speech and gesture relatedness, the study needs to be scrutinized closely. First, this study examined the role of visual-spatial and verbal resources in processing incongruent speech-gesture information (i.e., the semantic relationship between the speech and the gesture in the spontaneous discourse). Although the congruency between gesture and speech was manipulated, the relatedness between the speech and the picture probe was not. The actor's speech was always related to the picture probe for the target related pictures. This makes it hard to distinguish the differential sensitivity to unrelated spoken vs. gestural information, because there was not unrelated speech condition. The lack of evidence for the role of verbal WM resources might be due to lacking this manipulation. Second, participants were instructed to respond if picture probes were related only to the utterance, encouraging them to focus on speech and disregard gestures. Again, this makes it hard to delineate the attentional sensitivity to visual vs. verbal sources of information, since the task already prioritizes the processing of one over the other. Third, this paradigm is an off-line task, requiring participants to wait until a picture probe has appeared on the screen to respond. However, participants might process both types of information during inter-stimulus-interval, making it hard to distinguish the differential processing speed for gesture vs. speech. For this reason, the use of an online task that requires participants to respond during encoding phase (i.e., while they see gesture and hear speech) is more appropriate for the purpose of the current study.

3.2. The present study

The current study examined how verbal and visual-spatial cognitive resources relate with processing semantic information expressed concurrently in visual (i.e., gesture) vs. auditory (i.e., speech) modalities. We employed an online mismatch paradigm that induces cognitive load in visual vs. verbal domains at the encoding phase (Kelly et al., 2010a). As an index of verbal and visual-spatial cognitive resources, we employed WM tasks. WM capacity is critical in the brief maintenance, processing, and integration of information (Atkinson & Shiffrin, 1968; Baddeley &

Hitch, 1974; Baddeley, 2000). However, WM has a limited processing capacity; thus, individual differences in WM capacity index performance especially under interference and cognitive load (Baddeley, 1986, 1992; Conway, Cowan, & Buntig, 2001; Engle, 2002; Kane & Engle, 2003; Kane, Bleckley, Conway, & Engle, 2001). For example, the complex cognitive tasks such as multimedia learning environments require learners to process and integrate multimodal information and the multimodal encoding of information improves memory and learning (Clark & Paivio, 1991; Moreno & Mayer, 2002). Yet, according to the cognitive load theory (Paas, Renkl, & Sweller, 2003, 2004; Paas, van Gog, & Sweller, 2010; Sweller, 1988), the outcomes of multimodal encodings can sometimes be detrimental due to the limited processing capacity of WM (Chandler & Sweller, 1991; Sweller, 1988). For example, learners with low verbal abilities (i.e., low WM capacity) had reduced overall text comprehension when given explanatory notes to facilitate vocabulary learning (Yeung, Jin, & Sweller, 1998). How learners process and benefit from multimodal information depends on the cognitive load imposed on the learners and people with higher processing capacities (i.e., WM capacity) overcome cognitive load better (Engle, 2002; Paas et al., 2010; Plaas, Chun, Mayer, & Leutner, 2003). We used three WM tasks to measure spatial and verbal WM capacities: Corsi Block Span Task for visual WM, and Digit Span Task and Operation Span Task for verbal WM.

First, we aim to replicate Kelly et al. (2010) findings in a different language - Turkish. Previous research has suggested cross-cultural differences in the frequency of gesture use (Azar, Backus, & Özyürek, 2019; Kita, 2009). If this is the case, gesture comprehension, and more specifically attention to gestures, might show cross-cultural differences as well. Thus, the first motivation of the current study is to replicate the original findings of Kelly and colleagues in a different language and culture in which gesture processing and production might differ from

English culture. Second, we investigate how spatial and verbal cognitive resources (as measured by spatial and verbal WM tasks) relate to listeners' sensitivity in processing concurrent visual (i.e., gesture) vs. verbal (i.e., speech) information under cognitive load.

In line with the previous study (Kelly et al., 2010a), we predicted that error rates would be higher and response latencies would be longer for trials in which gesture and speech conveyed incongruent information in relation to one another (i.e., the GU and the SU conditions) compared to when they expressed congruent information (i.e., the baseline condition). For individual differences, we only focused on the relation between the action prime and the target speech/gesture and disregarded the congruency between the gesture and the speech within the target video. We only analyzed the incongruent trials in which gesture and speech were always incongruent to one another; yet the relatedness between the action prime and the target varied (either gesture or speech was unrelated to the action prime). The task required participants to retain an action information and compared this representation with the following target spoken and gestural information that were presented concurrently. Then, participants responded whether at least one source of information was relevant to the action prime (i.e., without responding exactly which channel was relevant). To this end, participants had to process both targets of information to respond. In all unrelated trials, participants had to maintain and process an incongruent (i.e., unrelated) information in relation to the preceding action prime conveyed through one of the two modalities. Thus, we expected a differentiation between the GU and the SU conditions depending on participants' visual-spatial and verbal skills because they posed extra processing load in different modalities. WM capacity indexes performance under interference and people with higher visual-spatial vs. verbal abilities overcome visual-spatial vs. verbal cognitive load better, respectively (Engle, 2002; Paas et al., 2010; Plaas et al., 2003). In line with this, we expected that people with higher visual-spatial abilities would be better at handling extra processing load in the visual domain. Then, if gestures stem from visual-spatial representations (e.g., Hadar & Butterworth, 1997; Kita, 2000; Krauss et al., 2000), people with higher visual-spatial abilities would process gestures better under extra visual processing load (i.e., when they needed to maintain two mismatching visual information in WM). More specifically, we expected that people with higher spatial WM capacity would have lower error rates and shorter response latencies compared to people with lower spatial WM capacity in visual cognitive load (i.e., in the GU condition). On the other hand, if verbal WM capacity was related with the processing of verbal information, people with higher verbal WM would be better at handling extra processing load in the verbal domain. We expected that people with higher verbal abilities would process speech better under extra verbal processing load (i.e., when they needed to maintain two mismatching verbal information in WM). More specifically, we expected that people with higher verbal WM capacity would have lower error rates and shorter response latencies compared to people with lower WM capacity in verbal cognitive load (i.e., in the SU condition).

3.3. Methods

3.3.1. Participants

We recruited 83 participants (50 females, M_{age}= 21.7 years, $M_{education}$= 15.3 years) from Koç University, Istanbul in return of either course credit or monetary compensation. Six participants were discarded: two participants reported using psychiatric medication at the time of testing, one participant was ambidextrous, one participant reported having difficulty in focusing the experimental task, and two participants were discarded due to experimenter error. The final sample consisted of 77 participants (46 females, M_{age}= 21.5 years, age range= 18-29, $M_{education}$= 15.8 years, education range= 12-22 years). All participants in the final sample were right-handed, had normal or corrected-to-normal vision, and were not taking any psychiatric or neurological medication at least six months prior to the testing. All participants gave informed consent before the experiment, which was approved by the Institutional Review Panel for Human Subjects of Koç University.

3.3.2. Stimuli & Tasks

3.3.2.1. Experimental task

The experimental task was adapted from the priming task used in Kelly and colleagues (2010a). The stimuli in this task consisted of 1 s videos of common daily actions with an instrument (i.e., action prime, e.g., cutting a piece of paper with scissors) followed by a 500 ms blank screen, and finished by 1 s videos of an actor making a gesture and uttering a word (i.e., target video). Out of 16 actions (i.e., chop, cut, dial, hammer, knock, saw, scrub, shake, slice, squeeze, stir, turn, twist, type, wipe, wring) that were used in Kelly et al. (2010a), only 10 of them were used in Turkish as explained in the following Pretest section. The auditory sound files were recorded with a native Turkish female, and they were merged with video clips in iMovie. The audio files were inserted 200 ms after the onset of the target video (i.e., gestures preceded the onset of words by 200 ms).

There were 9 conditions for each action prime, yielding 90 trials in total. Out of these 9 trials, 5 of them were experimental trials (10*5=50 in total) in which the information either in speech or gesture was related with the action prime. For the baseline condition, both the speech (e.g., saying "cut") and the gesture (e.g., gesturing "cut") were related with the action prime (e.g., cutting a paper). However, in the other 4 experimental conditions (i.e., incongruent conditions), only one piece of information (either speech or gesture) in the target video was related with the action prime; thus, speech and gesture created a local incongruence between them. The semantic incongruency between speech and gesture was either weak or strong. For example, for the gesture-unrelated condition (GU), the actor's speech was related with the action prime (e.g., saying *cut* to the prime video of cutting paper), but the gesture was either weakly (e.g., gesturing *chop*) or strongly (e.g., gesturing *turn*) unrelated with the action prime. For the speech-unrelated condition (SU), the actor's gesture was related with the action prime (e.g., gesturing *cut* to the priming video of cutting paper) but speech was either weakly (e.g., saying *chop*) or strongly (e.g., saying *turn*) unrelated with the action prime (see Figure 3.1). Please note that the "*unrelated"* in gesture- and speech-unrelated conditions denotes the relationship between the action prime and the gesture/speech in the target video, but not the relationship between the speech and the gesture within the target video. In all incongruent conditions, speech and gesture were always incongruent to one another whereas the relatedness of the speech and gesture in relation to the preceding action prime differed. The remaining 4 trials were fillers (4*10= 40 filler trials) in which neither the gesture nor the speech of the actor was related with the action prime.

Pretest. A pretest was conducted to decide on which action primes to use and the Turkish labels associated with each prime. Five participants watched all 16 action primes used in Kelly et al. (2010) and spontaneously produced 3 verbs referring to those actions. They were notified that these verbs should be comprised of one single word without any object name. Out of these responses, we chose the 3 most frequently given answers for each action prime and asked another 5 participants to choose the best verb defining the action prime out of these 3 options in a forced choice task. At the end of this pilot study, the action prime of "sawing" was excluded since there is no one-word verb referring to this action in Turkish. The action primes of "squeezing" and "wring" were also excluded because the same word in Turkish ("*sıkmak"*) was used for both actions. We also had to discard other 3 actions (i.e., knock, shake, and slice) because these three actions were paired with the actions discarded during the pretest in creating video pairs. Thus, we proceeded with 10 action primes in the present study. The full final list of 10 action primes and their Turkish labels can be found in Appendix.

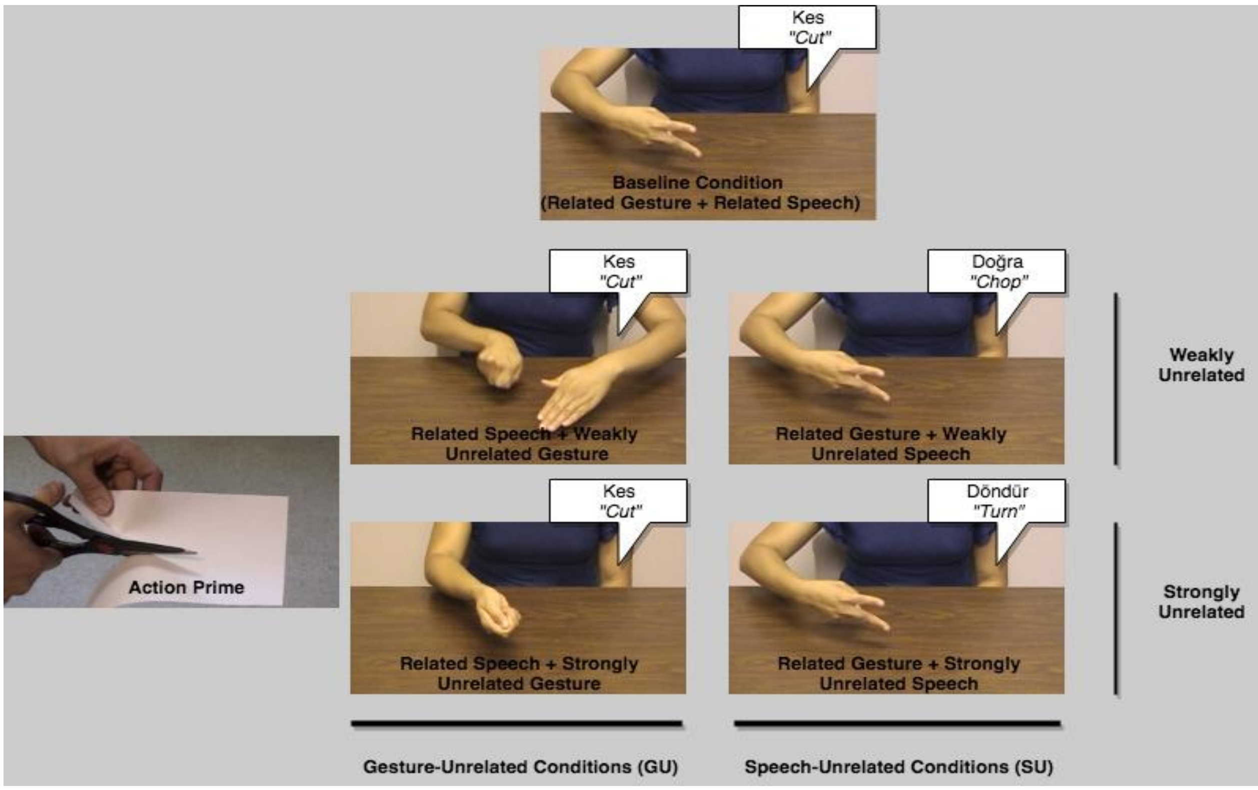


Figure 3.1. Adapted from Kelly, Maris, & Özyürek (2010). The figure illustrates 5 experimental conditions. Each trial begins with a 1 s video of an action prime, followed by a target video in which an actor makes a hand gesture and utters a word. In the baseline condition, both speech and gesture are related with the action prime, creating a local speech-gesture congruence within the target video. In the other 4 experimental conditions, only one piece of information is related with the action prime whereas the other source of information is either weakly or strongly unrelated with the action prime. In these conditions, gesture and speech conveyed incongruent information to one another. The words are written in Turkish with their English translation in italics.

3.3.2.2. Individual difference tasks

Corsi Block Span Task. The Corsi Block Span Task (adapted from Corsi, 1972) is a widely used test to assess visual-spatial working memory. In this task, an asymmetric array of 9 blue squares was presented on the screen, and in each trial, some of the squares flashed in a sequence. One block at a time turned yellow for 1 second, with an interblock time of 0.5 second. Participants waited till all squares flashed in sequence and were instructed to click each square in the same or reversed order that the flashes had occurred. Sequences started from 3 and proceeded to 9 flashes, advancing to the next level by successfully producing the flashes in each level. Each participant's block span would be the highest level in which she successfully produced all the flashes in the correct order.

Digit Span Task. The Digit Span Task, adapted from Wechsler (1939), measures storage capacity of verbal WM. In this task, participants saw a sequence of digits (e.g., 3, 7, 4, 2), presented one at a time at the rate of one digit per second in the center of the screen. Each digit appeared only once within a sequence. At the end of the sequence, participants had to recall them by selecting the digits from a circle either in the same or reversed order they were presented. The task began with 3 sequences of digits for forward Digit Span Task and 2 sequences for backward Digit Span Task, participants moved up or down a level depending on their performance. Participants' scores were the maximum number of digits recalled correctly, and the maximum number of digits recalled correctly before making two consecutive errors.

Operation Span Task. To assess the processing capacity for verbal WM, we employed the Operational Span Task (Turner & Engle, 1989). This task is predictive for performance in a wide range of cognitive tasks and traces individual differences in the processing capacity of verbal WM since it requires operations that are very similar to real-world activities (Conway et al., 2005; Engle, 2002; La Pointe & Engle, 1990). This task required participants to keep a series of letters in memory while solving simple arithmetic questions. One sequence consisted of a simple arithmetic question and one letter. For each sequence, participants were first presented with a simple arithmetic equation such as "(4*2)-3=?". They were instructed to click anywhere on the screen as soon as they finished mentally solving the problem. After the click, a number appeared on the screen, and participants judged whether the number was the correct solution of the arithmetic problem or not (e.g., press "Correct" if the number is "5"). After their response, a letter (e.g., K) appeared on the screen for 1 second. The first trial consisted of 2 sequences (i.e., arithmetic problem – letter – arithmetic problem – letter) and the number of sequences increased each trial, up to 7 sequences. At the end of each trial, participants were asked to select the letters in the correct order among an array of letters presented on the screen. We used the number of letters correctly recalled throughout the task as a measure for the processing capacity of verbal WM.

3.3.3. Procedure

Participants were tested in a dimly lit room where they were seated in front of a computer. They first completed the experimental task. In the experimental task, they were given a brief verbal instruction on how to perform the task. They were instructed to press "V" if they thought any part of the target video (speech, gesture, or both) was related to the action depicted in the first video (i.e., action prime) and to press "N" if they thought neither speech nor gesture in the target video was related to the action prime. It was emphasized that it is important to be accurate and fast at the same time while responding. They were also notified that they could even respond during the streaming of the target video. If they had not responded till the end of the target video, they would be notified that they should be faster in responding. Participants were presented with a fixation cross during inter-trial interval that lasted for 1 second. All participants completed 10 practice trials that were not included in the actual task. All trials were randomized, and participants did not get any feedback for their responses. The task was presented by using E-Prime 3 on a 13” Dell Laptop. There was no break during the experimental task, and it lasted approximately 15 minutes.

After the experimental task, participants completed WM tasks in a fixed order: forward and backward Digit Span Tasks, forward and backward Corsi Block Span Tasks, and the Operation Span Task. All WM tasks were presented in Inquisit 5 on a 21.5” iMac. Participants completed 2 practice trials before forward and backward Digit Span Tasks. There were no practice trials for the Corsi Span Tasks. For the Operation Span Task, participants completed 8 practice trials for remembering letters, 15 practice trials for solving arithmetic equations, and 4 practice trials for the task itself. The average responding time for solving arithmetic equations were calculated during the practice trial and in the actual task, the trials in which participants exceeded this average solving time for responding were skipped and counted as speed errors. The completion of all WM tasks lasted approximately 30 minutes. The entire session lasted about 50-55 minutes.

3.4. Analyses & Results

3.4.1. The experimental task – replication analyses

As stated earlier, the experimental task was adapted from Kelly and colleagues (2010a). One of the aims of the study was to replicate their findings in a different language (i.e., Turkish). To this end, we first present the exact analyses computed in the original article. We analyzed response time (RT, in milliseconds) and the proportion of errors (i.e., incorrect responses). Only experimental trials were included in the final analyses, making 50 trials per participant (5 experimental condition * 10 items in each) condition). In total, there were 3850 (50 trials * 77 participants) responses for accuracy and RT. Yet, for the RT measure, we discarded outliers that were two standard deviations of the mean (N=13 scores, .003%) and responses that were incorrect

(*N*=675, 17%), leaving 3162 trials for the RT measures for the final analyses. In line with Kelly et al.'s (2010) analyses, we tested if there is a bidirectional interaction between speech and gesture. In the first analyses, we asked if local speech-gesture incongruence within the target video hinders comprehension by conducting two separate repeated measures ANOVAs (accuracy and RT) to compare baseline condition with the average of all other four incongruent conditions. In the second analysis, we asked if there is a main effect of the modality– whether unrelated gestural vs. spoken information hindered comprehension more and conducted two separate repeated measures ANOVAs (accuracy and RT) with two factors: modality (speech- vs. gesture- unrelated trials) and strength (weakly vs. strongly unrelated).

Analysis 1. There was a statistically significant difference in reaction times (RT) between the baseline and all the incongruent conditions, F (1, 76) = 24.78, p< .001, η= .25. The RTs were longer in the incongruent conditions compared to the baseline (M_{diff}= 86.26 ms, see Figure 3.2). The error rates yielded the same pattern. There was a statistically significant difference for error rates between the incongruent and baseline conditions, F (1, 76) = 35.61, p< .001; η= .32. Participants made more errors in the incongruent conditions compared to the baseline condition (M_{diff}= 10.39%). Participants were slower and more error prone when speech and gesture conveyed incongruent information to one another compared to when they conveyed congruent information (i.e., baseline condition; see Figure 3.3).

Analysis 2. For reaction times, there was a main effect of the modality, F (1, 76) = 10.83, p< .01, η= .12. The RTs were longer in the GU trials compared to the SU trials, M_{diff}= 83.23 ms. There was not a main effect of the strength, F (1,76) = 1.52, p= .22, and there was no interaction between the modality and the strength, F (1, 76) = 2.40, p= .12 (see Figure 3.2). For the error rates, there was a main effect the modality, F (1, 76) = 6.36, p< .05, η= .08. Participants made more errors in the GU trials compared to the SU trials, M_{diff}= 4.54%. There was also a main effect of the strength, F (1, 76) = 9.21, p< .01, η= .11. Yet, this main effect was qualified by an interaction between the modality and the strength, F (1, 76) = 4.01, p< .05, η= .05. The strongly unrelated conditions yielded more error rates compared to the weakly unrelated conditions only in the SU trials, M_{diff}= 7.66% (see Figure 3.3).

In sum, we replicated the results of the Analysis 1. As in Kelly and colleagues, we found that participants were slower and less accurate in the incongruent conditions (averaged) compared to the baseline condition. Yet, the results of the Analysis 2 revealed a slightly different set of findings. Kelly et al. (2010a) found higher RTs for the SU compared to the GU condition. However, we found that participants were slower and less accurate when gestural information was unrelated to the preceding action prime (GU) compared to when speech conveyed unrelated information (SU). Yet, it is noteworthy that the current study achieved the parallel pattern for the RT and the error rate measures (see Figures 3.2 and 3.3).

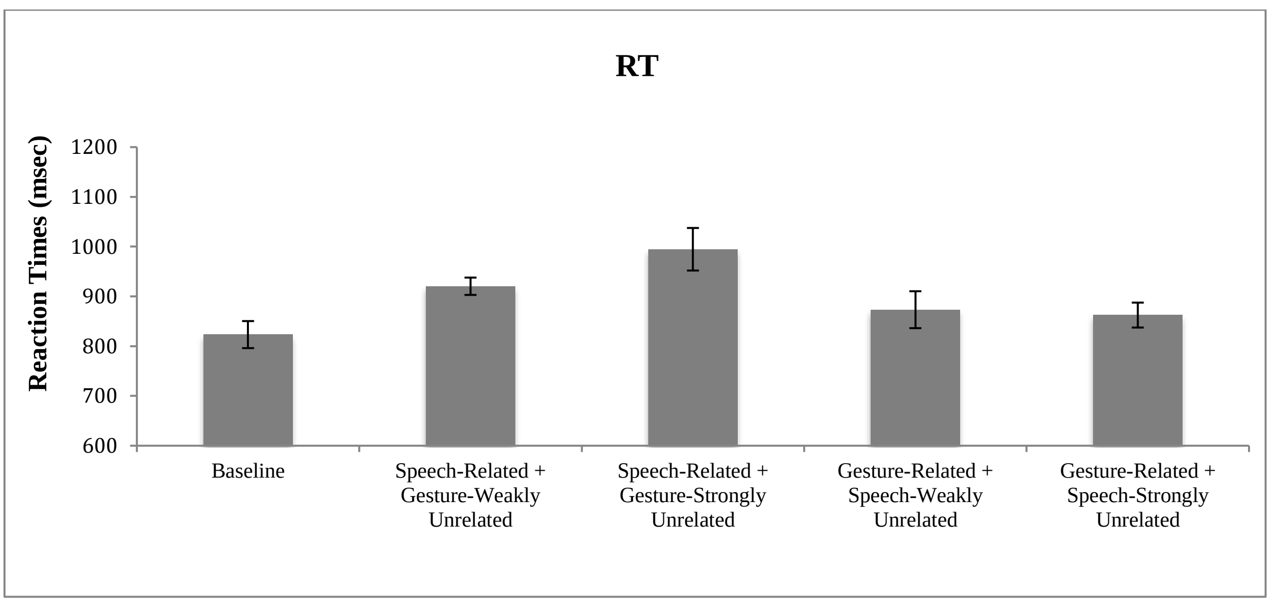


Figure 3.2. The results of the replication analyses for response latencies (RT) as a function of condition. Error bars represent standard errors.

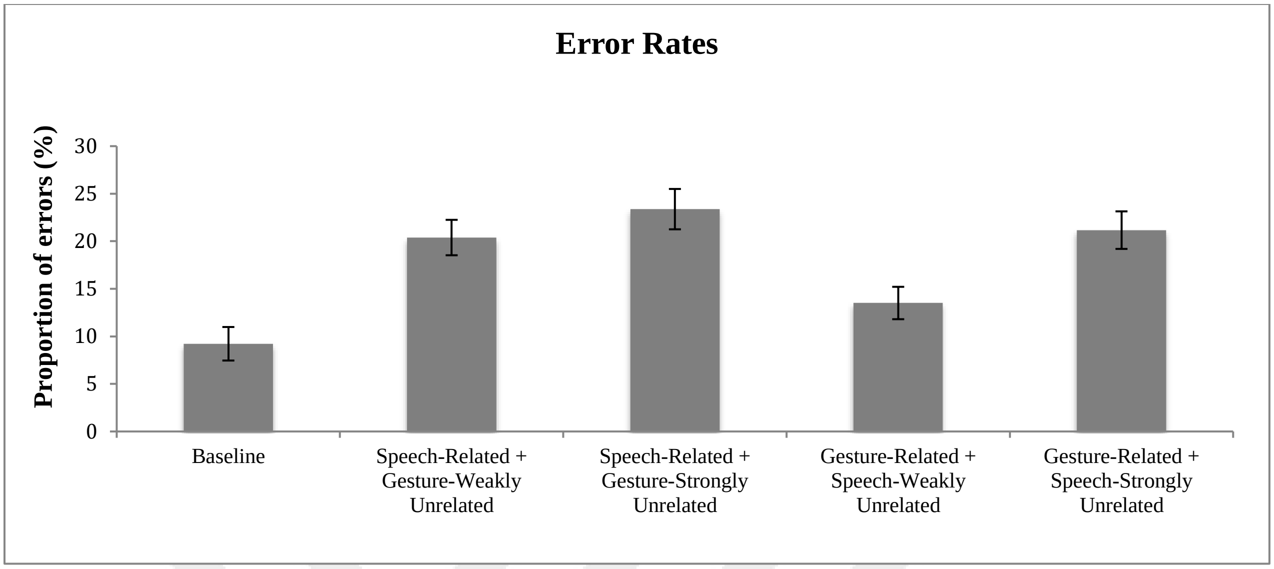


Figure 3.3. The results of the replication analyses for proportion of errors as a function of condition. Error bars represent standard errors.

3.4.2. Individual differences

Age and the number of years of education did not differ between males and females, *p*s> .05. There was also no statistically significant difference between males and females in any of the working memory scores, all *p*s> .05. Descriptive statistics for each individual difference task can be found in Table 3.1. Table 3.2 depicts the correlations among these scores.

Table 3.1. Descriptive statistics of scores for each individual difference task.

	Mean	Median	*SD*	Range
Operation Span Task				
Total Score	45.97	49	17.37	0-75
Total Number of Correct Letters	60.99	64	12.06	9-75
Math Errors	4.84	4	3.18	0 - 15
Corsi Block Span Task				
Forward Span	6.54	7	1.13	4-9
Backward Span	5.74	6	1.28	3-9
Digit Span Task				
Forward Span	8.13	8	1.14	5-11
Backward Span	8.05	8	1.40	5-12

Note: SD= standard deviation

Table 3.2. Correlation matrix for WM tasks, age, and education.

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
1. OSpan Total Score	1								
2. OSpan Correct Letters	.89**	1							
3. OSpan Math Errors	-.39**	-.47**	1						
4. Corsi Forward Span	.06	.08	-.02	1					
5. Corsi Backward Span	.03	-.04	.16	.08	1				
6. Digit Forward Span	.52**	.39**	-.10	.24**	.32	1			
7. Digit Backward Span	.44**	.28**	.01	0	.23**	.65**	1		
8. Age	.11	.13	.05	-.20	-.10	-.12	-.11	1	
9. Education	.08	.05	.02	-.19	-.19	-.04	-.08	.87**	1

Note: OSpan= Operation span task. Significance codes: ***p*< .01

3.4.3. Main analyses & Results

The main analyses comprised of two multi-level mixed effects models, one with the accuracy (i.e., correct, or incorrect) and the other with the congruency effect (i.e., RT for incongruent minus baseline) as the outcome variables. The multi-level mixed effects models were employed to examine how individual difference scores relate with the outcome variables. This method was appropriate to incorporate all data variance in our analyses. For the accuracy, we employed a logistic version of the model because the outcome variable was binary. The analyses were performed with lme4 package (Bates, Machler, Bolker, & Walker, 2014) on R (R Core Team, 2013).

The random effects included random intercepts of items (did some items yield more errors and increased congruency effect?) and participants (were some participants more prone to have errors and increased congruency effect?). To test the significance of different models, we used chi-square tests (i.e., likelihood ratio tests) for model comparisons. We added each predictor variable as fixed effects consecutively to the model to compare whether the inclusion of the new predictor significantly improved (i.e., explained the variance) the model compared to the previous model. To this end, we started with a null model (i.e., random item-intercept and participant-intercept only model), added one of the predictor variables as a fixed effect to the null model and compared it against the null model. Each predictor variable was added consecutively to the model and compared against the previous model containing all the other fixed effects except the one added in the corresponding step. The fixed effects included the modality (gesture-unrelated vs. speech-unrelated), the strength (weak vs. strong), the two-way interaction between these two, the Corsi Forward Block Span, the number of correct letters remembered in the Operation Span (O-Span) Task, and the two-way interactions between these WM scores and the modality, consecutively.

We scaled all continuous predictors (i.e., individual difference scores). This was done to simplify the interpretation and to alleviate the possible convergence problems for fitted models (Bates et al., 2019, p.17). To do so, we subtracted the average score from each participant's score and divide it by the standard deviation of the corresponding individual difference task. For visual WM, we used scaled scores of the Corsi Forward Block Span. For verbal WM, we had two measures: The Digit Span Task and the O-Span Task. Yet, forward and backward Digit Span scores were correlated with both Corsi Forward Block Span and the O-Span scores. The inclusion of the Digit Span scores would violate the assumptions of regression because of collinearity. Thus, we used the scaled scores of the number of letters correctly recalled in the O-Span Task as a measure of verbal WM capacity.

3.4.3.1. Accuracy

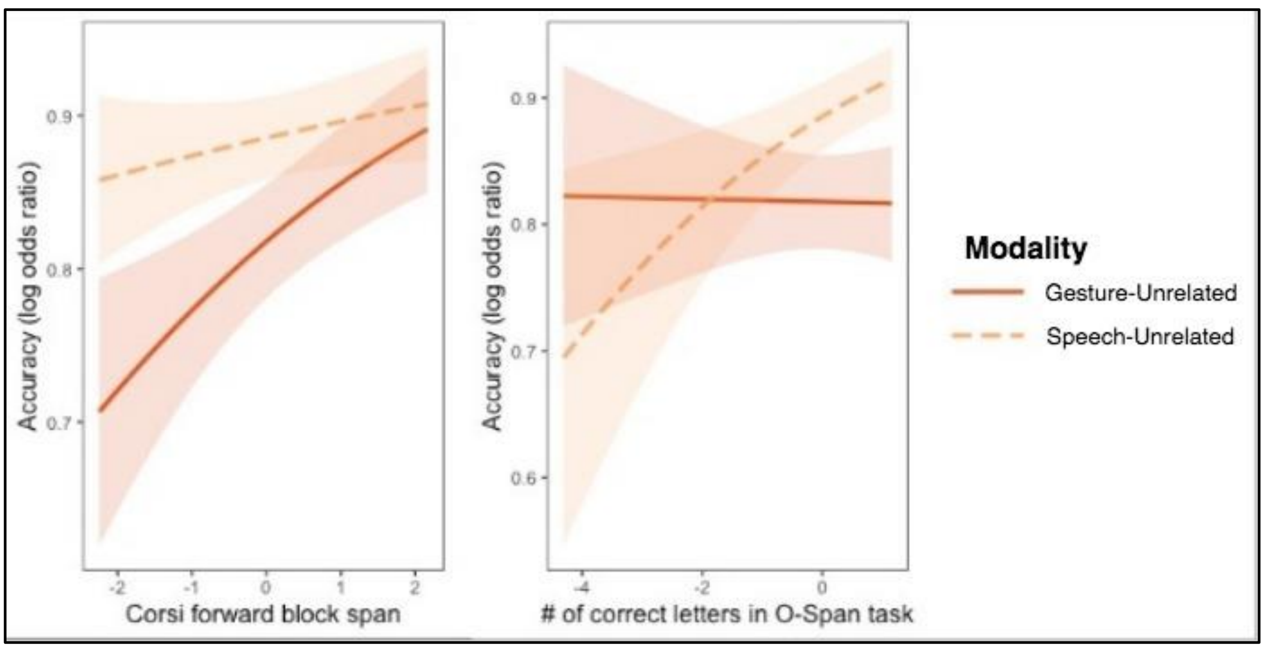
For accuracy, we used all experimental conditions, except the baseline condition. There were only unrelated conditions (with local speech-gesture incongruence) in the logistic version of the model, yielding 3080 responses (4 unrelated conditions * 10 items in each * 77 participants) in total. Out of these, 604 of the responses (19.6%) were incorrect.

The SU trials were associated with higher accuracy (i.e., higher proportions of correct responses) by .54 ± .14, $\chi^2(1) = 10.93$, $p < .001$. The inclusion of the strength improved the model, $\chi^2(1) = 15.08$, $p <$ 001. The strongly unrelated trials decreased accuracy by -.19 ± 13. Yet, the interaction between the modality and the strength was associated with the accuracy, $\chi^2(1) = 4.21$, $p < .05$. That is, the strongly unrelated trials were associated with extra decreased accuracy by -.40 ± .19 in the SU condition. The Corsi Forward Block span affected accuracy, $\chi^2(1) = 5.86$, $p < .05$, increasing it by .28 ± .09. The inclusion of the number of correct letters recalled in the O-Span Task did not improve the model significantly, $\chi^2(1) = 2.55$, $p = .11$. Although the inclusion of the two-way interaction between the Corsi Forward Block Span and the modality did not significantly improve the model, $\chi^2(1) = 2.31$, $p = .13$, the coefficients of the final model revealed a marginal significance ($p = .07$). The Corsi Forward Block span increased accuracy more in the GU trials by .29 ± .09. The interaction between the number of correct letters in the O-Span Task and the modality significantly improved the model, $\chi^2(1) = 10.85$, $p < .001$. The number of letters correctly recalled in the O-Span Task was associated with accuracy, increasing it by .29 ± .09 only in the SU condition (see Figure 3.4b). A summary of the final model's estimated coefficients and their corresponding standard errors can be found in Table 3.3.

Table 3.3. Model summary for accuracy.

	Coefficient	*SE*
FIXED EFFECTS		
Intercept	1.50***	.15
Modality (= Speech-Unrelated, SU)	.54***	.14
Strength (= Strongly Unrelated)	-.19*	.13
Corsi Forward Span	.28**	.09
O-Span score	-.01	.09
Modality × Strength	-.40*	.19
Modality ×Corsi Forward Span	-.17[†]	.09
Modality × O-Span score	.29***	.09
	Variance	
RANDOM EFFECTS		
Intercepts		
Subject	.36	
Item	.09	

Note: Significance codes= ***p< .001, **p< .01, *p< .05, [†]p< .1



Figures 3.4a and 3.4b. The relationship between log odds of accuracy (i.e., correct responding) and scaled scores of the Corsi forward block span (Figure 3.4a) and the number of correct letters in the O-Span Task (Figure 3.4b) as a factor of the modality (GU vs. SU). The hues around regression lines represent 95% confidence intervals. The range of the x-axes are different since the observed range differs across two tasks.

In sum, these analyses showed that participants were less accurate in the GU condition compared to SU condition. The strongly unrelated trials yielded less accuracy compared to the weakly unrelated trials, yet the difference was higher in the SU condition. For the individual difference scores, the Corsi Forward Block span (i.e., spatial WM) was associated with increased accuracy in both types of trials. Yet, this increase was higher (i.e., the slope was steeper) for the GU condition (see Figure 3.4a). The number of letters correctly recalled in the O-Span Task (i.e., verbal WM) was associated with increased accuracy only in the SU condition (see Figure 3.4b).

3.4.3.2. Congruency effect (Response times)

For the congruency effect, we calculated the difference in RTs between unrelated conditions and the baseline condition. To do so, we subtracted the RT of each item in the baseline condition from the RT of the corresponding item in every unrelated condition. This yielded 2250 responses in total. We already discarded outliers and trials with incorrect responses. Thus, for the missing data points, the difference scores were also missing.

The modality influenced the congruency effect, $\chi^2(1) = 12.67$, $p< .001$. The SU condition lowered the congruency effect by -71.57 ± 39.79 msec. The inclusion of the strength did not improve the model, $\chi^2(1) = 2.37$, $p= .12$. Yet, in the final model, strongly unrelated trials were associated with increased congruency effect by 74.65 ± 41.09 msec, at a marginal significance level of $p= .07$. The interaction between the modality and the strength was not associated with the congruency effect, $\chi^2(1) = 1.13$, $p= .29$. The Corsi Forward Block span and the number of correct letters in the O-Span Task were not associated with the congruency effect, $\chi^2(1) = .03$, $p= .85$ and $\chi^2(1) = .20$, $p= .66$, respectively. Also, the effect of the Corsi Forward Block span and the O-Span score on the congruency effect did not change across the GU and the SU conditions (see Figures

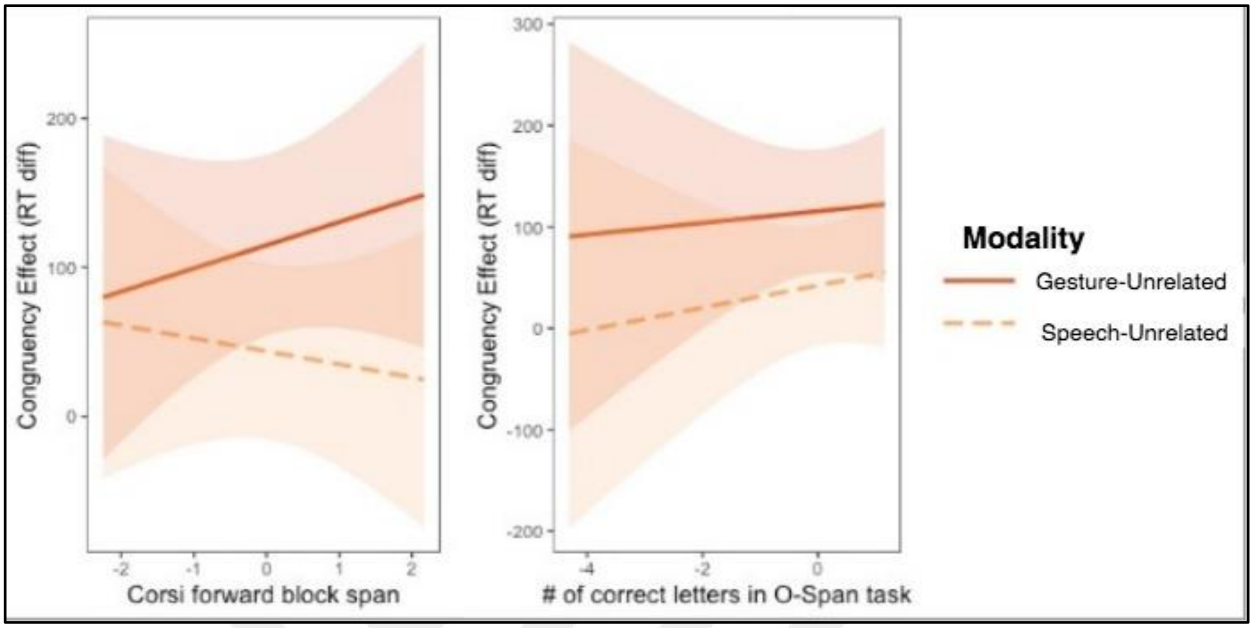
3.5a and 3.5b). The inclusion of two-way interaction terms did not significantly improve the model, $\chi^2(1) = .68$, $p = .41$ and $\chi^2(1) = .03$, $p = .87$, respectively. A summary of the final model's estimated coefficients and their corresponding standard errors can be found in Table 3.4.

In sum, we found that the congruency effect (i.e., incongruent RT minus baseline RT) was higher in the GU condition compared to the SU condition. None of the WM scores were associated with this congruency effect.

Table 3.4. Model summary for congruency effect (RT for unrelated minus baseline in msec)

	Coefficient	***SE***
FIXED EFFECTS		
Intercept	114.50**	37.07
Modality (= Speech-Unrelated, SU)	-71.57†	39.79
Strength (= Strongly Unrelated)	74.65	41.09
Corsi Forward Span	15.58	23.77
O-Span score	5.77	25.42
Modality × Strength	-59.88	57.12
Modality ×Corsi Forward Span	-24.29	29.07
Modality × O-Span score	5.27	32.01
	Variance	
RANDOM EFFECTS		
Intercepts		
Subject	9090	
Item	4222	

Note: Significance codes= ***p< .001, **p< .01, *p< .05, †p< .1



Figures 3.5a and 3.5b. The relationship between congruency effect (RT of each unrelated condition minus baseline) and scaled scores of the Corsi forward block span (Figure 3.5a) and the number of correct letters in the O-Span Task (Figure 3.5b) as a factor of the modality (GU vs. SU). The hues around regression lines represent 95% confidence intervals. The range of the x-axes are different since the observed range differs across two tasks.

3.5. Discussion

The current study investigated the role of individual differences in verbal and visual-spatial resources as measured through verbal (the number of letters correctly recalled in the O-Span Task) and spatial WM (the forward Corsi Block Span Task) capacities on how people process concurrent gestural vs. spoken expressions during an online task (Kelly et al., 2010a). The results corroborated previous findings on the bidirectional interaction between speech and gesture during comprehension. More importantly, our results suggested individual differences in the differential processing of concurrent gestural vs. spoken information depending on the propensity to use visual-spatial vs. verbal cognitive resources.

We employed the adapted version of Kelly et al.’s (2010a) paradigm used in an English-speaking sample. As stated earlier, the first motivation of the current study was to replicate earlier findings on the bidirectional relationship between speech and gesture in Turkish. Evidence suggests cross-cultural differences in terms of the production and perception of co-speech gestures (Azar et al., 2019; see Kita, 2009 for review). For example, monolingual Chinese speakers use spontaneous gestures less frequently compared to monolingual American English speakers, and English-French bilinguals gesture more when speaking in French, a high-gesture language, than when speaking in English (Laurent & Nicoladis, 2015; So, 2010). Thus, the replication of findings on the interactive nature of gesture and speech in different languages and cultures is important. In this regard, one of the strengths of the current study is to test these claims and the validity of the online priming paradigm (Kelly et al., 2010a) in a different language.

It is important to note that although we used the same experimental task with the same videos, the stimuli that were used in the current study were still somewhat different. As explained earlier, several action primes were discarded since there was not a single Turkish verb describing some videos. Despite this, we replicated the main findings reported in Kelly et al.'s (2010a) paper with a slightly different set of results regarding the modality. First, Kelly et al. (2010a) found that local speech-gesture incongruence yielded higher error rates and longer RTs compared to the baseline condition. This finding was fully replicated in the current study. Yet, we also found that participants were slower and more error-prone when target gestures were unrelated to action prime (GU) compared to when target speech was unrelated (SU). We found the exact same pattern in both RTs and error rates, which was lacking in the original study. Kelly et al. (2010a), on the other hand, found higher RTs for speech-unrelated compared to the gesture-unrelated trials, whereas there was no difference between two conditions in terms of the error rates. The discrepancy between these findings regarding the modality might be due to the smaller sample size used in Kelly et al.'s study and they might not capture the effect (which was visible in both error rate and

RT measures in the current study) due to reduced statistical power. Although we used the adapted version of the priming task used in Kelly et al., it consisted of a different set of stimuli, and the task was employed in a different language and culture. Turkish-speaking adults might exhibit different patterns in processing gestural vs. spoken expressions compared to English-speaking adults. The reason and the mechanisms behind these cross-linguistic and cross-cultural differences are out of the scope of the current paper.

Why would there be an asymmetric relationship between speech and gesture as our results indicated? In particular, why would unrelated gestural information hinder comprehension to a greater extent than unrelated speech? One possibility might be the isomorphism between actions in the action prime video and the gestures in the target videos. In the experimental task, participants were presented with videos of action primes consisting of different actions carried out with an instrument, followed by a target video in which an actor makes a gesture and utters a word. Yet, the related gestures were nearly identical to the previous action, except for the absence of the instrument (i.e., object). Due to higher isomorphism between the actions in the preceding video and the gestures in the target videos, it might be harder for participants to suppress visual-motoric information. Second, gestures preceded the onset of the spoken expressions by 200 ms. This might lead to the gestural information to be processed earlier, causing more adverse effects in comprehension. Nevertheless, we find this quite unlikely since gestures mostly precede spoken expressions in natural communication (Leonard & Cummins, 2011; Morrel-Samuels & Krauss, 1992; Obermeier, Holle, & Gunter, 2011). The stimuli were prepared both in the current study and in Kelly et al.'s (2010a) study on purpose to make it as ecologically valid as possible.

Our results showed a main effect of the strength of (un)relatedness only when the target speech conveyed unrelated information. There was no difference between strongly and weakly unrelated gesture conditions. This might again be due to higher isomorphism between the depicted actions in the action prime videos and the target gestures. For example, when participants are judging the relatedness between a visually presented action information (e.g., seeing someone to chop vegetables) and a slightly unrelated gesture (e.g., seeing someone to make a cutting gesture), they might readily and easily decide whether these two visual information are identical to each other. This might be the reason why there was no difference between strongly and weakly unrelated gesture trials.

The second motivation of the current study was to uncover the cognitive mechanisms governing attention and/or sensitivity to process gestural and spoken expressions when they were presented together (i.e., during multimodal comprehension). Several behavioral, electrophysiological, and neuroimaging studies have repeatedly demonstrated that listeners are sensitive to co-speech gestures that are semantically analyzed and integrated with the concurrent speech during comprehension (Cassell et al., 1999; Kelly, Barr, Church, & Lynch, 1999; Kelly et al., 2004; Wu & Coulson, 2005; see Özyürek, 2014 for a review). Studies on multisensory integration, in general, and speech-gesture integration, in particular, suggested individual differences in the ability to integrate audio-visual information and gesture perception (Beauchamp et al., 2004; Nagels, Kircher, Steines, Grosvald, & Straube, 2015; Nath & Beauchamp, 2012; Schmalenbach, Bilio, Kircher, van Kemenade, & Straube, 2017; Werner & Noppeney, 2010; Wu & Coulson, 2014a, 2014b). Moreover, gesture perception and the ability to integrate gestural information with that of the speech are impaired in elderly adults (Cocks et al., 2011; Thompson, 1995; Thompson & Guzman, 1999) and in several clinical groups such as people with autism (Hubbard et al., 2012; Silverman, Bennetto, Campana, & Tanenhaus, 2010), schizophrenia (Nagels, Kircher, Grosvald, Steines, Straube, 2018), and aphasia (Cocks, Byrne, Prichard, Morgan,

& Dipper, 2018). Insights into which mechanisms contribute to these individual differences and deficits are just beginning to emerge.

To this end, the current study aimed to fill this gap by asking how individual differences in the propensity to use verbal and spatial cognitive resources (as measured through verbal and spatial WM capacity) related with differential processing of gestural vs. spoken information during comprehension. Although we did not find any effect for the congruency effect (incongruent RTs minus baseline RTs), accuracy analyses suggested differential processing of gestural vs. spoken expressions depending on the individual differences in verbal and spatial WM capacities. First, we found that people with higher spatial WM capacity (higher score in the Corsi forward block span) were better at resolving local speech-gesture incongruence (i.e., higher accuracy) compared to people with lower spatial WM capacity. Logistic mixed effects analyses also revealed a marginally significant interaction with the modality. This positive relationship between spatial WM capacity and accuracy was especially higher under visual cognitive load (i.e., in the GU condition). That is, people with higher spatial WM capacity were better at maintaining and processing different visual semantic information (i.e., gestures). In contrast, people with higher verbal WM capacity (i.e., higher number of letters correctly recalled during the O-Span Task) were more accurate under verbal cognitive load. That is, people with higher verbal WM capacity were better at maintaining and processing different verbal semantic information (i.e., speech). In general, the results suggest that the processing capacity for the concurrent gestural vs. spoken expressions depend on visual-spatial and verbal skills.

These results could also be interpreted in a different way. We predicted and found that people with higher verbal WM capacity were more accurate in the SU condition, and people with higher visual WM capacity were more accurate in the GU condition. Nevertheless, in the SU condition, participants needed to process a related gesture target together with an unrelated speech target. In the GU condition, this time, they needed to process a related speech target simultaneously with an unrelated gesture target. Then, were people with higher verbal WM capacity better at processing related visual information or unrelated verbal information? Likewise, were people with higher visual WM capacity better at processing related verbal information or unrelated visual information?

Although both explanations can be plausible, we reasoned that the observed effects of WM would be on the processing of interfering information that taxes WM resources by creating extra processing load in the respective modality (Engle, 2002; Kane et al., 2001; Kane & Engle, 2003). Participants were required to process both types of target information to respond, thus, they did not ignore the unrelated target. Nevertheless, along with the processing of the related information, they still needed to maintain an unrelated information in working memory and resolve this incongruence to decide whether that target was relevant to the action prime or not. Due to cognitive load created by the maintenance and resolution of incongruent information in the experimental task and the role of WM in processing information under high processing load (e.g., visual, or verbal interference; Baddeley, 2000; Conway et al., 2001; Engle, 2002; Kane & Engle, 2003; Kane et al., 2001), we reasoned that the observed effects of WM on performance is due to the processing efficiency in the domain in which there was processing load (i.e., in the unrelated targets).

These results suggest that spatial resources are integral to process gestural information whereas verbal resources are essential to process spoken information. These findings are, in part, in line with a previous study indicating an important role for visual-spatial WM capacity in gesture-speech integration process (Wu & Coulson, 2014a). In Wu and Coulson study, people with higher visual-spatial WM capacities benefited more from congruent gestures, whereas there was no relation between verbal WM capacity and the gesture congruency. As mentioned earlier, the spoken expressions were always related with the picture probe for the target probes in Wu and Coulson's study, hence, the relatedness between the speech and the following picture probe was not manipulated. The lack of evidence for the role of verbal WM resources in Wu and Coulson (2014a) might be due to the lack of this manipulation. To this end, the current study, with the employment of both types of unrelated information (speech and gesture), asked if verbal WM might relate with the processing of speech. Second, participants were asked to focus on speech and disregard the gestural information in Wu and Coulson study. This manipulation is not ecologically valid since in natural multimodal communication, listeners automatically attend to both types of communicative signals. Also, when participants are asked to pay more attention to one channel of information (i.e., prioritize the processing of one over the other), it is hard to understand people's natural tendencies in processing gestural vs. spoken expressions. Third, Wu and Coulson used an off-line task, which requires participants to wait for responding until a picture probe has appeared on the screen. However, participants have extra time to process both types of information and resolve any type of mismatches till the appearance of the picture probe, making it more difficult to describe processing speed for gesture and speech. In this regard, in the current study, we employed an online task in which participants were presented with and asked to attend to both gestural and spoken information.

Our results suggested that people with higher visual-spatial cognitive resources are more sensitive to gestures, which was evident in the higher efficiency in resolving and processing unrelated gestural information. People with higher verbal cognitive resources, on the other hand, are more sensitive to speech, leading to better processing of unrelated spoken information. These results support the hypothesis that visual-spatial resources are primary for gesture processing whereas verbal resources do not play much role in the processing of gestural information. Overall, our results suggested that individual differences in visual-spatial and verbal WM capacity relate with how listeners differentially attend to gestural vs. spoken expressions during multimodal comprehension.

These findings are in line with previous research showing individual differences in learning from visual vs. verbal materials (Mendelson, 2004; Riding & Ashmore, 1980; Riding, Burton, Reese, & Sharratt, 1995). We all differ in how we interact with our environment, extract, and perceive information from it. People have relative predispositions to acquire, process, and organize information (Alfred & Kraemer, 2017; Jonassen & Grabowski, 1993; Kirby, Moore, & Schofield, 1988; Kraemer, Rosenberg, & Thompson-Schill, 2009; Riding et al., 1995). They show differential sensitivity to and gain differential benefit from visual-spatial vs. verbal information based on whether they have visual vs. verbal cognitive styles to process and encode information (Ausburn & Ausburn, 1978; Riding & Ashmore, 1980; Kirby et al., 1988; Kiat & Belli, 2018; Kóc-Januchta, Höffler, Thoma, Prechtl, & Leutner, 2017, Meneghetti, Labate, Grassano, Ronconi, & Pazzaglia, 2014; for cf. see Kollöffel, 2012). For example, students with higher spatial abilities benefited from the presence of 3-D models when learning cell biology whereas students with lower spatial abilities became cognitively loaded with the presence of 3-D models and did not benefit from the use of these visual-spatial aids (Huk, 2006). Although prior research suggested that gesture perception and gesture-speech integration is automatic (Kelly et al., 2010a; 2010b), there is accumulating evidence suggesting that this process can be modulated by several internal and external factors (e.g., Drijvers & Özyürek, 2017; Holle, Obleser, Rueschemeyer, Gunter, 2010; Holler et al., 2014; Kelly, Ward, Creigh, & Bartolotti, 2007; Obermeier, Dolk, & Gunter, 2012; Obermeier, Kelly, & Gunter, 2015; Straube, Green, Jansen, Chatterjee, & Kircher, 2010). For example, listeners in suboptimal communicative situations (e.g., noise, Drijvers & Özyürek, 2017) and individuals who chronically experience these situations (e.g., hearing impaired people; Obermeier et al., 2012) are more sensitive to gestural information. The current study extends this research by showing that not only contextual factors, but also individual differences in the propensity to use visual-spatial vs. verbal resources may be related to gesture and speech perception.

Why is it important to uncover the mechanisms governing gesture and speech perception? Several studies have demonstrated beneficial effects of gestures on language comprehension, encoding, and learning for both children and adults, and in different contexts extending from lab environment to instructional classroom settings (Alibali & Nathan, 2012; Aussems & Kita, 2019; Beattie & Shovelton, 1999; Kim, Roth, & Thom, 2011; Roth & Welzel, 2001; Sekine & Kita, 2015; Sekine, Sowden, & Kita, 2015; Macedonia, Muller, Friederici, 2011; Yeo, Ledesma, Nathan, Alibali, & Church, 2017). However, our results suggest that not everyone is equally sensitive to gestures and benefit from gestures. To achieve optimal learning and comprehension, individual differences should be taken into consideration. Research emphasizing the cognitive style as a learner characteristic might benefit instructional design and help attain better learning outcomes (Ausburn & Ausburn, 1978). Considering these, to our knowledge, this is the first study addressing the individual differences in the differential sensitivity to speech vs. gestures and give insight into who may benefit more from gestural vs. spoken expressions in multimodal comprehension and learning settings.

One of the strengths of the current study is the employment of the mixed effects analyses that takes all data variance into account (Baayen, Davidson, & Bates, 2008). Yet, this study has some limitations. One caveat of the current study is that the study involved healthy, collegeeducated participants from a university, which accepts top 15 to 20% of their cohort. Thus, although we achieved variability to a certain extent in our individual differences tasks, a more comprehensive sample in which scores in verbal and spatial WM tasks exhibit full variability might give more definitive results. Future studies can address this issue in different samples with verbal and visual-spatial problems such as people with speech deficits, language learners, and clinical groups with spatial and visual processing impairments. Moreover, the role of visual-spatial vs. verbal resources on the sensitivity to gesture vs. speech can be investigated by manipulating the requirements of the task at hand. For example, future studies can investigate how attention to and the comprehension of gestural and spoken expressions change under visual-spatial vs. verbal cognitive load. Second, although the experimental task was proved to be a reliable index to measure sensitivity to gestures, gesture-speech incongruence is rare in natural communication. Moreover, the current experimental task does not require participants to actively integrate gestural information with spoken information. Yet, in certain situations, successful communication requires integration of gestural semantics with that of speech, such as when gestures express non-redundant and additional information that is expressed in speech (Church, Garber, & Rogalski, 2007; Goldin-Meadow & Sandhofer, 1999; Kelly, 2001). Future studies can employ different paradigms requiring participants to integrate gestural semantics to that of the speech to achieve successful comprehension. Research on the role of visual-spatial and verbal resources on gesture comprehension can be extended to different contexts including more natural everyday settings like instructional classroom settings.

3.6. Conclusion

In conclusion, this study examined the individual differences of using visual-spatial vs. verbal resources on people’s sensitivity to gestural vs. spoken message. We demonstrated that individuals with higher spatial WM capacity (but not higher verbal WM capacity) were more efficient in processing gestures. In contrast, individuals with higher verbal WM capacity were more perceptive to unrelated information in spoken expressions. These findings suggest that people's sensitivity to different types of information can be related to their tendency to use visual-spatial vs. verbal cognitive resources during the encoding, maintenance, and the processing of information.

CHAPTER 4

INFORMATIVENESS OF GESTURES GUIDES VISUAL ATTENTION DURING SPATIAL LANGUAGE COMPREHENSION

This chapter is based on Özer, D., Karadöller, D., Özyürek, A., & Göksun, T. (in preparation). Informativeness of gestures guides visual attention during spatial language comprehension.

Abstract

Gestures help speakers and listeners during communication and thinking, particularly for visual-spatial information. Speakers use gestures to complement their speech when communicating spatial information, such as describing spatial relations between objects (e.g., saying "The pencil is *here*" and gesturing to the right side to express that the pencil is on the speaker's right). Visual information conveyed by gestures enhances listeners' comprehension. Whether and how listeners allocate overt visual attention to gestures in different contexts is mostly unknown. The present study investigated how the relative information value of gestures in relation to the accompanying speech guides listeners' visual attention to gestures. We asked whether and how (1) listeners gazed at gestures when they conveyed complementary vs. redundant information to the speech and (2) overt gaze allocation to gestures related to how much information listeners glean from those gestures. We demonstrated that listeners fixated to gestures more when they expressed complementary (i.e., specifying speech) than redundant information to the accompanying speech. Overt visual attention to gestures did not predict how much listeners benefited from observing gestures. These results suggest that the relative communicative importance of gestures guides listeners' visual attention; however, semantic information conveyed by gestures can be abstracted peripherally without necessitating direct visual attention.

4.1. Introduction

Individuals use their hands to think and communicate (McNeill, 1992), particularly for spatial information, such as relative spatial relations between objects (e.g., Alibali, 2005; Beattie & Shovelton, 1999a, 2001; Dargue, Sweller, & Jones, 2019; Hostetter, 2011; Lavergne & Kimura, 1987). Speakers also use gestures to complement their speech when describing spatial relations (e.g., Beattie & Shovelton, 2006; Holler & Stevens, 2007). For example, gestures might complement spoken deictic constructions. When a speaker says, "The pencil is *here*," and gestures to the right side to express that the pencil is on the speaker's right (e.g., Emmorey & Casey, 2001). In turn, listeners can benefit from observing those gestures when comprehending objects' relative position (Beattie & Shovelton, 1999a, 1999b, 2001, 2002; Holler, Shovelton, & Beattie, 2009). However, little is known about how listeners allocate visual attention to gestures during the comprehension of spatial relations between objects, particularly when gestures express complementary information (i.e., information present only in gesture) in relation to the accompanying speech. This will be the focus of the present study. In the next sections, we will discuss the current literature on the visual attention to gestures and the role of gestures in the communication of spatial information, before presenting the current study.

4.1.1. Visual attention to gestures

In face-to-face interaction, addressees mainly look at their interlocutors' faces, as maintained mutual gaze between interlocutors is a culture-specific social norm (Argyle & Cook, 1976; Argyle & Graham 1976; Gobel, Kim, & Richardson, 2015). Eye-tracking research showed that listeners/viewers fixate only a minority of speakers' gestures and spend approximately 90% of the total viewing time fixating on speakers' faces (i.e., mostly the nose bridge and the eye area) in both live face-to-face and video interactions (Beattie, Webster, & Ross, 2010; Gullberg & Holmqvitz, 1999, 2006; Gullberg & Kita, 2009; but see Nobe, Hayamizu, Hasegawa, & Takahashi, 1997, 2000 for increased overt gaze allocation to gestures when interacting with an anthropomorphic agent). Gazing at gestures did not predict the information uptake from those gestures (e.g., Beattie et al., 2010; Gullberg & Kita, 2009). These studies suggest that listeners do not need to directly fixate on gestures to glean information from them as they can be processed peripherally. These findings are also in line with the sign language literature, showing that peripheral perception is sufficient for the signs to be processed in parallel with overt visual attention to the face (e.g., Emmorey, Thompson, & Colvin, 2009; Swisher, Christie, & Miller, 1989). However, several kinematic features of gestures modulate listeners' overt gaze allocation. Listeners allocate more overt visual attention to gestures that were executed in the peripheral gesture space (e.g., away from the body/torso) compared to ones that were performed in the central gesture space (e.g., closer to face and body/torso), especially for the vertical axis (Gullberg & Holmqvitz, 1999). Gestures with slower strokes (i.e., the most meaningful part of the gesture) and longer post-stroke holds attract more fixations (Gullberg & Holmqvitz, 2006; Gullberg & Kita, 2009; Nobe et al., 1997, 2000). Additionally, listeners look at gestures more if speakers themselves looked at their own gestures first (i.e., auto-fixations; Gullberg & Holmqvitz, 1999; Gullberg & Kita, 2009; but see Gullberg & Holmqvitz, 2006) as speakers' gaze direct listeners' attention to the target of their gaze.

The relative communicative or information value of gestures in relation to the accompanying speech might also modulate listeners' visual attention to gestures. When the spoken channel is not sufficient for comprehension, listeners might rely more strongly on visual information expressed by gestures and allocate more direct attention to them. Listeners benefit from gestures more in challenging listening conditions in which speech intelligibility is hindered, such as in degraded vs. clear speech comprehension (e.g., Drijvers & Özyürek, 2017, 2018; Holle et al., 2010), when listeners have hearing impairments (e.g., Obermeier, Dolk, & Gunter, 2012), or when communicating in a non-native vs. native language (e.g., Church, Ayman-Nolley, & Mahootian, 2004; Dahl & Ludvigsen, 2014; Drijvers & Özyürek, 2018, 2020; Kelly, McDevitt, & Esch, 2009; Macedonia, Müller, & Friederici, 2011; Sueyoshi & Hardison, 2005). Indeed, studies suggest that the degree of language experience modulates how much listeners allocate overt visual attention to gestures (e.g., Drijvers, Vaitonyté, & Özyürek, 2019; but see Gullberg & Holmqvitz, 1999). Listeners fixated more on gestures when watching narratives told in a partially comprehensible (i.e., French) or completely incomprehensible language (i.e., Russian) compared to when watching narratives told in their native language (i.e., Flemish; Rimé, Boulanger, & d'Ydewalle, 1988). A recent study investigated how native and non-native listeners allocated overt visual attention to gestures in clear and degraded speech when comprehending Dutch action verbs (e.g., to salt, Drijvers et al., 2019). They showed that non-native listeners looked at gestures more both during clear and degraded speech comprehension. Moreover, only native, but not non-native, listeners' gazing at gestures predicted how much they benefitted from observing gestures during degraded speech comprehension. These findings suggest that although non-native listeners gazed at gestures more compared to native listeners, they might be more hindered by degraded speech and need more cues (i.e., phonological cues in visible speech) to aid comprehension during degraded speech (Drijvers et al., 2017, 2018).

Another factor indexing the relative communicative or informational value of gestures is the complementary distribution of information across speech and gesture (i.e., the degree of the redundancy between information expressed by gesture vs. speech). Although speech and gesture express a unified meaning representation, gestures have different representational properties from speech. Gestures are realized in the visual-spatial domain and convey meaning holistically, whereas speech conveys meaning in a linear and segmented fashion (McNeill, 1992). Therefore, gesture and speech rarely express completely redundant information, and in most cases, gestures add semantic information to the accompanying speech. For example, consider a speaker who says, "There is a ball," and makes a gesture as if she is holding a ball with her two hands. In this case, both speech and gesture represent the same referent (i.e., ball), and the redundancy between these two is high. However, gesture still conveys additional information to the one expressed in speech, such as the ball's size. A two-hand gesture might represent a bigger ball such as a volleyball compared to a single-hand gesture representing a smaller ball such as a tennis ball. Therefore, the semantic redundancy between speech and gesture lies on a continuum (Goldin-Meadow, 2003). At the one end of the continuum are cases in which there is minimal semantic overlap between speech and gesture. Gestures often complement and provide additional meaning to spoken expressions on a semantic level (e.g., Beattie & Shovelton, 2006; de Ruiter, Bangerter, & Dings, 2012; Gerwing & Allison, 2009; Melinger & Levelt, 2004). Children and adults use complementary gesture-speech combinations both at clause- and word-level (e.g., Alibali, Evans, Hostetter, Ryan, & Mainela-Arnold, 2009; Beattie & Shovelton, 2006; Church & Goldin-Meadow, 1986; Holler & Stevens, 2007; Hostetter & Alibali, 2011; Özçalışkan & Goldin-Meadow, 2005). In turn, listeners pick up gestural information to disambiguate speech and combine spoken and gestural semantics for successful comprehension of a given message (e.g., Cocks, Morgan, & Kita, 2010; Demir-Lira et al., 2018; Kelly et al., 1999; Sekine, Sowden, & Kita, 2015; Sekine et al., 2020).

The redundancy of information expressed by gesture vs. speech can affect listeners' attention to gestures. Related to the current work, a recent study (Yeo & Alibali, 2017) investigated whether listeners allocated more direct visual attention to gestures that express information necessary to disambiguate fluent vs. disfluent speech (i.e., speech with filled pauses "*um*"). In this study, the actress said, "The triangle changed color," and made a gesture of an upward-pointing triangle followed by a shape array containing either both upward- and downward-pointing triangles (i.e., non-redundant gesture) or only an upward-pointing triangle (i.e., redundant-gesture). The total fixation duration to gestures did not differ across redundant vs. non-redundant gestures. However, when analyzed whether listeners fixated to gestures at least once in a dichotomous way (0-1), they found that listeners tended to fixate to gestures more when they expressed non-redundant compared to redundant information. Moreover, they tended to gaze at gestures more when speech was disfluent than when it was fluent (Yeo & Alibali, 2017). Although analyses in this study should be taken cautiously as they measured whether listeners fixated to the gesture space at least once across different conditions, results suggest that perceived communicative importance of gestures emerging from the intelligibility of the alternative channel of expression (disfluent vs. fluent speech) or the complementary nature of gestures to the accompanying speech (redundant vs. non-redundant gestures) modulates how listeners direct visual attention to gestures.

4.1.2. Gestures in spatial context

Gestures have a special role in spatial cognition (Alibali, 2005; Dargue et al., 2019; Hostetter, 2011). Gestures help both speakers and listeners when expressing, communicating, and thinking about spatial information such as route descriptions (e.g., Allen, 2003; Arslan & Göksun, in press; Emmorey, Tversky, & Taylor, 2000; Özer et al., 2017; So, Shum, & Wong, 2015), relative position of objects (e.g., Beattie & Shovelton, 1999a, 1999b, 2001, 2006; Driskell & Radtke, 2003; Göksun, Lehet, Malykhina, & Chatterjee, 2013; Holler et al., 2009; McNeil, Alibali, & Evans, 2000), mental rotations (e.g., Chu & Kita, 2008, 2011; Emmorey & Casey, 2001; Göksun, Goldin-Meadow, Newcombe, & Shipley, 2013; Hostetter, Alibali, & Bartholomew, 2011), or describing motion in space (e.g., Kita & Özyürek, 2003; Kartalkanat & Göksun, 2020; Yeo & Alibali, 2018). Gestures are also more effective than spatial speech in learning and encoding spatial information (Chu & Kita, 2011; Ehrlich, Levine, & Goldin-Meadow, 2006; So et al., 2015).

Gestures are realized in the visual-spatial domain. They can afford the representation of visual-spatial information better than speech, in which meaning is represented hierarchically in a linear format (McNeill, 1992). Therefore, speakers design their multimodal utterances in a way that each modality conveys information that is appropriate for their expressive properties. Given gesture's expressive potential, speakers tend to convey visual-spatial information such as object size, object location, manner of movement, and spatial relations in gestures rather than in speech (Beattie & Shovelton, 2006; Emmorey & Casey, 2001; Holler & Stevens, 2007; Kita & Özyürek, 2003). Indeed, research showed that speakers encode spatial information such as object size only in their gestures rather than in speech, especially when the size information is crucial (Beattie & Shovelton, 2006) or when they talk to interlocutors with whom the size information is new (Holler & Stevens, 2007). Moreover, when describing mental rotation of objects, speakers tend to express rotation direction only in their gestures that complement deictic anaphoric constructions (e.g., "Turn it this way") in their concurrent speech (Emmorey & Casey, 2001). These spoken deictic constructions such as verbal pointers or framing devices refer to the speakers' gestures and signal listeners that an essential part of the message is conveyed through gestures. For example, a speaker who says, "The boy is moving *like this*," can produce up and down movement with a hand to represent the manner of motion, which is complementary to the spoken information. In such a multimodal utterance, gestures express the critical complementary information to disambiguate the accompanying speech. One spatial context in which speakers often use gestures to complement their speech with such spoken deictic constructions is the descriptions of relative spatial relations between objects (e.g., saying "The candle is *here*" while making a gesture that shows the right side to express that the candle is on the speaker's right side, Karadöller, Sümer, & Özyürek, in press). The focus of the current paper is on the visual attention to gestures that complement spoken deictic constructions (i.e., *here)* when comprehending relative spatial relations between objects.

Speakers use an abundant number of gestures to describe relative spatial relations (e.g., left-right, on-under, front-behind, next to) between a figure object (i.e., the object whose relative position to be located) and a ground object (i.e., the reference object; Göksun et al., 2013; Karadöller, in press; Talmy, 1983). Observing those gestures, in turn, enhance listeners' comprehension of such descriptions of spatial relations (e.g., Beattie & Shovelton, 1999a, 1999b, 2001; Holler et al., 2009). Nevertheless, the facilitative effects of observing gestures might be different for different types of spatial relations. Some spatial relations such as *left-right* (i.e., *viewpoint-dependent*) require viewpoint alignment and pose a challenge to both speakers and listeners compared to *viewpoint-independent* spatial relations such as *on-under* (Galati & Avraamides, 2013; Galati, Michael, Mello, Greenauer, & Avraamides, 2013; Keysar, Barr, Balin, & Brauner, 2000). Speakers of languages that tend to use the relative frame of reference to talk about spatial relations (e.g., Turkish, English, and many Indo-European languages) face a viewpoint coordination problem (e.g., Miller & Johnson Laird, 1976). This coordination problem arises from the position of interlocutors when they interact as facing each other.

In most cultures, speakers (both signers and non-signing hearers) describe left-right directionality from their own egocentric perspective (e.g., Miller & Johnson-Laird, 1976, but see Mainwaring, Tversky, Ohgishi, & Schiano, 1995). Hearing non-signer addressees mostly interpret speakers' gestures egocentrically as well, showing a lack of viewpoint alignment (Pyers, Perniss, & Emmorey, 2015). The gestural description of left-right directionality from the speaker's egocentric perspective creates an incongruence between the veridical position in which the gesture is made and the intended meaning. Thus, listeners/viewers experience an additional visual-spatial load when they process gestures that do not align with their own perspective. In line with this, research suggests that gestural expressions of viewpoint-dependent (left-right) vs. viewpoint-independent spatial relations have differential effects on comprehension. Observing gestures that express spatial relations requiring perspective-taking (such as *on, under, in*) enhance listeners' comprehension (e.g., Beattie & Shovelton, 1999a, 1999b, 2001; Holler et al., 2009) whereas gestures that express spatial relations requiring perspective-taking (such as left-right) hinder listeners' comprehension (Gullberg & Kita, 2009; Hostetter, Murch, Rothschild, & Gillard, 2018; Pyers et al., 2015). An open question is how listeners allocate visual attention to and benefit from gestures during comprehension of spatial relations that require viewpoint alignment or not.

4.2. The present study

We examine whether and how listeners allocate visual attention to gestures that complement spoken deictic constructions during spatial language comprehension. More specifically, we ask (1) whether listeners gaze at gestures more when they express critical complementary information (i.e., when gestures are coupled with non-informative ambiguous speech) compared to redundant information (i.e., when gestures are coupled with informative speech) in relation to the accompanying speech, and (2) whether and how gazing at gestures relates to how much listeners benefit from observing gestures during the comprehension of viewpoint-dependent (left-right) and viewpoint-independent (on-under) spatial relations. To this end, we used eye-tracking to record participants' eye movements.

To investigate these questions, we employed an offline comprehension paradigm in which we presented participants with short video clips of an actress describing either viewpoint-dependent (*left-right*) or viewpoint-independent (*on-under*) spatial relations between two objects in three different conditions: (1) only in speech without making any gesture (speech-only, e.g., saying *right* while standing still), (2) both in speech and in gesture (redundant-gesture, e.g., saying *right* while gesturing to the right side), and (3) only in gesture with ambiguous speech (complementary-gesture, e.g., saying *here* while gesturing to the right side). The actress spoke two sentences: the first sentence introduced the figure and the ground objects (e.g., "There is a vase and a candle"), whereas the second sentence introduced the relative spatial relation of the figure object (e.g., "The candle is on the *right*"). The actress made the gesture during the second sentence preceding the spatial term. After watching the videos, participants were asked to choose the picture that best depicts the described spatial relation among four alternatives. We are interested in accuracy, reaction times (RTs), and gazing gestures, as measured by the percent number of fixations to the gestural space.

For behavioral performance, we predict that observing gestures will have different effects on listeners' comprehension across viewpoint-dependent vs. viewpoint-independent spatial relations. We expect that participants will have higher accuracies and shorter RTs when spatial information is given bimodally (redundant-gesture) compared to unimodally (speech-only) for *on-under* gestures (Beattie & Shovelton, 1999a, 1999b, 2001, 2002; Holler et al., 2009; Hostetter, 2011). For left-right gestures, however, we reason that observing those gestures from the speaker's egocentric perspective will hinder listeners' comprehension by indexing a mismatching location from the listener/viewer's perspective (Hostetter et al., 2018; Pyers et al., 2015). In this case, we expect that participants will have lower accuracies and longer RTs in redundant-gesture condition compared to speech-only condition for *left-right* gestures. We also expect that accuracies will be the lowest, and RTs will be the longest when the critical spatial information is expressed only in gesture (complementary-gesture) compared to redundant-gesture and speech-only conditions. However, we expect this difference to be higher for viewpoint-dependent spatial relations (left-right) than viewpoint-independent ones (on-under).

For eye gaze, we reason that listeners will allocate more overt visual attention to gestures as their relative information value in relation to the accompanying speech increases. We expect that listeners will fixate to gestures more in the complementary-gesture condition than redundant-gesture condition for both types of spatial relations. However, we expect this effect to be present only during the second sentence, in which the critical spatial information is expressed in speech and gesture. We expect to see no difference during the first sentence that is comparable across conditions and does not express any spatial information. Also, we expect no difference in gaze allocation for left-right vs. on-under gestures across conditions. Last, we predict that gazing at gestures will relate to increased information uptake from gestures. Specifically, we predict that listeners will have higher accuracies and shorter RTs when they gaze at gestures more for both types of spatial relations. However, we expect this effect to be more salient for the complementary-gesture condition than redundant-gesture condition, in which the critical spatial information can be abstracted from both speech and gesture.

4.3. Methods

4.3.1. Participants

We recruited 51 native Turkish-speakers (36 females, M_{age}=21.2 years, SD_{age}=2.61, age range=18-30 years, $M_{\text{education}}$=15.4 years, $SD_{\text{education}}$=2.55) from Koç University, Istanbul in return for either course credit or monetary compensation. All participants were right-handed, had normal or corrected-to-normal vision, normal hearing, and were not taking any psychiatric or neurological medication at least six months before the testing. All participants gave informed consent before the experiment, which was approved by the Institutional Review Panel for Human Subjects of Koç University.

4.3.2. Materials

The experimental task consisted of short video clips of an actress describing different spatial relations between two objects. In these videos, the actress uttered two sentences and sometimes made gestures. The experimental trials consisted of two types of spatial relations: viewpoint-dependent spatial relations (i.e., *left-right*) vs. viewpoint-independent spatial relations (i.e., *on-under*). The filler trials consisted of the spatial relation of "*in."* Each spatial relation was presented in three different conditions: (1) In the *speech-only condition (SO)*, the actress uttered the sentence with the specific spatial term without any hand gesture (e.g., the actress said that "There is a vase and a candle. The candle is on the *right*" while standing still), (2) in the *redundant-gesture condition (RG)*, the actress uttered the sentence with the specific spatial term and made a gesture describing that relationship (e.g., the actress said that "There is a vase and a candle. The candle is on the *right*" while showing her right-side), and (3) in the *complementary-gesture condition (CG)*, the actress uttered the sentence with an adverbial demonstrative (non-specific, ambiguous spatial term) and made a gesture specifying the spatial relation (e.g., the actress said that "There is a vase and a candle. The candle is *here*" while showing her right-side to express that the candle is on the right of the vase.) Figure 4.1 depicts the different conditions in the experiment.

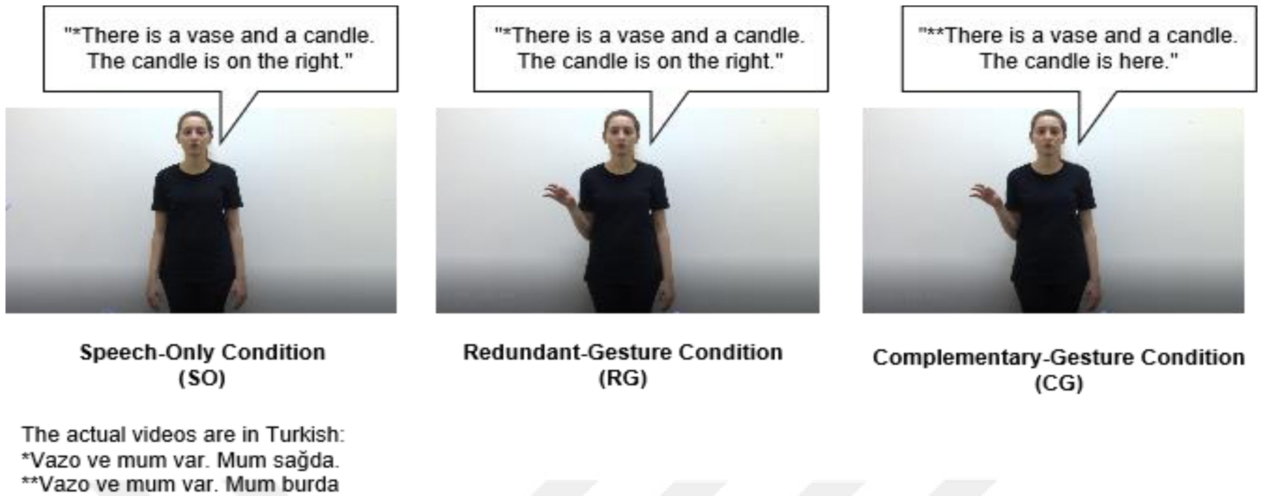


Figure 4.1. Different conditions in the experimental task.

All videos displayed the actress from the head to the knees, appearing in the same starting position (i.e., in the middle of the screen) with hands casually hanging on each side of the body. The actress wore black clothes, and the background was white. In each video, the actress uttered two sentences: the first sentence introduced the ground and the figure objects, respectively (e.g., "There is a vase and a candle"), whereas the second sentence introduced the relative spatial location of the figure object (e.g., "The candle is on the left/right/on/under/in" or "The candle is here"). To make utterances as natural as possible, we used a dataset in which Turkish-speaking adults described relative spatial relations between objects in spontaneous speech (Karadöller, et al., in press). In line with this, the first introduction sentence always presented the ground object first, followed by the figure object (e.g., "there is a vase [GR] and a candle [FIG]"). The second sentence presented the figure object and its relative spatial position without naming the ground object (e.g., "the candle [FIG] is on the right"). The actress made the gestures during the second sentence with her right hand and retracted her arm to the initial position. For left-right gestures, the actress extended her arm up to the side of her left and right arm. For on-under gestures, the actress made the gestures around her torso, right hand showing slightly above to the face for showing "on," and slightly below to the feet for "under." All videos were 6-seconds long. We inserted the audio files of speech-only videos (i.e., the videos in which the actress does not make any gesture) to the videos in which the actress used a gesture. When a speaker makes a gesture, the prosodic prominence of the gesture's referent is altered in speech (Krahmer & Swerts, 2007). To eliminate the possible confounding effects of this altered prosodic prominence among conditions, we recorded speech-only versions for each trial with a gesture component and then swapped the audio files.

4.3.3. Procedure

Participants were tested in a dimly lit soundproof room on a 17" Acer laptop on which a mouse, headphones, and the eye-tracking device were connected. Participants were instructed to watch the short video clips and then choose the picture that best depicted the spatial relation described in the video among four alternatives. In the response screen, they were presented with four pictures showing different spatial relations between the same figure and the ground object. Ground objects were always in the center of the pictures, and the figure object's location in relation to the ground object changed in each picture. They were asked to choose the correct picture by clicking with the mouse as accurately and fast as possible (see Figure 4.2 for a sample response screen). In a canonical trial, participants were presented with a "Get Ready!" screen for 800 ms, followed by a brief preview of the response screen (i.e., picture array) for 500 ms to familiarize participants with the appearances of the figure and the ground objects before they listened to the spatial descriptions. Then, participants saw the fixation cross for 1000 ms, followed by the video for 6000 ms. After the video, the self-timed response screen appeared, and the participants clicked on one of the pictures in the array with the mouse. Figure 4.3 depicts a sample trial.

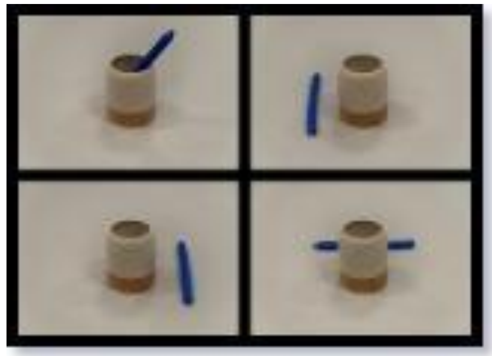


Figure 4.2. A sample response screen for the experimental task.

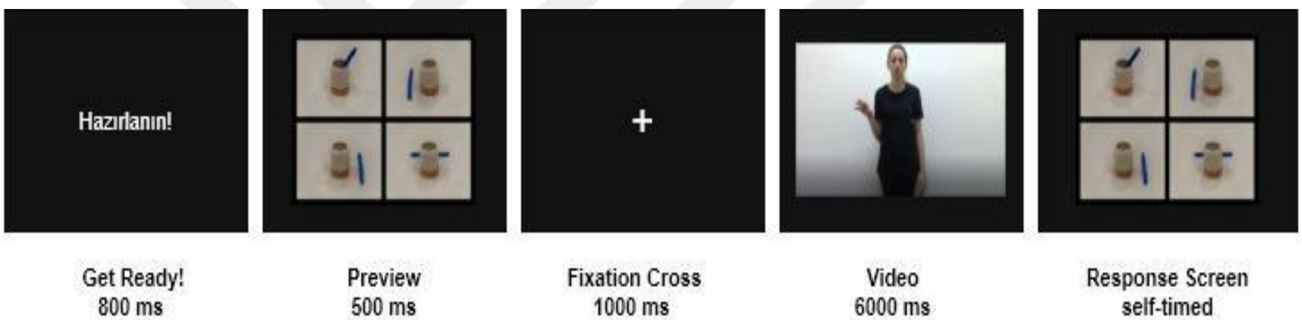


Figure 4.3. A sample trial in the experimental task.

Before starting the experimental task, participants went through familiarization and practice sessions. In the familiarization session, participants saw seven pictures depicting different spatial relations between two different objects and written labels for each spatial relation (i.e., left, right, front, behind, on, under, in; please note that front-behind were not used in the video stimuli, however, these spatial relations were depicted in the pictures as foils during the response screen). Later, the experimenter demonstrated the task with five examples for each spatial relation used in the experiment (i.e., left-right, on-under, in). After familiarization, the participants completed 12 practice trials and were given oral feedback (correct-incorrect) by the experimenter. After these trials, participants have seated approximately 70 cm away from the computer and instructed to limit their movements during the session. We used Tobii Pro X3-120 to monitor eye-movements at a sampling frequency of 120 Hz. Participants underwent a 9-point calibration and validation procedure. The calibration proceeded until the discrepancy between the calibration point, and the participant's eye gaze was < 1°.

The participants completed 150 trials in total, ten trials for each spatial relation and condition: 10*5 spatial relations (right, left, on, under, in) *3 conditions (speech-only, redundant gesture, complementary gesture). Out of 150 trials, 120 were experimental trials, and the rest ("in") were filler items. All trials were presented randomly in E-Prime 3.0 with Tobii extension. The procedure lasted approximately 30 minutes in total.

4.4. Analyses & Results

We only analyzed 120 experimental trials (excluding filler items of *in*) per participant. In total, there were 6120 responses (120*51 participants). For RT, we discarded outliers that were two standard deviations above or below the mean (*N* = 196, 3.2%) and incorrect responses (N = 438, 7.16%). See Table 4.1 for the number of analyzed items in each condition. For eye-tracking data, we discarded trials with a track loss of 25% and higher and trials in which all the fixations were in non-AOI space for the time-of-interest.

4.4.1. Eye-tracking analysis

The eye movements were automatically coded as fixations and saccades using Tobii Pro Lab algorithms with fixation filters. We defined two areas of interest (AOIs): face and gestural space (see Figure 4.4). The gestural space AOI started from the upper neck of the actress and ended at the hand wrists when arms hanged casually on her sides (Drijvers et al., 2019). We created the length of the gestural space AOI based on the x-coordinates that correspond to the furthest point in which a gesture was made on the screen. The size and the location of the AOIs were equal across videos. We were interested in the proportion of the number of fixations to the gestural space and the face AOIs. The fixations that fell outside of these two AOIs (approximately 16% of total fixations) were not analyzed. Only for descriptive purposes, we reported the proportion of the number of fixations and the proportion of the total fixation duration to each AOI relative to the total number of fixations to the entire screen (i.e., AOI + non-AOI spaces) and the total time of the video, respectively (see Table 4.2). However, for analyses, we calculated the proportion of the number of fixations to the gestural space relative to the total number of fixations to the gestural space and the face AOIs (i.e., fixations to gestures divided by the sum of fixations to the gesture + face).

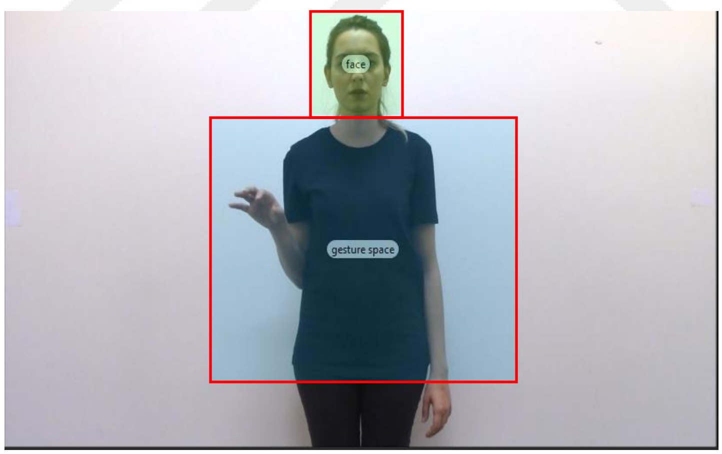


Figure 4.4. Areas of interests: face and gestural space

4.4.2. Mixed-effects analyses

We used multi-level mixed-effects models to incorporate all data variance in our analyses. We used the logistic version of the model for binary outcome variables (i.e., accuracy). All analyses were performed with *lme4* package (Bates, Mächler, Bolker, & Walker, 2014) on R (R Core Team, 2013).

For each analysis, we started with a null model (i.e., the item- and the subject- intercepts model) and added each fixed effect and the interactions between fixed effects one-by-one to the model. To this end, each predictor variable was added consecutively to the model and compared against the previous model containing all the other fixed effects except the one added in the corresponding step (Baayen, Davidson, & Bates, 2008). Since one of our factor variables has three levels (i.e., condition), we used Bonferroni-corrected pairwise comparisons with *emmeans and contrast* functions in *emmeans* package in R (Length, 2020).

First, we asked *(Section 4.4.3.1.),* how task performance differed across conditions and spatial relation types. In Model 1, the outcome variable was accuracy and in Model 2, the outcome variable was RT. The fixed factors included the main effects of the condition and type, and the two-way interaction between them. Second, we asked *(Section 4.4.3.2.)* how gazing at gestures differed across conditions and spatial relation types for the entire video (Model 3), only for the first sentence that was equal across trials and during which there were no gestures (Model 4), and only for the second sentence in which the actress uttered the critical spatial term and made a gesture (Model 5). In these analyses, the outcome variable was the proportion of fixations to the gestural space, and the fixed effects included the main effects of the condition, the type, and the two-way interaction between them. Last, we asked *(Section 4.4.3.3.)*, how gazing at gestures during the second sentence in the videos affected task performance (accuracy in Model 6 and RT in Model 7) across different conditions and spatial relation types. The fixed effects in these models included the main effects of the condition, the type, the proportion of fixations to the gestural space, and the two- and three-way interactions between these variables. Since there were no gestures in the SO condition, we only included the RG and the CG conditions in Models 6 and 7.

4.4.3. Results

4.4.3.1. The experimental task performance

Descriptive statistics about accuracy and RT for each condition can be found in Table 4.1.

Table 4.1. Descriptive statistics for accuracy and RT in general and across conditions.

	N	*Mean*	*SD*
Accuracy			
SO Condition	2040	.95	.22
RG Condition	2040	.96	.18
CG Condition	2040	.87	.34
RT (in msec)			
SO Condition	1906	1089	505
RG Condition	1914	1094	504
CG Condition	1666	1132	545

Note: SD= Standard Deviation, SO=Speech-Only, RG=Redundant-Gesture, CG=Complementary-Gesture

For accuracy (Model 1), there was a main effect of the condition, $\chi^2(2) = 48.25$, $p < .001$. CG condition was associated with lower accuracy by -1.11 ± .18 compared to SO condition ($z = -6.14$, $p < .001$) and by -1.44 ± .19 compared to RG condition, $z = -7.56$, $p < .001$. There was no difference between RG and SO conditions, $z = 1.59$, $p = .34$. The addition of the spatial relation type significantly improved the model, $\chi^2(1) = 12.17$, $p < .001$. Across all conditions, on-under trials were associated with decreased accuracy by -.57 ± .15 compared to left-right trials, $z = -3.69$, $p < .001$. The addition of the interaction between condition and spatial relation type did not improve the model, $\chi^2(1) = 2.03$, $p = .36$. Figure 4.5 shows accuracy across different conditions and spatial relation types.

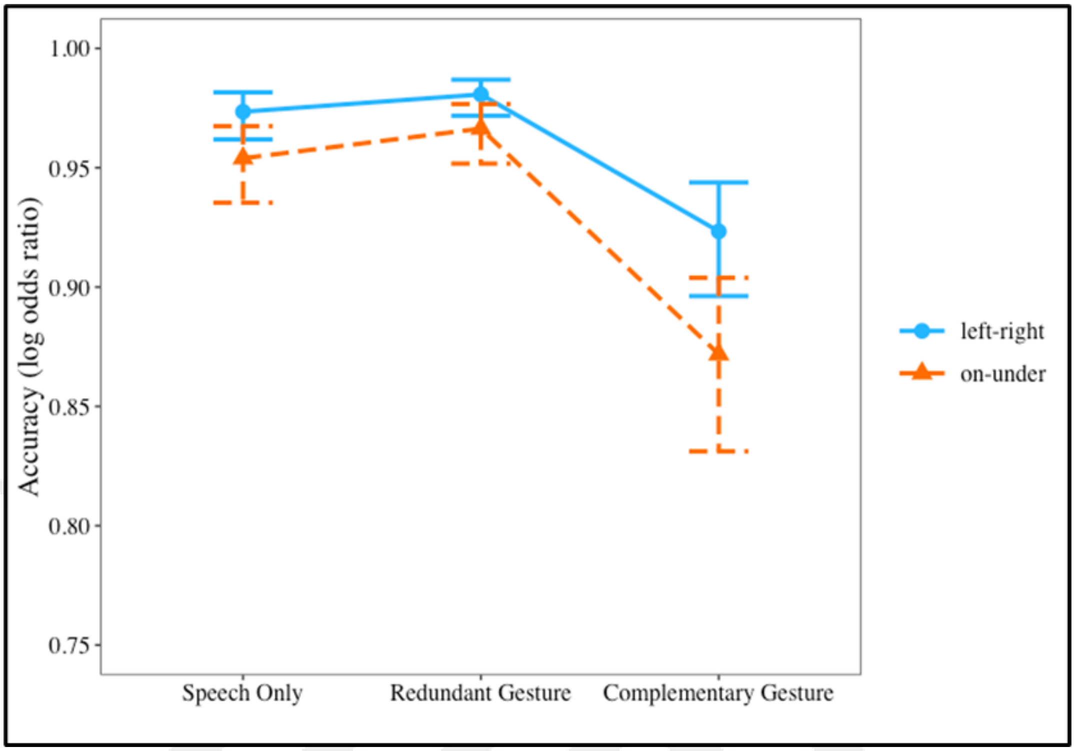


Figure 4.5. Accuracy across different conditions and spatial relation types (Model 1). The brackets represent 95% confidence intervals.

For RT (Model 2), there was no main effect of the condition, $\chi^2(2) = 4.33$, $p = .11$. The addition of the spatial relation type improved the model significantly, $\chi^2(1) = 26.98$, $p < .001$. Across all conditions, on-under trials were associated with increased RTs by 156.28 ± 28.42 compared to left-right trials, $t\ (113) = 5.50$, $p < .001$. There was no interaction between condition and spatial relation type, $\chi^2(2) = .36$, $p = .84$. Figure 4.6 shows RTs across different conditions and spatial relation types.

In sum, we found that participants have lower accuracy in CG condition compared to RG and SO conditions. Participants also have lower accuracy and slower RTs for on-under trials compared to left-right trials. There were no differences between RG and SO conditions for either accuracy or RT.

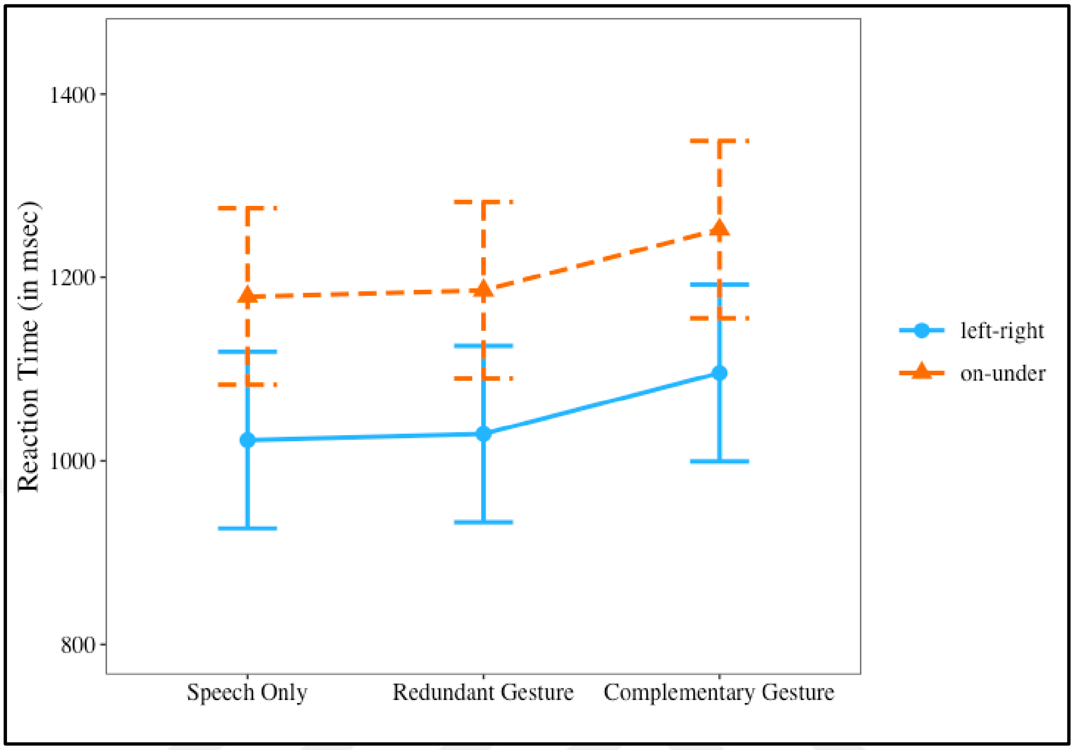


Figure 4.6. RT across different conditions and spatial relation types (Model 2). The brackets represent 95% confidence intervals.

4.4.3.2. Eye-tracking results

Table 4.2 presents descriptive statistics for the proportions of the total number of fixations and the fixation duration to the face and the gestural space AOIs across different conditions. Participants, on average, spent 23% of the total viewing time fixating to the gestural space and around 70% of the total viewing time fixating to the face.

Table 4.2. Descriptive statistics for the percentages of the total fixation duration and the fixation number for the face and the gestural space AOIs across different conditions.

		Total Fixation Duration		Total Fixation Number	
		Mean	*SD*	*Mean*	*SD*
SO Condition	Face %	71.1	26.2	52	24.7
	Gestural Space %	17.6	22.2	27.6	23.2
RG Condition	Face %	68.7	23.9	49.3	23
	Gestural Space %	23.5	22	36.1	23.3
CG Condition	Face %	65.9	23.5	46.2	22.9
	Gestural Space %	27.9	22.2	39.2	23.5

Note: SO=Speech-Only, RG=Redundant-Gesture, CG=Complementary-Gesture

We first analyzed the proportion of fixations to the gestural space throughout the entire video (Model 3; see Figure 4.7). Please note that although there were no gestures in the SO condition, we still measured eye gaze to the gestural space in this condition to have a baseline measure against the other two conditions with gestures. For eye gaze in the entire video, there was a main effect of the condition, $\chi^2(2) = 89.55$, $p < .001$. CG condition was associated with increased proportion of fixations to the gestural space by .12 ± .01 compared to SO condition ($z = 11.28$, $p < .001$) and by .04 ± .01 compared to RG condition, $z = 3.51$, $p < .01$. There was also a difference between RG and SO conditions. RG condition was associated with increased proportions of fixations to the gestural space by .08 ± .01 compared to SO condition, $z = 7.78$, $p < .001$. The addition of the spatial relation type did not improve the model, $\chi^2(1) = .78$, $p = .38$ There was no interaction between condition and spatial relation type, $\chi^2(2) = .36$, $p = .84$.

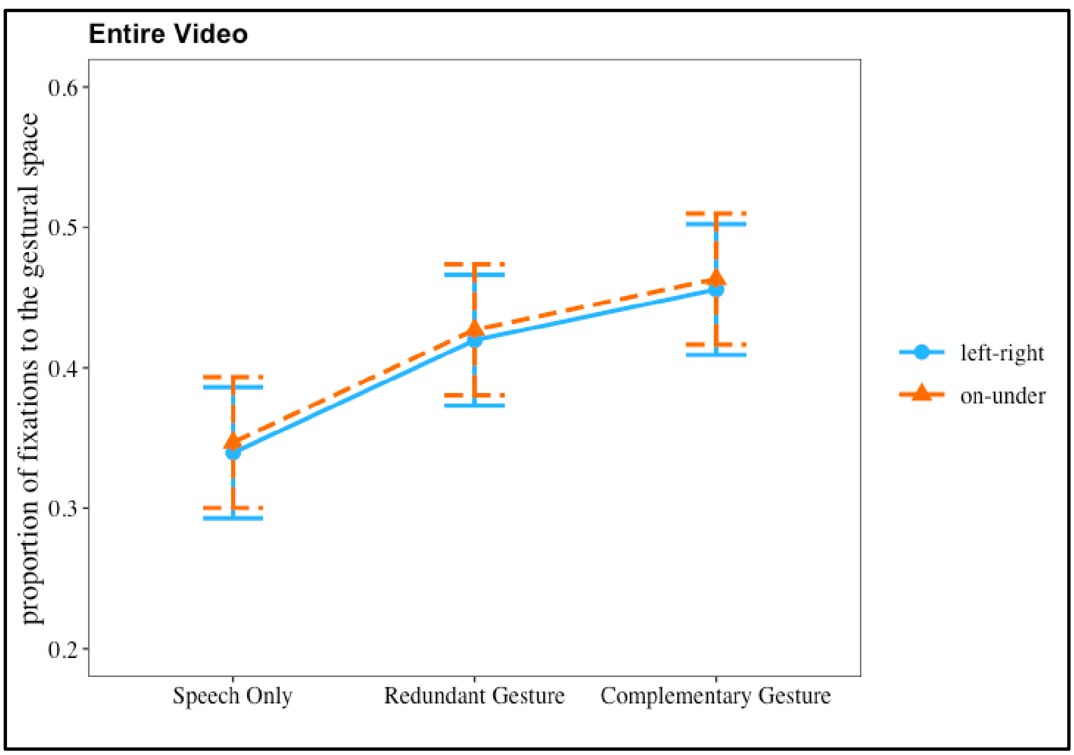


Figure 4.7. The proportion of the number of fixations to the gestural space AOI for the entire video across different conditions and spatial relation types (Model 3). The brackets represent 95% confidence intervals.

We also analyzed the proportion of fixations to the gestural space separately for the first and the second sentences. For the first sentence (i.e., introductory sentence that was equal across trials and during which there was no gestures; Model 4; see Figure 4.8), there was no main effect of the condition, $\chi^2(2) = .75$, $p = .69$. The was no main effect of the spatial relation type, $\chi^2(1) = .41$, $p = 52$. There was no interaction between condition and spatial relation type, $\chi^2(5) = 4.52$, $p = .48$.

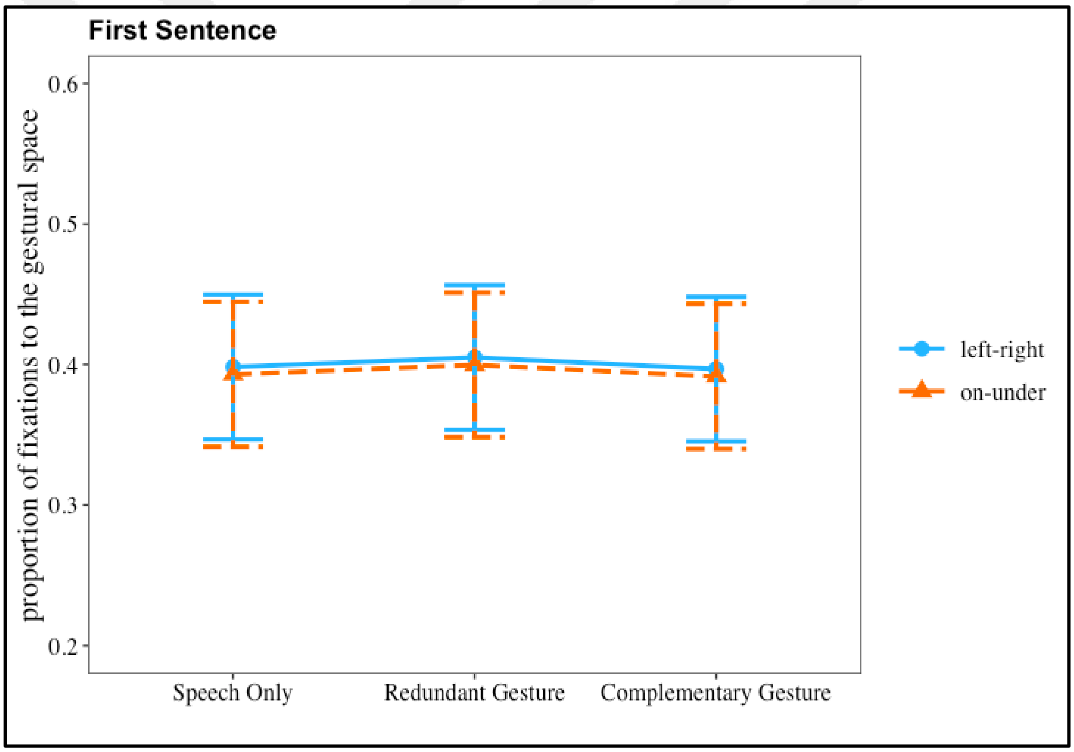


Figure 4.8. The proportion of the number of fixations to the gestural space AOI only for the first sentence across different conditions and spatial relation types (Model 4). The brackets represent 95% confidence intervals.

We then analyzed the proportion of fixations to the gestural space during the second sentence in which the actress uttered the critical spatial term and made a gesture (Model 5; see Figure 4.9). For the second sentence, there was a main effect of the condition, $\chi^2(2) = 120$, $p = < .001$. For both spatial relation types, CG condition was associated with increased proportion of fixations to the gestural space by .20 ± .01 compared to SO condition (z = 14.33, p < .001) and by .06 ± .01 compared to RG condition, z = 4.48, p < .001. The proportion of fixations to gestural space was higher in RG condition by .14 ± .01 compared to SO condition, z = 9.85, p < .001. There was also a main effect of the spatial relation type, $\chi^2(1) = 4.70$, p = < .05. Across all conditions, on-under trials were associated with increased proportions of fixations to the gestural space by .02 ± .01 compared to left-right trials, $t(119) = 2.19$, p < .05. There was no interaction between condition and spatial relation type, $\chi^2(2) = 1.20$, p = .55.

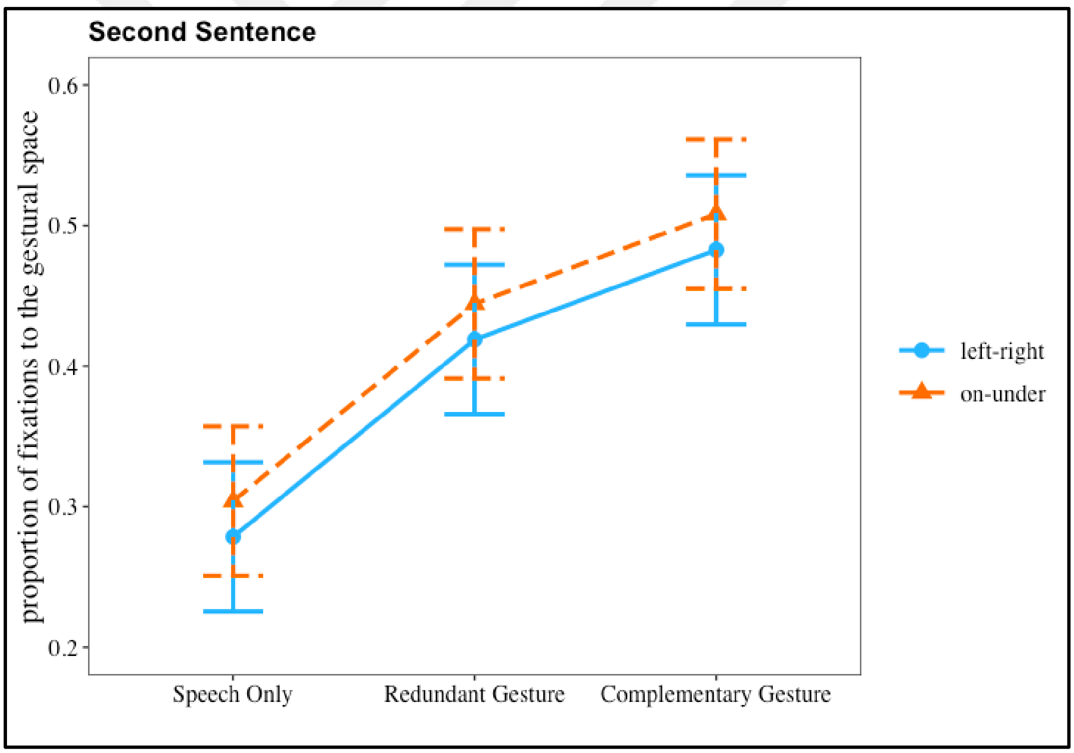


Figure 4.9. The proportion of the number of fixations to the gestural space AOI only for the second sentence across different conditions and spatial relation types (Model 5). The brackets represent 95% confidence intervals.

In sum, we found that listeners fixated to gestures more in CG condition compared to RG condition when the entire video analyzed. Yet, this effect is mainly driven by the second part of the video (i.e., the second sentence including the critical spatial term and gestures). During the second sentence, participants fixated to gestures more in CG condition compared to RG condition. Also, they fixated to the gestural space for on-under trials compared to left-right trials.

4.4.3.3. Eye tracking and task performance results

We tested how gazing at gestures predicted participants' accuracy across different conditions and spatial relation types (Model 6; see Figure 4.10). There was no main effect of gazing gestures, $\chi^2(1) = .61$, $p = .44$. There was a main effect of the condition, $\chi^2(1) = 39.32$, $p = < .001$. Across both spatial relation types, CG condition was associated with lower accuracy by -1.43 ± .20 compared to RG condition, $z = -7.44$, $p < .001$. There was also a main effect of the spatial relation type, $\chi^2(1) = 8.18$, $p = < .01$. Across both conditions, on-under trials were associated with lower accuracies by -.57 ± .19 compared to left-right trials, $z = -3.04$, $p < .001$. There were no two-way interactions between gazing at gestures and condition ($\chi^2(1) = 1.41$, $p = .23$), between gazing at gestures and spatial relation type ($\chi^2(1) = 1.67$, $p = .20$), and between condition and spatial relation type, $\chi^2(1) = 1.73$, $p = .19$. There was also no three-way interaction between gazing at gestures, condition, and spatial relation type, $\chi^2(4) = 5.07$, $p = .28$.

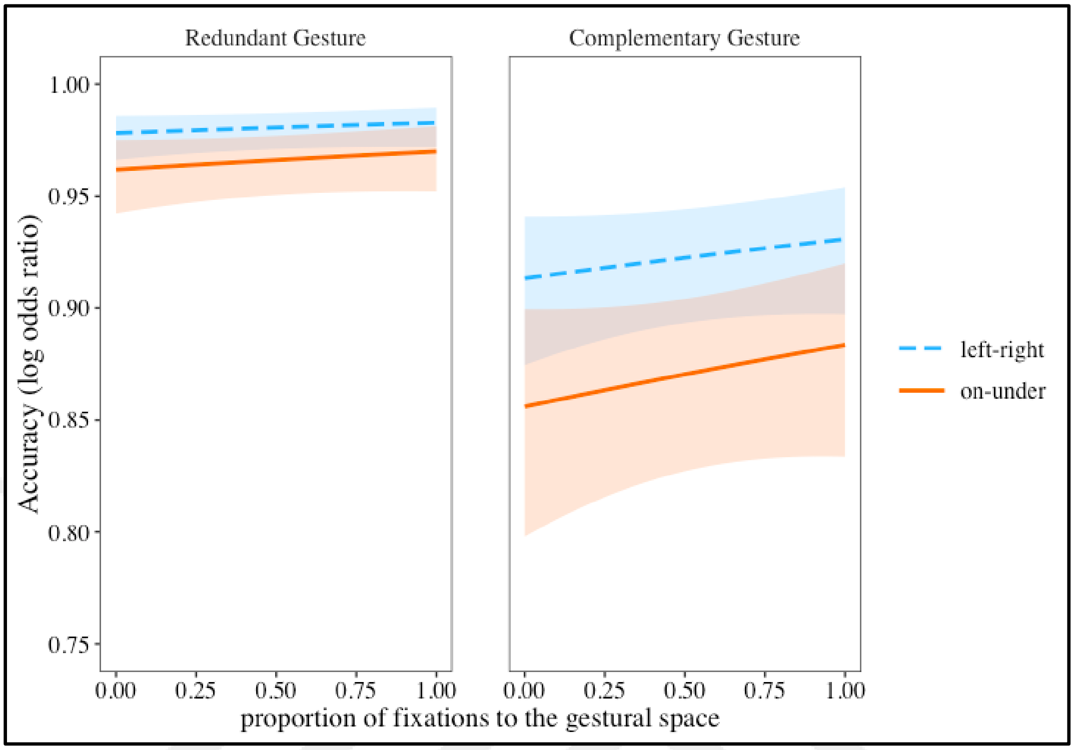


Figure 4.10. Accuracy as a function of gesture gazing across different conditions and spatial relation types (Model 6). The x-axis shows the proportion of fixations to the gestural space. The hues around the lines represent 95% confidence intervals.

Last, we tested how gazing at gestures relates to RTs across different conditions and spatial relation types (Model 7, see Figure 4.11). There was a main effect of gazing at gestures; however, this effect was qualified by an interaction with spatial relation type, $\chi^2(1) = 15.53$, $p = < .001$. Slope analyses showed that gazing at gestures more was associated with increased RTs by 113.90 ± 38.57 only for on-under trials, $t = 2.95$, $p < .001$. There was no statistical relation between gazing gestures and RTs for left-right trials, $t = 1.44$, $p = .15$. There was no main effect of condition, $\chi^2(1) = 2.63$, $p = .11$. There was a main effect of spatial relation type, $\chi^2(1) = 13.84$, $p = < .001$. Across both conditions, on-under trials were associated with increased RTs by 145.44 ± 37.43 compared to left-right trials, $t(75) = 3.89$, $p < .001$. There were no two-way interactions between gazing at gestures and condition ($\chi^2(1) = 2.71$, $p = .10$) and between condition and spatial relation type, $\chi^2(1) = .19$, $p = < .67$. There was no three-way interaction between gazing at gestures, condition, and spatial relation type, $\chi^2(3) = 4.99$, $p = .17$. In summary, we found that fixating to gestures was associated with increased RTs only for on-under trials.

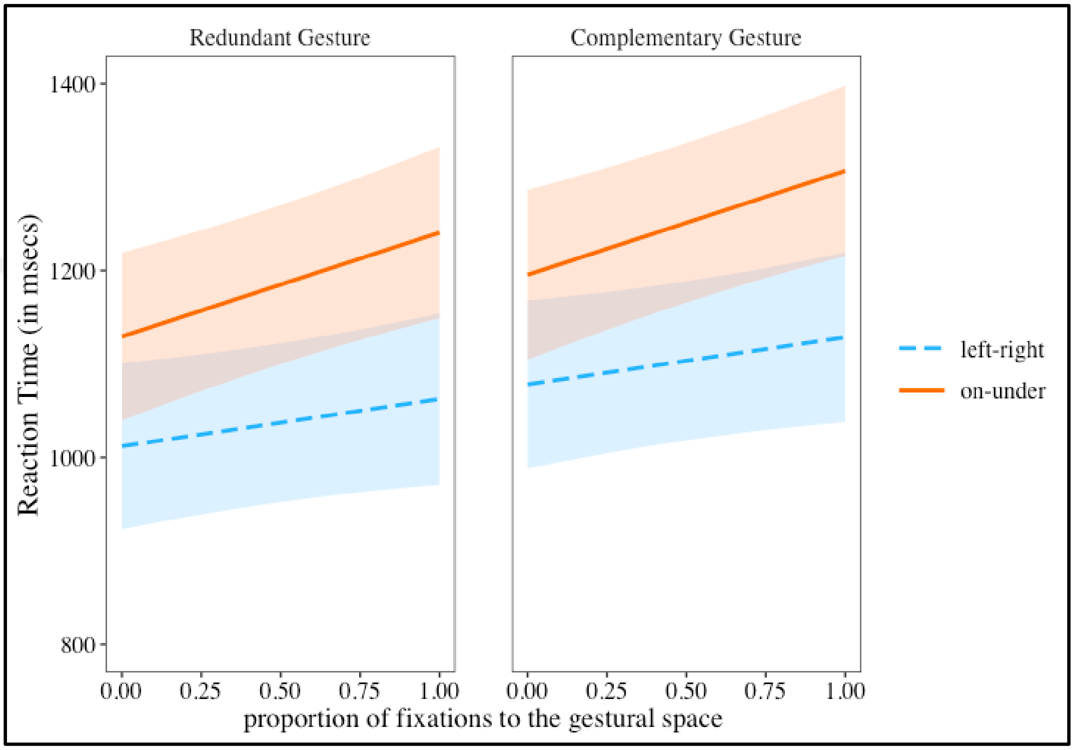


Figure 4.11. RT as a function of gesture gazing across different conditions and spatial relation types (Model 7). The x-axis shows the percent number of fixations to the gestural space. The hues around the lines represent 95% confidence intervals.

4.5. Discussion

The current study investigated whether and how (1) listeners allocated overt visual attention to gestures as a function of the information value of those gestures in relation to the accompanying speech, and (2) gazing at gestures related to how much listeners benefitted from observing gestures during the comprehension of viewpoint-dependent (left-right) and viewpoint-independent (on-under) spatial relations. We expected that listeners would attend to gestures more as their relative information value increased and gazing at gestures would be related to increased comprehension.

Results showed that on the comprehension task, observing left-right or on-under gestures along with informative speech did not change listeners' comprehension compared to when they did not see gestures. Listeners' comprehension got hindered when the only source of information was gestures. The eye-tracking findings demonstrated that listeners allocated more overt visual attention to gestures that complemented spoken deictic constructions ("*here*") than gestures that were redundant with the accompanying speech. We failed to find any evidence of gazing at gestures related to increased information uptake from those gestures.

4.5.1. Task performance: How does observing gestures affect listeners' comprehension of different types of spatial relations?

We employed an offline comprehension paradigm in which participants were presented with unimodal or bimodal (gesture + speech) utterances (e.g., Holler et al., 2019; Hostetter et al., 2018; Kelly, Barr, Church, & Lynch, 1999) when comprehending spatial relations between objects. We predicted that observing gestures along with speech would facilitate listeners' comprehension for viewpoint-independent spatial relations (on-under; Beattie & Shovelton, 1999a, 1999b, 2001; Holler et al., 2009; Hostetter et al., 2018). For viewpoint-dependent spatial relations (left-right), we expected that seeing gestures would hinder comprehension as left-right gestures that are executed from the speaker's own egocentric perspective would express a mismatching location from the speaker's perspective (Gullberg & Kita, 2009; Hostetter et al., 2018; Pyers et al., 2015). However, our results showed that accuracies and RTs did not differ between bimodal (speech + gesture, i.e., redundant-gesture) and unimodal (speech-only) encodings for both types of spatial relations.

Gestures communicate meaning and observing gestures along with speech facilitates comprehension and learning (see Dargue et al., 2019 and Hostetter, 2011 for meta-analyses, e.g., Beattie & Shovelton, 2005; Church et al., 2001; Driskell & Radtke, 2003; Holler et al., 2009; Kelly et al., 1999; Valenzeno, Alibali, & Klatzky, 2003). Yet, many studies reported null effects and showed that observing gestures along with speech did not affect listeners' comprehension and learning (e.g., Goldin-Meadow & Singer, 2003; Kelly & Goldsmith, 2004; Krauss, Dushay, Chen, & Rauscher, 1995; McNeil et al., 2000). Moreover, Hostetter (2011) showed that observing gestures benefit comprehension more when gestures express non-redundant information to the accompanying speech, and when listeners are verbally less proficient (children vs. adults). In line with this, research suggests that gestures are particularly helpful in challenging communication situations such as in degraded speech comprehension or for non-native listeners (e.g., Drijvers et al., 2018; Holle et al., 2010; Sueyoshi & Hardison, 2005). However, in the current study, we tested verbally competent adults who were native listeners in clear speech comprehension. The average accuracies for redundant-gesture (96%) and speech-only (95%) conditions showed a ceiling effect, suggesting that participants found the task easy, and there may not be room for gestural enhancement.

For left-right gestures, contrary to our predictions, listeners' comprehension did not get hindered by observing gestures. As in the case of on-under gestures, we did not find a difference between speech + gesture (i.e., redundant-gesture) and speech-only conditions for left-right directionality. However, it is important to note that our participants were explicitly told that the speaker is making gestures from her own egocentric perspective, and thus, they needed to take the speaker's perspective. Earlier studies did not provide clear instructions on how to interpret the speaker's egocentric gestures that express complex spatial arrays with geometric shapes (e.g., Hostetter et al., 2018). The finding that observing left-right gestures did not hinder comprehension might be due to participants having learned and familiarized with viewpoint alignment during the practice session.

We found that participants had lower accuracies and longer RTs for on-under compared to left-right across all conditions. Earlier research suggested that, in general, viewpoint-dependent spatial relations (left-right) are harder compared to viewpoint-independent ones (on-under) as they require viewpoint alignment (Galati et al., 2013; Keysar et al., 2000). Developmental research also showed that the acquisition and comprehension of left-right directionality come at a later timepoint in development and follow a more protracted trajectory compared to on-under for both speaking and signing children (e.g., Karadöller et al., in press; Turan, Kobaş, & Göksun, in press). There might be several reasons for the current finding. First, our participants were explained and practiced how to align with the speaker's egocentric left-right gestures. Although they were given practice for all types of spatial relations, they might benefit more from training for left-right directionality as they initially found it more difficult than on-under trials and could have paid extra attention to them. Second, they might find it challenging to distinguish on vs. under gestures as they were executed in a smaller gesture space with subtle differences in hand movements as opposed to left-right gestures that were executed in a larger gesture space and indexed the location more broadly with gross hand/arm movements. Third, on-under spatial relations in our stimuli set included contact between objects, whereas objects always stand apart from each other in left-right spatial relations. Especially for "under" relations, in most of the pictures, only a portion of the figure object was visible as it was occluded by the ground object. This might hinder participants' performance as they had problems in distinguishing objects. The use of computer-generated spatial arrays of on-under relations with no contact between objects (e.g., a pencil standing above a vase with no contact) might have yielded different results. Fourth, people might have certain beliefs on which figure-ground object pairs afford on-under spatial relations. We generally tend to think of the bigger object as the ground when conceptualizing on-under relations with contact. For example, we tend to conceptualize that "the pencil is on the luggage" instead of "the luggage is under the pen." Although the ground objects were always bigger compared to figure objects in our stimuli set, participants might still get confused on whether and how figure-ground object pairs afford on-under relations compared to left-right relations that is devoid of this problem. Future work can pay attention to these stimuli characteristics and the use of space for spatial relations while designing their studies.

In line with our prediction, results showed that participants were less accurate when speech is ambiguous, and the only source of information was gestures (complementary-gesture) compared to the other two conditions in which speech expressed the spatial relation (redundant-gesture and speech-only) for both types of spatial relations. This suggests that participants in our study could abstract spatial information more readily from speech, and when the spoken channel is under-informative, their comprehension declined. For left-right gestures, seeing gestures that express a mismatching location from the listeners' own egocentric perspective might create an extra cognitive burden when the gestures are the only source of information than other conditions in which left-right directionality can already be abstracted from the speech. As in the case of on-under gestures, listeners might find it difficult to distinguish on vs. under gestures that are executed with subtle kinematic differences in a very small gesture space, especially when those gestures were the only source of information to comprehend the speaker's message successfully.

4.5.2. Eye gaze: How do listeners allocate visual attention to complementary vs. redundant gestures for different types of spatial relations?

Participants looked at the gestural space approximately 25% of the total viewing time when there was a gesture (excluding speech-only condition), averaged across conditions and types of spatial relations (see Table 4.2). This proportion is higher compared to previous research (e.g., 0.5% in Gullberg & Holmqvitz, 1999; 8% in Gullberg & Kita, 2009; 9.3% in Yeo & Alibali, 2017). However, earlier studies that report lower proportions of fixations to gestures measured eye-movements either in live interactions (e.g., Gullberg & Holmqvitz, 1999) or in videos with speakers selected from the corpus (i.e., the videos were not acted, e.g., Gullberg & Kita, 2009; but see Yeo & Alibali, 2017). The communication context in the current study was less natural and less social as listeners passively listened to speakers who uttered sentences and made gestures in a controlled manner (gestures were scripted, the actress showed no facial expressions and stood still across all trials) without interacting with her (i.e., in video-presentations). This might be one reason why participants in our study looked at the speaker's face less and fixated to other cues. One advantage of such controlled settings is the experimental control over speech content and gesture kinematics. Yet, interactional settings are important, and future studies can investigate how listeners attend to gestures that complement spoken deictics when they are in dialogue.

Another reason for the increased proportions of fixations to gestures in the current study compared to earlier research can emerge from cross-language differences. Previous research suggested differences in gesture use and processing across languages (see Kita, 2009 for review). Mediterranean cultures are usually relatively higher-gesture cultures. For example, Turkish speakers use more gestures than Dutch speakers (Azar, Backus, & Özyürek, 2020) and Italian speakers gesture more frequently compared to British English speakers (Cavicchio & Kita, 2013).

Moreover, Italian listeners benefit more from the speaker's gestures when decoding objects than English listeners (Graham & Argyle, 1975). Earlier eye-tracking research tested low-gesture languages (e.g., English in Beattie et al., 2010; Yeo & Alibali, 2017; Swedish in Gullberg & Holmqvitz, 1999, 2006; Dutch in Drijvers et al., 2020; Gullberg & Kita, 2009). Listeners might be more attentive to gestures in Turkish, which is a high-gesture language. Visual attention to gestures across different languages is an important open area for further investigation.

In line with the main prediction of the current study, our results showed that listeners' overt visual attention to gestures changed as a function of the relative information value of gestures in relation to the accompanying speech. Specifically, we found that listeners gazed at gestures more when they complemented ambiguous speech than when they expressed redundant information to the accompanying speech, for both types of spatial relations. More importantly, this effect was only present during the second sentence in which the critical spatial term was uttered, and a gesture was made. There was no difference in eye gazing across conditions during the first sentence, in which the actress only introduced the figure and the ground objects without making any gesture. This finding is in line with earlier research, suggesting that listeners attend to and benefit from gestures more under certain circumstances, particularly when gestures have greater importance that emerge from either internal (e.g., native vs. non-native listeners, Dahl & Ludvigsen, 2014; Drijvers & Özyürek, 2018; Macedonia & Kriegstein, 2012; Sueyoshi & Hardison, 2005; individual differences in cognitive resources, see Özer & Göksun, 2020a for review; Özer & Göksun, 2020b; Schubotz, Holler, Drijvers, & Özyürek, 2020) or external factors (e.g., spatial vs. non-spatial content, Alibali, 2005; Lavergne & Kimura, 1987; Lausberg & Kita, 2003; Rauscher et al., 1996; clear vs. degraded speech comprehension, Drijvers et al., 2017, 2018; Schubotz et al., 2020; Holle et al., 2010). One key factor in indexing the relative importance of gestures is the complementary distribution of information across speech and gesture. A meta-analysis (Hostetter, 2011) showed that non-redundant gestures that convey additional information to the accompanying speech are more communicative and benefit listeners' comprehension more than redundant gestures that express overlapping information with the accompanying speech. With the help of eye-tracking methodology, the current results extend these findings by demonstrating that semantic redundancy between speech and gesture also guides visual attention to gestures. Our results align with earlier eye-tracking research, showing that when gestures provide a more useful channel of information, listeners allocate more overt visual attention to those gestures as in the case of communicating in a non-native language (e.g., Drijvers et al., 2020) or when gestures resolve speech ambiguity (e.g., Yeo & Alibali, 2017).

The current study is among the first to investigate whether and how listeners visually attend to gestures in the context of spoken deictic constructions in a controlled way. When describing spatial information, speakers tend to convey important spatial information in their gestures and frequently use deictic constructions in speech (such as *this way, like this, here, this big)* that is complemented with gestures (e.g., Beattie & Shovelton, 2006; Emmorey & Casey, 2001; Holler & Stevens, 2007). Speakers refer to their own gestures with such deictic constructions, which in turn might direct listeners' visual attention. Our results demonstrate that the use of such deictic constructions (verbal pointers) signal the importance of gestures and directs listeners' visual attention to those gestures.

We predicted no difference in gaze allocation between left-right and on-under gestures as they were all executed in the central space with comparable kinematic features of equal velocities and without holds (Gullberg & Kita, 2009; Nobe et al., 2000). The actress made on-under gestures on her torso and left-right gestures in very close proximity to her torso, all of which were in the central gesture space (see Beattie et al., 2010; Gullberg, 2003; McNeill, 1992 for discussions on gesture space). The results showed that listeners gazed at on-under gestures more compared to left-right gestures across both redundant- and complementary-gesture conditions during the second sentence. This finding might emerge from differences in hand and arm movements across left-right and on-under gestures mentioned above. Although both types of gestures were made in the central gesture space, on-under gestures were executed in a smaller space with a subtly different hand movement. Listeners might have looked at on-under gestures more as it was harder to distinguish these on vs. under gestures. Left-right gestures, on the other hand, were executed in a larger space with gross hand/arm movements in an indexical manner to show left vs. right space. Listeners could have abstracted left-right information even from shoulder/upper-arm movements without necessarily fixating to the hand.

4.5.3. Eye gaze and comprehension: Whether gazing at gestures relate to how much listeners benefit from observing gestures?

We predicted that allocating more overt visual attention to gestures would be positively related to how much listeners benefit from observing gestures during the comprehension of both types of spatial relations. However, the results did not support this prediction. We found that gazing at gestures did not relate to gestural enhancement of comprehension. This finding is in line with earlier work (Beattie et al., 2010; Gulberg & Kita, 2009; but see Drijvers et al., 2019) and suggests that information conveyed through gestures can be abstracted peripherally without necessarily fixating to them. Gestures in the current study were indexical ones that represent the location in space. However, gestures that express fine-grained details with finger movements (such as gestures representing the objects' features) might demonstrate a different set of results. These gestures might attract more overt visual attention and gazing at those kinds of gestures might predict how much listeners benefit from them. We found that gazing at on-under gestures resulted in increased RTs across both redundant- and complementary-gesture conditions. As discussed above, we believe that this might emerge from the properties of on-under gestures in the current study. Listeners fixated to on-under gestures more as it was harder to distinguish them, which would eventually lead to increased RTs as opposed to left-right gestures that can be abstracted from gross arm movements.

One important area that should be investigated in further research is the mechanisms governing individual differences in visual attention to gestures. Speakers and listeners exhibit variation in how frequently they use gestures, how they attend to gestures, and how much they benefit from using and observing gestures during communication and thinking dependent upon individual differences in cognitive skills (e.g., Chu, Meyer, Foulkes, & Kita, 2014; Özer & Göksun, 2020b; Schubotz et al., 2020; Wu & Coulson, 2014; see Özer & Göksun, 2020a for review). It is unknown whether and how individual differences in cognitive abilities relate to how listeners allocate visual attention to gestures. For example, people who have high spatial skills coupled with low verbal abilities might rely more on gestures (Hostetter & Alibali, 2011; Özer & Göksun, 2020b). Do these people allocate more overt visual attention to gestures compared to people with lower spatial skills, and how it results in how much information listeners uptake from those gestures? It is also important to uncover how individual differences in cognitive resources interact with different external factors such as speech content (e.g., spatial vs. non-spatial, Hostetter, 2011), complexity (e.g., objects with low vs. high verbal codability, Kita & Davies, 2009), or relative information value of gestures (e.g., complementary vs. redundant) in determining visual attention to gestures.

4.6. Conclusion

This study investigated listeners' visual attention to gestures as a function of the relative information value of gestures in relation to the accompanying speech. We demonstrated that listeners allocated more overt visual attention to gestures when they complement accompanying speech than when they expressed redundant information with the speech. However, direct visual attention to gestures did not relate to how much listeners uptake information from them. These results suggest that the perceived information value of gestures as referred by the spoken deictic constructions might direct listeners' attention to gestures; however, listeners can glean visual semantics conveyed through gestures peripherally without fixating to them.

CHAPTER 5

GENERAL DISCUSSION

This dissertation aimed to investigate whether and how listeners' attention to and processing of gestures present variation across individuals and situations. The short answer is: yes. How listeners employ gestures during comprehension exhibits variation depending on internal factors such as individuals' cognitive dispositions and external factors such as the perceived communicative value of gestures across different contexts. In this dissertation, I demonstrated that individual differences in visual-spatial abilities, and the semantic overlap between speech and gesture affect listeners' attention to gestures. Yet, as is often the case in research, the picture is a little more complicated than that. Many factors interact with each other to yield different patterns in these processes. Below I summarize and discuss the core findings and then suggest areas open to further investigation.

5.1. Summary and the discussion of the main findings

In **Chapter 2,** I reviewed the evidence on inter-individual variation in gesture use and processing to understand the field's current state. I addressed evidence on how individual differences in cognitive skills related to (1) *gesture use*: how much speakers use spontaneous gestures and to what extent they benefit from using gestures during speaking, task solving, and learning, and (2) *gesture processing*: to what extent listeners attend to and process gestures and how much they benefit from observing gestures during comprehension and learning from a *functionalist* perspective. Gestures serve many functions for speakers and listeners across different contexts (Cook & Fenn, 2017; Kita et al., 2017; Novack & Goldin-Meadow, 2017). These functions interact with individuals' cognitive dispositions and inform us about the possible cognitive mechanisms governing these processes. First, gestures serve communicative functions as they provide an alternative channel of expression between interlocutors. Second, gestures help speakers and listeners maintain, manipulate, and package visual-spatial information and reduce cognitive load when speaking and thinking (e.g., Kita et al., 2017; Marstaller & Burianova, 2013). Research in the gesture production literature suggests that speakers with lower communicative proficiencies (e.g., non-native vs. native speakers, lower vs. higher verbal skills, or children vs. adults) benefit more from using gestures during speaking (Colletta et al., 2010; Gillespie et al., 2014; Gullberg, 2010; Hostetter & Alibali, 2007; Smithson et al., 2011; Wermelinger et al., 2020). Additionally, speakers with lower visual-spatial skills benefit more from using gestures when speaking and thinking about spatial relations, such as when they describe how to solve mental rotations or when they encode routes (e.g., Göksun et al., 2013; Pouw et al., 2016; So et al., 2015; Trafton et al., 2016). In Chapter 2, I argued that speakers might employ gestures as a compensatory tool to manage their limited cognitive abilities during speaking and thinking. This does not entail that gestures replace already impaired cognitive abilities; instead, I suggest that gestures might help people manage their cognitive resources when they are intact (see Chapter 2 for discussion). However, higher use of gestures by speakers with lower cognitive resources might not be due to the active strategy of compensation. Instead, people with lower cognitive resources might employ gestures more because of their adaptive communication skills, such as relying more on verbal resources than gestural ones. That is, differences in gesture use and processing based on the use of cognitive resources can be related to individuals' employment of different strategies. Although I found this plausible, it is hard to distinguish these two accounts as data in the current literature is correlational.

What happens on the listener's side? Chapter 2 revealed little research on the comprehension domain, and it is unknown how individual differences in different cognitive abilities relate to listeners' attention to and processing of gestures during language comprehension. Based on this, as a next step, I investigated how individual differences in cognitive abilities relate to listeners' attention to gestures during language comprehension.

In **Chapter 3**, I examined the role of visual-spatial and verbal cognitive resources in online processing of concurrent gestural vs. spoken information in a mismatch paradigm (Kelly et al., 2010). First, I replicated earlier findings on the bidirectional relationship between speech and gesture during comprehension in a different language, Turkish. The results showed that observing gestures and speech that expressed mismatching information hindered comprehension compared to the baseline condition in which speech and gesture expressed congruent information. Second, I demonstrated that this is not a monolithic process and individual differences exist in how listeners process concurrent gestural vs. spoken expressions. The results showed that when the visual-spatial cognitive load was high (i.e., when gestures expressed different and mismatching information to the preceding action representation), people with higher visual-spatial abilities were better at processing gestures. In contrast, people with higher verbal abilities were better at processing speech when the verbal cognitive load was high (i.e., when speech expressed mismatching information to the preceding action representation). I concluded that visual-spatial cognitive resources are critical to process visual semantic information conveyed by gestures, and people with higher visual-spatial skills can process gestures better, particularly when gestures expressed additional semantic information. This is in line with an earlier finding that showed observing gestures hindered comprehension when listeners' visual-spatial cognitive resources were taxed (Hostetter et al., 2018). These findings suggest that visual-spatial cognitive resources are critical for processing gestures, and people with higher visual-spatial capacities might benefit more from observing gestures compared to people with lower capacities for whom the additional processing load of observing gestures might be demanding (Aldugom et al., 2020; Kalyuga et al., 2003; Kartalkanat & Göksun, 2020; Kelly & Goldsmith, 2014; Post et al., 2013; Sweller, 2005; Wu & Coulson, 2014). Indeed, a recent study showed that college students with higher visual skills benefitted more from observing gestures along with speech (compared to speech-only) than students with lower skills during math learning (Aldugom et al., 2020). Observing gestures also facilitated children's narrative comprehension more for children with higher verbal skills (Kartalkanat & Göksun, 2019).

There appear to be differential patterns governing how lower cognitive resources relate to using vs. processing gestures. The current evidence suggests that multimodal encodings (speech + gesture) help speakers with lower cognitive resources when using gestures, whereas hinder listeners with lower cognitive resources when observing gestures. How can these two different sets of findings be reconciled? One possibility for this distinction might be related to how using vs. seeing gestures affect cognitive load. Gestures project speakers' internal representations to external space and provide visual and proprioceptive feedback for speakers that can be used to maintain and manipulate task-related visual-spatial information. This way, gestures reduce speakers' cognitive load when speaking and thinking (e.g., Goldin-Meadow et al., 2001; Marstaller & Burianova, 2013; Pouw et al., 2014; Wesp et al., 2001). Speakers have agency over their gestures that might help them offload part of the internal cognitive load to an external space. However, gestures do not provide visual or proprioceptive feedback to listeners as they do not have agency over manual action. Gestures provide an alternative source of information in visual-spatial modality to listeners. Processing the extra visual information conveyed by gestures and speech requires coordination between multiple sources of information. The additional processing load of gestures might be too demanding for people with lower cognitive capacities. As a result, different results appear when one examines how visual-spatial resources can associate with gesture use or processing.

How and to what extent listeners attend to speakers' gestures and benefit from observing them during comprehension and learning also vary across situations (Dargue et al., 2019; Holler et al., 2009; Holler et al., 2014; Hostetter, 2011; Nagels et al., 2015). Observing gestures facilitate listeners' comprehension more in certain circumstances, in which gestures provide a critical source of information for comprehension such as when they express spatial vs. non-spatial information (e.g., Beattie & Shovelton, 2001; Hostetter, 2011) or in degraded vs. clear speech comprehension (e.g., Drijvers et al., 2017, 2018). Another external factor that modulates gesture's communicative value is the semantic overlap between speech and gesture.

In **Chapter 4**, I asked how the relative information value of gestures as indexed by the accompanying speech context (ambiguous speech coupled with spoken deictic constructions vs. unambiguous speech) affect how listeners attend to gestures by employing both behavioral and eye-tracking methodologies. Eye-tracking provides a good method to investigate how listeners' allocation of attention to multiple communicative cues is reflected in gaze behavior (Huettig, Rommers, & Meyer, 2011; Posner, 2016). I asked whether and how (1) listeners attended to gestures differently when they expressed complementary vs. redundant information in relation to the accompanying speech, and (2) individual differences in overt visual attention to gestures predicted how much listeners gleaned information from those gestures during the comprehension of spatial language. I found that listeners fixated to gestures more when they complemented spoken deictic constructions (e.g., saying "*here*" complemented with a gesture showing the right side) compared to when they expressed redundant information to the accompanying speech (e.g., saying "*right*" and gesturing to the right side). The use of spoken deictics in speech might refer to the speaker's gestures and increase the perceived communicative value of gestures, which results in increased attention allocation by listeners. This finding suggests that gestures that add semantic information to the spoken utterances are perceived to be more communicative by listeners. The results also demonstrated that overt gaze allocation to gestures did not predict gestural enhancement of comprehension. I concluded that listeners attend to gestures more as the relative information and communicative value of gestures increase. However, gestures can be processed peripherally and do not necessitate direct visual attention.

This dissertation contributes to the current literature by demonstrating that the so-called beneficial effects of gestures are not monolithic and show variation across individuals and situations. The gesture theories and current practices in the gesture literature mostly downplayed the role of individual differences and treated them as error variance. This leads to generalized assumptions that gesturing and the communicative and cognitive benefits of gestures are invariant across individuals and different contexts. However, as evidenced in this dissertation, speakers and listeners employ gestures differently across different contexts that interact with their own cognitive dispositions. This dissertation's main theoretical contribution is to pave the way for taking all these different internal and external factors into consideration to have a full picture of when, why, and for whom gestures exhibit their supposed roles. Based on this theoretical contribution, this dissertation also entails a practical implication. This line of research would inform several areas in which gestures are incorporated to foster learning, such as classrooms. Teachers are generally encouraged to use gestures as they teach. However, teachers' gestures might not benefit all students equally. Learners with lower cognitive abilities might even get hindered by those gestures. The instructional programs should be tailored according to the cognitive dispositions and needs of the learners for optimal learning outcomes.

5.2. Future directions

Many questions remain unanswered in the current state of the art. In this section, I suggest several lines of research that can be addressed in future work.

First, future research should bridge different internal and external factors to have a full picture of *why, when,* and *for whom* gestures are beneficial. These factors might interact with each other to yield different patterns. The diverse set of results present in the literature on how individual differences in cognitive resources relate to gesture employment might emerge from the use of different paradigms in different contexts (e.g., spatial vs. non-spatial tasks, see Chapter 2 for discussion). For example, given that gestures are more adept at expressing visual-spatial information (Beattie & Shovelton, 1999a, 2001; Hostetter, 2011), do speakers with lower visual-spatial skills gesture more only when speaking and thinking about spatial information compared to non-spatial information? Similarly, do listeners with higher visual-spatial skills benefit from observing gestures only in a spatial context (Aldugom et al., 2020). I believe that investigating how individual differences in cognitive resources relate to gesture use and processing across different communication contexts would provide a better picture of when and for whom gestures are important.

Second, future research can study the role of different cognitive abilities in gesture processing by employing different paradigms. Mismatch paradigms are useful to measure how listeners integrate concurrent spoken vs. gestural information (e.g., Kelly et al., 2010; Özyürek et al., 2007; Wu & Coulson, 2014). However, mismatches are rare in natural communication. Future studies can test how visual-spatial abilities relate to how much listeners benefit from observing gestures by using comprehension paradigms in which participants are presented with bimodal vs. unimodal encodings (Aldugom et al., 2020; Kartalkanat & Göksun, 2020). In most cases, gestures add information to the accompanying speech, and the additional semantic information conveyed by gestures facilitate listeners' comprehension differentially depending on their cognitive dispositions. Related to these two suggestions, I currently investigate individual differences in the gestural enhancement of comprehension when gestures express complementary vs. redundant information by employing the same experimental paradigm used in Chapter 4. I specifically examine the role of spatial working memory capacity and perspective-taking ability on how much listeners benefit from observing redundant vs. complementary gestures during the comprehension of spatial relations. I believe this study will provide evidence on how individuals' visual-spatial skills interact with different gesture-speech contexts during spatial language comprehension.

Third, future research should address why, when, and who employs gestures in interactional settings. Studies presented in this dissertation used experimental paradigms in which an actress passively described actions or spatial relations. In most cases, these paradigms reduce the natural qualities of human face-to-face communication. In natural interactive communication settings, interlocutors exchange a diverse set of multimodal cues that are inextricably linked to each other (Holler & Levinson, 2019; Holler, Kendrick, & Levinson, 2019). Examining how individuals' cognitive dispositions relate to their attention to and processing of speakers' gestures when all multimodal cues are present in an interactive manner (e.g., gestures coupled with social eye gaze) would be promising. Additionally, certain cognitive dispositions might manifest themselves differently in interactional settings. For example, communicative fluency might be more important when people are in dialogue. Future research in interactional settings would have many practical implications as most natural communication occurs in such situations, such as classrooms (Alibali & Nathan, 2007; Singer & Goldin-Meadow, 2005; Valenzeno et al., 2003).

Fourth, it is important to see how combinations of different cognitive abilities interact with gesture use and processing. Individuals show variation in the propensity to use different cognitive resources. For example, speakers with high visual-spatial skills coupled with low verbal abilities used gestures the most (Hostetter & Alibali, 2007, 2011). Future research can test how capacities in different cognitive resources interact to yield different patterns in gesture use and processing for the same individuals.

Last, it is unknown how individuals' own gesture use relates to how and to what extent they attend to and process speakers' gestures. The neural instantiation of gesture perception in children was affected by a history of gesture production, suggesting that gesture processing might depend on sensorimotor processes (Wakefield et al., 2013). Future studies can investigate the production-perception cycle of gestures within individuals across behavioral and neuroimaging studies. For example, do people who benefit more from using gestures during speaking also benefit more from observing gestures during comprehension? These questions can be tested in different populations at different developmental timepoints or with clinical groups (e.g., Parkinson's disease, Cleary et al., 2011; Göksun, Özer, & Akbıyık, in press; Humphries, Holler, Crawford, Herrera, & Poliakoff, 2016) to understand how these two processes associate or dissociate from each other. These studies would benefit the current literature to a great extent as they would provide control over individuals' characteristics and would help to delineate the mechanisms governing gesture production vs. comprehension.

5.3. Conclusion

Speakers and listeners employ gestures during communication and thinking. This dissertation investigated factors that could contribute to variation in speakers' and listeners' employment of gestures across individuals and situations. As gestures reduce speakers' cognitive load, speakers might use gestures as a compensatory tool during speaking and thinking that interact with individuals' cognitive dispositions. For listeners, visual-spatial cognitive resources are critical for processing co-speech gestures and observing gestures might hinder comprehension for listeners with lower visual-spatial cognitive resources as the processing of additional visual information is demanding. Moreover, gestures that add critical semantic information to the accompanying speech are more communicative than gestures that convey redundant information as was present in the increased overt gaze allocation. Overall, this dissertation provides novel findings and suggests that gesture use and processing are not monolithic processes. Current theories and experimental practices should consider variation across individuals and situations to have a full picture of when, why, and for whom gestures are beneficial.

REFERENCES

Akbıyık, S., Karaduman, A., Göksun, T., & Chatterjee, A. (2018). The relationship between co-speech gesture production and macrolinguistic discourse abilities in people with focal brain injury. *Neuropsychologia*, *117*, 440-453.

Akhavan, N., Göksun, T., & Nozari, N. (2018). Integrity and function of gestures in aphasia. *Aphasiology*, *32*(11), 1310-1335.

Alamillo, A. R., Colletta, J. M., & Guidetti, M. (2013). Gesture and language in narratives and explanations: the effects of age and communicative activity on late multimodal discourse development. *Journal of Child Language*, *40*(3), 511-538.

Aldugom, M., Fenn, K., & Cook, S. W. (2020). Gesture during math instruction specifically benefits learners with high visuospatial working memory capacity. *Cognitive Research: Principles and Implications*, *5*(1), 1-12.

Alfred, K. L., & Kraemer, D. J. (2017). Verbal and visual cognition: Individual differences in the lab, in the brain, and in the classroom. *Developmental Neuropsychology*, *42*(7-8), 507-520.

Alibali, M. W., Flevares, L. M., & Goldin-Meadow, S. (1997). Assessing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of Educational Psychology*, *89*(1), 183.

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, *44*(2), 169-188.

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation*, *5*(4), 307-331.

Alibali, M. W., & Nathan, M. J. (2007). *Teachers' gestures as a means of scaffolding students' understanding: Evidence from an early algebra lesson*. In R. Goldman, R. Pea, B. Barron, & S. J. Derry (Eds.), Video research in the learning sciences (pp. 349–365). Mahwah, NJ: Erlbaum.

Alibali, M. W., Evans, J. L., Hostetter, A. B., Ryan, K., & Mainela-Arnold, E. (2009). Gesture–speech integration in narrative: Are children less redundant than adults? *Gesture*, *9*(3), 290-311.

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the Learning Sciences*, *21*(2), 247-286.

Alfred, K. L., & Kraemer, D. J. (2017). Verbal and visual cognition: Individual differences in the lab, in the brain, and in the classroom. *Developmental Neuropsychology*, *42*(7-8), 507-520.

Andersen, G. J., & Ni, R. (2008). Aging and visual processing: Declines in spatial not temporal integration. *Vision Research*, *48*(1), 109-118.

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Oxford, England: Cambridge University Press.

Argyle, M., & Graham, J. A. (1976). The central Europe experiment: Looking at persons and looking at objects. *Environmental Psychology and Nonverbal Behavior*, *1*(1), 6-16.

Arslan, B. & Göksun, T. (in press). Aging, working memory, and mental imagery: Understanding gestural communication in younger and older adults. *Quarterly Journal of Experimental Psychology*.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, Vol. 2, pp. 89-195.

Ausburn, L. J., & Ausburn, F. B. (1978). Cognitive styles: Some information and implications for instructional design. *Educational Technology Research and Development*, *26*(4), 337-354.

Aussems, S., & Kita, S. (2019). Seeing iconic gestures while encoding events facilitates children's memory of these events. *Child Development*, *90*(4), 1123-1137.

Austin, E. E., & Sweller, N. (2014). Presentation and production: The role of gesture in spatial communication. *Journal of Experimental Child Psychology*, *122*, 92-103.

Austin, E. E., & Sweller, N. (2017). Getting to the elephants: Gesture and preschoolers' comprehension of route direction information. *Journal of Experimental Child psychology*, *163*, 1-14.

Austin, E. E., & Sweller, N. (2018). Gesturing along the way: Adults' and preschoolers' communication of route direction information. *Journal of Nonverbal Behavior, 42*(2), 199-220.

Azar, Z., Backus, A., & Özyürek, A. (2019). General-and language-specific factors influence reference tracking in speech and gesture in discourse. *Discourse Processes*, *56*(7), 553-574.

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.

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.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation,* Vol. 8, pp. 47-89.

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.

Baddeley, A. D. (1992). Working memory. *Science*, 255, 556–559.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory. *Trends in Cognitive Sciences*, *4*(11), 417-423.

Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. *The Handbook of Child Language*, *30*, 96-151.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. Retrieved from http://arxiv.org/abs/1406.5823.

Bates, D., Mächler, M., Bolker, B., Walker, S., Bojesen, R.H., Singmann, H., … Fox, J. (2019). Linear mixed-effects models using "Eigen" and S4. Retrieved from https://cran-r.project.org/web/packages/lme4/lme4.pdf.

Baxter, J. C., Winters, E. P., & Hammer, R. E. (1968). Gestural behavior during a brief interview as a function of cognitive variables. *Journal of Personality and Social Psychology*, *8*, 303.

Beattie, G., & Shovelton, H. (1999a). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica*, *123*(1-2), 1-30.

Beattie, G., & Shovelton, H. (1999b). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology*, *18*(4), 438-462.

Beattie, G., & Shovelton, H. (2001). An experimental investigation of the role of different types of iconic gesture in communication: A semantic feature approach. *Gesture*, *1*(2), 129-149.

Beattie, G., & Shovelton, H. (2002). What properties of talk are associated with the generation of spontaneous iconic hand gestures? *British Journal of Social Psychology*, *41*(3), 403-417.

Beattie, G., & Shovelton, H. (2006). When size really matters: How a single semantic feature is represented in the speech and gesture modalities. *Gesture*, *6*(1), 63-84.

Beattie, G., Webster, K., & Ross, J. (2010). The fixation and processing of the iconic gestures that accompany talk. *Journal of Language and Social Psychology*, *29*(2), 194-213.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, *41*(5), 809-823.

Bello, A., Capirci, O., & Volterra, V. (2004). Lexical production in children with Williams syndrome: Spontaneous use of gesture in a naming task. *Neuropsychologia*, *42*(2), 201-213.

Bernardis, P., & Caramelli, N. (2009, January). Meaning in words, gestures, and mental images. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 31, No. 31).

Biau, E., & Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain and Language*, *124*(2), 143-152.

Biau, E., & Soto-Faraco, S. (2015). Synchronization by the hand: the sight of gestures modulates low-frequency activity in brain responses to continuous speech. *Frontiers in Human Neuroscience*, *9*.

Blake, J., Myszczyszyn, D., Jokel, A., & Bebiroglu, N. (2008). Gestures accompanying speech in specifically language-impaired children and their timing with speech. *First Language*, *28*(2), 237-253.

Bohn, M., Kordt, C., Braun, M., Call, J., & Tomasello, M. (in press). Learning novel skills from iconic gestures: A developmental and evolutionary perspective. *Psychological Science*.

Brayne, C., Gao, L., Dewey, M., Matthews, F. E., & Ageing Study Investigators. (2006). Dementia before death in ageing societies—the promise of prevention and the reality. *PLoS Medicine*, *3*(10).

Broadway, J. M., & Engle, R. W. (2011). Individual differences in working memory capacity and temporal discrimination. *PLoS One*, *6*(10).

Calero, C. I., Shalom, D. E., Spelke, E. S., & Sigman, M. (2019). Language, gesture, and judgment: children's paths to abstract geometry. *Journal of Experimental Child Psychology*, *177*, 70-85.

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

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

Calvert, G., Spence, C., & Stein, B. E. (Eds.). (2004). *The handbook of multisensory processes*. MIT press.

Capirci, O., Contaldo, A., Caselli, M. C., & Volterra, V. (2005). From action to language through gesture: A longitudinal perspective. *Gesture*, *5*(1-2), 155-177.

Capirci, O., & Volterra, V. (2008). Gesture and speech: The emergence and development of a strong and changing partnership. *Gesture*, *8*(1), 22-44.

Cassell, J., McNeill, D., & McCullough, K. E. (1999). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics & Cognition*, *7*(1), 1-34.

Cavicchio, F., & Kita, S. (2013). English / Italian bilinguals switch gesture parameters when they switch languages. *Proceedings of the Annual Meeting of the Cognitive Science*, 35(35), 305–309.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), 293-332.

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General*, *137*(4), 706.

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, *140*(1), 102.

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: The role of cognitive abilities and empathy. *Journal of Experimental Psychology: General*, *143*(2), 694.

Chui, K. (2011). Do gestures compensate for the omission of motion expression in speech? *Chinese Language and Discourse*, *2*(2), 153-167.

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, *23*(1), 43-71.

Church, R. B., Kelly, S. D., & Lynch, K. (2000). Immediate memory for mismatched speech and representational gesture across development. *Journal of Nonverbal Behavior*, *24*(2), 151-174.

Church, R. B., Ayman-Nolley, S., & Mahootian, S. (2004). The role of gesture in bilingual education: Does gesture enhance learning? *International Journal of Bilingual Education and Bilingualism*, *7*(4), 303-319.

Church, R. B., Garber, P., & Rogalski, K. (2007). The role of gesture in memory and social communication. *Gesture*, *7*(2), 137-158.

Clark, H. (1996). *Using Language*. Cambridge University Press, Cambridge.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149-210.

Cleary, R. A., Poliakoff, E., Galpin, A., Dick, J. P., & Holler, J. (2011). An investigation of co-speech gesture production during action description in Parkinson's disease. *Parkinsonism & Related Disorders*, *17*(10), 753-756.

Clough, S., & Duff, M. C. (2020). The role of gesture in communication and cognition: Implications for understanding and treating neurogenic communication disorders. *Frontiers in Human Neuroscience*, 14:323.

Cocks, N., Morgan, G., & Kita, S. (2011). Iconic gesture and speech integration in younger and older adults. *Gesture*, *11*(1), 24-39.

Cocks, N., Byrne, S., Pritchard, M., Morgan, G., & Dipper, L. (2018). Integration of speech and gesture in aphasia. *International Journal of Language & Communication Disorders*, *53*(3), 584-591.

Cohen, R. L., & Borsoi, D. (1996). The role of gestures in description-communication: A cross-sectional study of aging. *Journal of Nonverbal Behavior*, *20*(1), 45-63.

Colgan, S. E., Lanter, E., McComish, C., Watson, L. R., Crais, E. R., & Baranek, G. T. (2006). Analysis of social interaction gestures in infants with autism. *Child Neuropsychology*, *12*(4-5), 307-319.

Colletta, J. M., Pellenq, C., & Guidetti, M. (2010). Age-related changes in co-speech gesture and narrative: Evidence from French children and adults. *Speech Communication*, *52*(6), 565-576.

Conway, A. R., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*(2), 331-335.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769-786.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, *106*(2), 1047-1058.

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes*, *27*(4), 594-610.

Cook, S. W., & Fenn, K. M. (2017). The function of gesture in learning and memory. *Gesture Studies*, *7*, 129-153.

Copeland, D. E., & Radvansky, G. A. (2007). Aging and integrating spatial mental models. *Psychology and Aging*, *22*(3), 569.

Corsi, P. M. (1972) *Human memory and the medial temporal region of the brain* (Unpublished doctoral dissertation). McGill University, Montreal, Canada.

Dahl, T. I., & Ludvigsen, S. (2014). How I see what you're saying: The role of gestures in native and foreign language listening comprehension. *The Modern Language Journal*, *98*(3), 813-833.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*(1), 1-18.

Dargue, N., Sweller, N., & Jones, M. P. (2019). When our hands help us understand: A meta-analysis into the effects of gesture on comprehension. *Psychological Bulletin*, *145*(8), 765.

Dargue, N., & Sweller, N. (2020). Learning Stories Through Gesture: Gesture's Effects on Child and Adult Narrative Comprehension. *Educational Psychology Review*, *32*(1), 249-276.

de Fockert, J. W., Rees, G., Frith, C. D., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science*, *291*(5509), 1803-1806.

Demir, Ö. E., Fisher, J. A., Goldin-Meadow, S., & Levine, S. C. (2014). Narrative processing in typically developing children and children with early unilateral brain injury: Seeing gesture matters. *Developmental Psychology*, *50*(3), 815.

Demir, Ö. E., Levine, S. C., & Goldin-Meadow, S. (2015). A tale of two hands: Children's early gesture use in narrative production predicts later narrative structure in speech. *Journal of Child Language*, *42*(3), 662-681.

Demir-Lira, Ö. E., Asaridou, S. S., Raja Beharelle, A., Holt, A. E., Goldin-Meadow, S., & Small, S. L. (2018). Functional neuroanatomy of gesture–speech integration in children varies with individual differences in gesture processing. *Developmental Science*, *21*(5), e12648.

de Nooijer, J. A., Van Gog, T., Paas, F., & Zwaan, R. A. (2013). Effects of imitating gestures during encoding or during retrieval of novel verbs on children's test performance. *Acta Psychologica*, *144*(1), 173-179.

de Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis. *Topics in Cognitive Science*, *4*(2), 232-248.

Dick, A. S., Goldin-Meadow, S., Solodkin, A., & Small, S. L. (2012). Gesture in the developing brain. *Developmental Science*, *15*(2), 165-180.

Dimeck, P. T., Roy, E. A., & Hall, C. R. (1998). Aging and working memory in gesture imitation. *Brain and Cognition*, *37*(1), 124-127.

Dimitrova, N., Özçalışkan, Ş., & Adamson, L. B. (2016). Parents' translations of child gesture facilitate word learning in children with autism, Down syndrome, and typical development. *Journal of Autism and Developmental Disorders*, *46*(1), 221-231.

Drijvers, L., & Özyürek, A. (2017). Visual context enhanced: The joint contribution of iconic gestures and visible speech to degraded speech comprehension. *Journal of Speech, Language, and Hearing Research*, *60*(1), 212-222.

Drijvers, L., & Özyürek, A. (2018). Native language status of the listener modulates the neural integration of speech and iconic gestures in clear and adverse listening conditions. *Brain and Language*, *177*, 7-17.

Drijvers, L., Vaitonytė, J., & Özyürek, A. (2019). Degree of language experience modulates visual attention to visible speech and iconic gestures during clear and degraded speech comprehension. *Cognitive Science*, *43*(10), e12789.

Drijvers, L., & Özyürek, A. (2020). Non-native listeners benefit less from gestures and visible speech than native listeners during degraded speech comprehension. *Language and Speech*, *63*(2), 209-220.

Driskell, J. E., & Radtke, P. H. (2003). The effect of gesture on speech production and comprehension. *Human Factors*, *45*(3), 445-454.

Dror, I. E., & Kosslyn, S. M. (1994). Mental imagery and aging. *Psychology and Aging*, *9*(1), 90.

Ehrlich, S. B., Levine, S. C., & Goldin-Meadow, S. (2006). The importance of gesture in children's spatial reasoning. *Developmental Psychology*, *42*(6), 1259.

Eielts, C., Pouw, W., Ouwehand, K., Van Gog, T., Zwaan, R. A., & Paas, F. (2018). Co-thought gesturing supports more complex problem solving in subjects with lower visual working-memory capacity. *Psychological Research*, 1-12.

Emmorey, K., Tversky, B., & Taylor, H. A. (2000). Using space to describe space: Perspective in speech, sign, and gesture. *Spatial Cognition and Computation*, *2*(3), 157-180.

Emmorey, K., & Casey, S. (2001). Gesture, thought and spatial language. *Gesture*, *1*(1), 35-50.

Emmorey, K., Thompson, R., & Colvin, R. (2009). Eye gaze during comprehension of American Sign Language by native and beginning signers. *Journal of Deaf Studies and Deaf Education*, *14*(2), 237-243.

Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*. Cambridge University Press, Cambridge, UK.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19-23.

Evans, J. L., Alibali, M. W., & McNeil, N. M. (2001). Divergence of verbal expression and embodied knowledge: Evidence from speech and gesture in children with specific language impairment. *Language and Cognitive Processes*, *16*(2-3), 309-331.

Feyereisen, P., & Havard, I. (1999). Mental imagery and production of hand gestures while speaking in younger and older adults. *Journal of Nonverbal Behavior*, *23*(2), 153-171.

Frick-Horbury, D. (2002). The effects of hand gestures on verbal recall as a function of high-and low-verbal-skill levels. *The Journal of General Psychology*, *129*(2), 137-147.

Galati, A., & Avraamides, M. N. (2013). Flexible spatial perspective-taking: conversational partners weigh multiple cues in collaborative tasks. *Frontiers in Human Neuroscience*, *7*, 618.

Galati, A., Michael, C., Mello, C., Greenauer, N. M., & Avraamides, M. N. (2013). The conversational partner's perspective affects spatial memory and descriptions. *Journal of Memory and Language*, *68*(2), 140-159.

Galati, A., Weisberg, S., Newcombe, N., & Avraamides, M. N. (2015). Individual differences in spatial ability influence the effect of gesturing on navigation and spatial memory. *Proceedings of Gesture and Speech in Interaction-4th edition (GESPIN 4)*, 119-124.

Galati, A., Weisberg, S. M., Newcombe, N. S., & Avraamides, M. N. (2018). When gestures show us the way: Co-thought gestures selectively facilitate navigation and spatial memory. *Spatial Cognition & Computation*, *18*(1), 1-30.

Gerwing, J., & Allison, M. (2009). The relationship between verbal and gestural contributions in conversation: A comparison of three methods. *Gesture*, *9*(3), 312-336.

Gillespie, M., James, A. N., Federmeier, K. D., & Watson, D. G. (2014). Verbal working memory predicts co-speech gesture: Evidence from individual differences. *Cognition*, *132*(2), 174-180.

Glasser, M. L., Williamson, R. A., & Özçalışkan, Ş. (2018). Do children understand iconic gestures about events as early as iconic gestures about entities? *Journal of Psycholinguistic Research*, *47*(3), 741-754.

Glosser, G., Wiley, M. J., & Barnoskir, E. J. (1998). Gestural communication in Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, *20*(1), 1-13.

Gobel, M. S., Kim, H. S., & Richardson, D. C. (2015). The dual function of social gaze. *Cognition*, *136*, 359-364.

Goldin-Meadow, S., & Sandhofer, C. M. (1999). Gestures convey substantive information about a child's thoughts to ordinary listeners. *Developmental Science*, *2*(1), 67-74.

Goldin-Meadow, S., & Saltzman, J. (2000). The cultural bounds of maternal accommodation: How Chinese and American mothers communicate with deaf and hearing children. *Psychological Science*, *11*(4), 307-314.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, *12*(6), 516-522.

Goldin-Meadow, S. (2003). *How our hands help us think.* Cambridge, MA: Harvard University Press.

Goldin-Meadow, S., & Singer, M. A. (2003). From children's hands to adults' ears: Gesture's role in the learning process. *Developmental Psychology*, *39*(3), 509.

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, *64*, 257-283.

Göksun, T., Hirsh-Pasek, K., & Golinkoff, R. M. (2010). How do preschoolers express cause in gesture and speech? *Cognitive Development*, *25*(1), 56-68.

Göksun, T., Lehet, M., Malykhina, K., & Chatterjee, A. (2013). Naming and gesturing spatial relations: Evidence from focal brain-injured individuals. *Neuropsychologia*, *51*(8), 1518-1527.

Göksun, T., Goldin-Meadow, S., Newcombe, N., & Shipley, T. (2013). Individual differences in mental rotation: what does gesture tell us? *Cognitive Processing*, *14*(2), 153-162.

Göksun, T., Lehet, M., Malykhina, K., & Chatterjee, A. (2015). Spontaneous gesture and spatial language: Evidence from focal brain injury. *Brain and Language*, *150*, 1-13.

Göksun, T., Özer, D., & Akbıyık, S. (in press). *Gesture in the Aging Brain.* In A. Morgenstern & S. Goldin-Meadow (Eds.), Gesture in Language. De Gruyter Mouton

Graham, J. A., & Heywood, S. (1975). The effects of elimination of hand gestures and of verbal codability on speech performance. *European Journal of Social Psychology, 5*(2), 189-195.

Gullberg, M., & Holmqvist, K. (1999). Keeping an eye on gestures: Visual perception of gestures in face-to-face communication. *Pragmatics & Cognition*, *7*(1), 35-63.

Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics & Cognition*, *14*(1), 53-82.

Gullberg, M., & Kita, S. (2009). Attention to speech-accompanying gestures: Eye movements and information uptake. *Journal of Nonverbal Behavior*, *33*(4), 251-277.

Gullberg, M. (2010). Methodological reflections on gesture analysis in second language acquisition and bilingualism research. *Second Language Research*, *26*(1), 75-102.

Gunter, T. C., & Bach, P. (2004). Communicating hands: ERPs elicited by meaningful symbolic hand postures. *Neuroscience Letters, 372*(1-2), 52-56.

Hadar, U., & Butterworth, B. (1997). Iconic gestures, imagery, and word retrieval in speech. *Semiotica*, *115*(1-2), 147-172.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, *9*(9), 416-423.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. Gazzaniga (Ed.), *The Cognitive Neuroscience, (*pp. 819 – 836). Boston: MIT Press.

Hagoort, P. (2013). MUC (memory, unification, control) and beyond. *Frontiers in Psychology*, *4*, 416.

Hilverman, C., Cook, S. W., & Duff, M. C. (2018). Hand gestures support word learning in patients with hippocampal amnesia. *Hippocampus*, *28*(6), 406-415.

Hilverman, C., Clough, S. A., Duff, M. C., & Cook, S. W. (2018). Patients with hippocampal amnesia successfully integrate gesture and speech. *Neuropsychologia*, *117,* 332-338.

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, *19*(7), 1175-1192.

Holle, H., Obleser, J., Rueschemeyer, S. A., & Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *Neuroimage*, *49*(1), 875-884.

Holle, H., Obermeier, C., Schmidt-Kassow, M., Friederici, A. D., Ward, J., & Gunter, T. C. (2012). Gesture facilitates the syntactic analysis of speech. *Frontiers in Psychology*, *3*.

Holler, J., & Stevens, R. (2007). The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Social Psychology*, *26*(1), 4-27.

Holler, J., Shovelton, H., & Beattie, G. (2009). Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context? *Journal of Nonverbal Behavior*, *33*(2), 73-88.

Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M., & Özyürek, A. (2014). Social eye gaze modulates processing of speech and co-speech gesture. *Cognition*, *133*(3), 692-697.

Holler, J., Kendrick, K. H., & Levinson, S. C. (2018). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic Bulletin & Review*, *25*(5), 1900-1908.
Holler, J., Kendrick, K. H., & Levinson, S. C. (2018). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic Bulletin & Review*, *25*(5), 1900-1908.
Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639-652.

Hostetter, A. B., & Hopkins, W. D. (2002). The effect of thought structure on the production of lexical movements. *Brain and Language*, *82*(1), 22-29.

Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: Relations between verbal and spatial skills and gesture production. *Gesture*, *7*(1), 73-95.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495-514.

Hostetter, A. B., & Alibali, M. W. (2011). Cognitive skills and gesture–speech redundancy: Formulation difficulty or communicative strategy? *Gesture*, *11*(1), 40-60.

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, *137*(2), 297.

Hostetter, A. B., Alibali, M. W., & Bartholomew, A. (2011). Gesture during mental rotation. In L. Carlson, C. Hoelscher & T. F. Shipley, (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1448–1453). Austin, TX: Cognitive Science Society.

Hostetter, A., & Sullivan, E. (2011). Gestures are produced during spatial tasks. In L. Carlson, C. Hoelscher & T. F. Shipley, (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Hostetter, A. B., & Potthoff, A. L. (2012). Effects of personality and social situation on representational gesture production. *Gesture*, *12*(1), 62-83.

Hostetter, A. B., & Alibali, M. W. (2018). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review*, *26*(3), 721-752.

Hostetter, A. B., Murch, S. H., Rothschild, L., & Gillard, C. S. (2018). Does seeing gesture lighten or increase the load? Effects of processing gesture on verbal and visuospatial cognitive load. *Gesture*, *17*(2), 268-290.

Hubbard, A. L., Wilson, S. M., Callan, D. E., & Dapretto, M. (2009). Giving speech a hand: Gesture modulates activity in auditory cortex during speech perception. *Human Brain Mapping*, *30*(3), 1028-1037.

Hubbard, A. L., McNealy, K., Scott-Van Zeeland, A. A., Callan, D. E., Bookheimer, S. Y., & Dapretto, M. (2012). Altered integration of speech and gesture in children with autism spectrum disorders. *Brain and Behavior*, *2*(5), 606-619.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151-171.

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience, 31*(1), 80-93.

Huk, T. (2006). Who benefits from learning with 3D models? The case of spatial ability. *Journal of Computer Assisted Learning*, *22*(6), 392-404.

Humphries, S., Holler, J., Crawford, T. J., Herrera, E., & Poliakoff, E. (2016). A third-person perspective on co-speech action gestures in Parkinson's disease. *Cortex*, *78*, 44-54.

Huyse, A., Leybaert, J., & Berthommier, F. (2014). Effects of aging on audio-visual speech integration. *The Journal of the Acoustical Society of America*, *136*(4), 1918-1931.

Iverson, J. M., Capirci, O., Longobardi, E., & Caselli, M. C. (1999). Gesturing in mother-child interactions. *Cognitive Development*, *14*(1), 57-75.

Iverson, J. M., Longobardi, E., Spampinato, K., & Cristina Caselli, M. (2006). Gesture and speech in maternal input to children with Down's syndrome. *International Journal of Language & Communication Disorders*, *41*(3), 235-251.

Iverson, J. M., Capirci, O., Volterra, V., & Goldin-Meadow, S. (2008). Learning to talk in a gesture-rich world: Early communication in Italian vs. American children. *First Language*, *28*(2), 164-181.

Iverson, J. M., & Braddock, B. A. (2011). Gesture and motor skill in relation to language in children with language impairment. *Journal of Speech, Language, and Hearing Research*, 54(1), 72-86.

Jonassen, D. H., & Grabowski, B. (1993). *Individual differences and instruction*. New York: Allen & Bacon.

Jorm, A. F., Korten, A. E., & Henderson, A. S. (1987). The prevalence of dementia: a quantitative integration of the literature. *Acta Psychiatrica Scandinavica*, *76*(5), 465-479.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, *99*(1), 122.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509-539.

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637-671.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*(1), 47.

Karadöller, D. Z., Ünal, E., Sümer, B., Göksun, T., Özer, D., & Ozyurek, A. (2019). Children but not adults use both speech and gesture to produce informative expressions of Left-Right relations. In *the 44th Annual Boston University Conference on Language Development (BUCLD 44)*.

Karadöller, D. Z., Sümer, B., & Özyürek, A. (in press). Effects and non-effects of late language exposure on spatial language development: Evidence from deaf adults and children. *Language Learning and Development*.

Kartalkanat, H., & Göksun, T. (2020). The effects of observing different gestures during storytelling on the recall of path and event information in 5-year-olds and adults. *Journal of Experimental Child Psychology*, *189*, 104725.

Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, *40*(4), 577-592.

Kelly, S. D. (2001). Broadening the units of analysis in communication: Speech and nonverbal behaviours in pragmatic comprehension. *Journal of Child Language*, *28*(2), 325-349.

Kelly, S. D., & Goldsmith, L. H. (2004). Gesture and right hemisphere involvement in evaluating lecture material. *Gesture*, *4*(1), 25-42.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language*, *89*(1), 253-260.

Kelly, S. D., Ward, S., Creigh, P., & Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain and Language*, *101*(3), 222-233.

Kelly, S. D., Manning, S. M., & Rodak, S. (2008). Gesture gives a hand to language and learning: Perspectives from cognitive neuroscience, developmental psychology, and education. *Language and Linguistics Compass*, *2*(4), 569-588.

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*(2), 313-334.

Kelly, S. D., Özyürek, A., & Maris, E. (2010a). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science, 21*(2), 260-267.

Kelly, S. D., Creigh, P., & Bartolotti, J. (2010b). Integrating speech and iconic gestures in a Stroop-like task: evidence for automatic processing. *Journal of Cognitive Neuroscience*, *22*(4), 683-694.

Kendon, A. (1994). Do gestures communicate? A review. *Research on Language and Social Interaction*, *27*(3), 175-200.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32-38.

Kiat, J. E., & Belli, R. F. (2018). The role of individual differences in visual\verbal information processing preferences in visual\verbal source monitoring. *Journal of Cognitive Psychology*, *30*(7), 701-709.

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*(2), 154-169.

Kim, M., Roth, W. M., & Thom, J. (2011). Children's gestures and the embodied knowledge of geometry. *International Journal of Science and Mathematics Education*, *9*(1), 207-238.

Kim, Z. H., & Lausberg, H. (2018). Koreans and Germans: Cultural Differences in Hand Movement Behaviour and Gestural Repertoire. *Journal of Intercultural Communication Research*, *47*(6), 439-453.

Kirby, J. R., Moore, P. J., & Schofield, N. J. (1988). Verbal and visual learning styles. *Contemporary Educational Psychology*, *13*(2), 169-184.

Kirschner, P. A. (2017). Stop propagating the learning styles myth. *Computers & Education*, *106*, 166-171.

Kirk, E., Pine, K. J., & Ryder, N. (2011). I hear what you say but I see what you mean: The role of gestures in children's pragmatic comprehension. *Language and Cognitive Processes*, *26*(2), 149-170.

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.

Kita, S. (2000) How representational gestures help speaking? In D. McNeill (Ed.) *Language and Gesture*. Cambridge, UK: Cambridge University Press.

Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes*, *24*(2), 145-167.

Kita, S., & Davies, T. S. (2009). Competing conceptual representations trigger co-speech representational gestures. *Language and Cognitive Processes*, *24*(5), 761-775.

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.

Kirby, J., Moore, P., & Shofield, N. (1988). Verbal and visual learning styles. *Contemporary Educational Psychology*, 13, 169–184.

Klooster, N. B., Cook, S. W., Uc, E. Y., & Duff, M. C. (2015). Gestures make memories, but what kind? Patients with impaired procedural memory display disruptions in gesture production and comprehension. *Frontiers in Human Neuroscience*, *8*, 1054.

Koć-Januchta, M., Höffler, T., Thoma, G. B., Prechtl, H., & Leutner, D. (2017). Visualizers versus verbalizers: Effects of cognitive style on learning with texts and pictures–An eye-tracking study. *Computers in Human Behavior*, *68*, 170-179.

Kollöffel, B. (2012). Exploring the relation between visualizer–verbalizer cognitive styles and performance with visual or verbal learning material. *Computers & Education*, *58*(2), 697-706.

Kozhevnikov, M., Hegarty, M., & Mayer, R. E. (2002). Revising the visualizer/verbalizer dimension: evidence for two types of visualizers. *Cognition & Instruction*, 20, 47–77.

Kraemer, D. J., Rosenberg, L. M., & Thompson-Schill, S. L. (2009). The neural correlates of visual and verbal cognitive styles. *Journal of Neuroscience*, *29*(12), 3792-3798.

Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology*, *61*(5), 743.

Krauss, R. M., Dushay, R. A., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology*, *31*, 533-552.

Krauss, R. M., & Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. *Gesture, Speech, and Sign*, *93*.

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: a process model. In D. McNeill (Ed.), *Language and gesture*. Cambridge, UK: Cambridge University Press.

La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(6), 1118.

Laurent, A., & Nicoladis, E. (2015). Gesture restriction affects French–English bilinguals' speech only in French. *Bilingualism: Language and Cognition*, *18*(2), 340-349.

Lausberg, H., & Kita, S. (2003). The content of the message influences the hand choice in co-speech gestures and in gesturing without speaking. *Brain and Language*, *86*(1), 57-69.

Lavergne, J., & Kimura, D. (1987). Hand movement asymmetry during speech: No effect of speaking topic. *Neuropsychologia*, *25*(4), 689-693.

LeBarton, E.S. & Iverson, J.M. (2017). Gesture’s role in learning interactions. In R.B. Church, M.W. Alibali, & S.D., Kelly (Eds.), *Why gesture? How the hands function in speaking, thinking and communicating* (pp. 331-351). John Benjamins Publishing.

Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, *26*(10), 1457-1471.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, *108*(3), 732-739.

Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, *32*(6), 982-998.

Macoun, A., & Sweller, N. (2016). Listening and watching: The effects of observing gesture on preschoolers’ narrative comprehension. *Cognitive Development*, *40*, 68-81.

Mainela-Arnold, E., Alibali, M. W., Ryan, K., & Evans, J. L. (2011). Knowledge of mathematical equivalence in children with specific language impairment: Insights from gesture and speech. *Language, Speech, and Hearing Services in Schools*, 42(1), 18-30.

Mainela-Arnold, E., Alibali, M. W., Hostetter, A. B., & Evans, J. L. (2014). Gesture–speech integration in children with specific language impairment. *International Journal of Language & Communication Disorders*, *49*(6), 761-770.

Mainwaring, S. D., Tversky, B., Ohgishi, M., & Schiano, D. J. (2003). Descriptions of simple spatial scenes in English and Japanese. *Spatial Cognition and Computation*, *3*(1), 3-42.

Marstaller, L., & Burianová, H. (2013). Individual differences in the gesture effect on working memory. *Psychonomic Bulletin & Review*, *20*(3), 496-500.

Mastrogiuseppe, M., Capirci, O., Cuva, S., & Venuti, P. (2015). Gestural communication in children with autism spectrum disorders during mother–child interaction. *Autism*, *19*(4), 469-481.

Mayberry, R. I., & Nicoladis, E. (2000). Gesture reflects language development: Evidence from bilingual children. *Current Directions in Psychological Science*, *9*(6), 192-196.

McNeil, N. M., Alibali, M. W., & Evans, J. L. (2000). The role of gesture in children's comprehension of spoken language: Now they need it, now they don't. *Journal of Nonverbal Behavior*, *24*(2), 131-150.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought.* Chicago, IL: University of Chicago Press.

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

McNeill, D. (2005). *Gesture and Thought.* Chicago, IL: University of Chicago Press.

Melinger, A., & Levelt, W. J. (2004). Gesture and the communicative intention of the speaker. *Gesture*, *4*(2), 119-141.

Mendelson, A. L. (2004). For whom is a picture worth a thousand words? Effects of the visualizing cognitive style and attention on processing of news photos. *Journal of Visual Literacy*, *24*(1), 1-22.

Mendelson, A. L., & Thorson, E. (2004). How verbalizers and visualizers process the newspaper environment. *Journal of Communication*, *54*(3), 474-491.

Meneghetti, C., Labate, E., Grassano, M., Ronconi, L., & Pazzaglia, F. (2014). The role of visuospatial and verbal abilities, styles, and strategies in predicting visuospatial description accuracy. *Learning and Individual Differences*, *36*, 117-123.

Miller, G.A., & Johnson-Laird, P.N. (1976). Language and perception. Cambridge MA: The Belknap Press of Harvard University.

Montepare, J., Koff, E., Zaitchik, D., & Albert, M.L. (1999). The use of body movements and gestures as cues to emotions in younger and older adults. *Journal of Nonverbal Behavior*, 23 (2), 133–152.

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94*(1), 156.

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 615.

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology*, 411-424.

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders, 20*(1), 115-128.

Nagels, A., Kircher, T., Steines, M., Grosvald, M., & Straube, B. (2015). A brief self-rating scale for the assessment of individual differences in gesture perception and production. *Learning and Individual Differences, 39,* 73-80.

Nagels, A., Kircher, T., Grosvald, M., Steines, M., & Straube, B. (2019). Evidence for gesture-speech mismatch detection impairments in schizophrenia. *Psychiatry Research*, *273*, 15-21.

Nagpal, J., Nicoladis, E., & Marentette, P. (2011). Predicting individual differences in L2 speakers' gestures. *International Journal of Bilingualism*, *15*(2), 205-214.

Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *Neuroimage*, *59*(1), 781-787.

Newcombe, N. S., Uttal, D. H., & Sauter, M. (2013). *Spatial development*. In P. D. Zelazo (Ed.), *Oxford library of psychology. The Oxford handbook of developmental psychology (Vol. 1): Body and mind* (p. 564–590). Oxford University Press.

Nicoladis, E., Pika, S., & Marentette, P. (2009). Do French–English bilingual children gesture more than monolingual children? *Journal of Psycholinguistic Research*, *38*(6), 573-585.

Nobe, S., Hayamizu, S., Hasegawa, O., & Takahashi, H. (1997, September). Are listeners paying attention to the hand gestures of an anthropomorphic agent? An evaluation using a gaze tracking method. In *International Gesture Workshop* (pp. 49-59). Springer, Berlin, Heidelberg.

Nobe, S., Hayamizu, S., Hasegawa, O., & Takahashi, H. (2000). Hand gestures of an anthropomorphic agent: Listeners' eye fixation and comprehension. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, *7*(1), 86-92.

Novack, M. A., & Goldin-Meadow, S. (2017). Gesture as representational action: A paper about function. *Psychonomic Bulletin & Review*, *24*(3), 652-665.

Obermeier, C., Holle, H., & Gunter, T. C. (2011). What iconic gesture fragments reveal about gesture–speech integration: When synchrony is lost, memory can help. *Journal of Cognitive Neuroscience*, *23*(7), 1648-1663.

Obermeier, C., Dolk, T., & Gunter, T. C. (2012). The benefit of gestures during communication: Evidence from hearing and hearing-impaired individuals. *Cortex*, *48*(7), 857-870.

Obermeier, C., Kelly, S., & Gunter, T.C. (2015) A speaker's gesture style can affect language comprehension: ERP evidence from gesture-speech integration. *SCAN* (10), 1236-1243.

Özçalışkan, Ş., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, *96*(3), 101-113.

Özçalışkan, Ş., & Goldin-Meadow, S. (2010). Sex differences in language first appear in gesture. *Developmental Science*, *13*(5), 752-760.

Özçalışkan, Ş., Levine, S. C., & Goldin-Meadow, S. (2013). Gesturing with an injured brain: How gesture helps children with early brain injury learn linguistic constructions. *Journal of Child Language*, *40*(1), 69.

Özçalışkan, Ş., Adamson, L. B., & Dimitrova, N. (2016). Early deictic but not other gestures predict later vocabulary in both typical development and autism. *Autism*, *20*(6), 754-763.

Özçalışkan, Ş., Adamson, L. B., Dimitrova, N., & Baumann, S. (2017). Early gesture provides a helping hand to spoken vocabulary development for children with autism, down syndrome, and typical development. *Journal of Cognition and Development*, *18*(3), 325-337.

Özçalışkan, Ş., Adamson, L. B., Dimitrova, N., & Baumann, S. (2018). Do parents model gestures differently when children's gestures differ? *Journal of Autism and Developmental Disorders*, *48*(5), 1492-1507.

Özer, D., Tansan, M., Özer, E. E., Malykhina, K., Chatterjee, A., & Göksun, T. (2017). The effects of gesture restriction on spatial language in young and elderly adults. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1471–1476). Austin, TX: Cognitive Science Society.

Özer, D., Göksun, T., & Chatterjee, A. (2019). Differential roles of gestures on spatial language in neurotypical elderly adults and individuals with focal brain injury. *Cognitive Neuropsychology*, *36*(5-6), 282-299.

Özer, D. & Göksun, T. (2020a). Gesture use and processing: a review on individual differences in cognitive resources. *Frontiers in Psychology,* 11: 573555.

Özer, D., & Göksun, T. (2020b). Visual-spatial and verbal abilities differentially affect processing of gestural vs. spoken expressions. *Language, Cognition and Neuroscience*, 35(7), 896-914.

Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, *19*(4), 605-616.

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130296.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*(1), 1-4.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*(1), 1-8.

Paas, F., Van Gog, T., & Sweller, J. (2010). Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review*, *22*(2), 115-121.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior*, *19*(2), 221-243.

Perry, M., Church, R. B., & Goldin-Meadow, S. (1992). Is gesture-speech mismatch a general index of transitional knowledge? *Cognitive Development*, *7*(1), 109-122.

Pika, S., Nicoladis, E., & Marentette, P. F. (2006). A cross-cultural study on the use of gestures: Evidence for cross-linguistic transfer? *Bilingualism: Language and Cognition*, *9*(3), 319-327.

Ping, R., & Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cognitive Science*, *34*(4), 602-619.

Posner, M. I. (2016). Orienting of attention: Then and now. *Quarterly Journal of Experimental Psychology*, *69*(10), 1864-1875.

Post, L. S., Van Gog, T., Paas, F., & Zwaan, R. A. (2013). Effects of simultaneously observing and making gestures while studying grammar animations on cognitive load and learning. *Computers in Human Behavior*, *29*(4), 1450-1455.

Pouw, W. T., 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*, 359.

Pouw, W. T., Mavilidi, M. F., van Gog, T., & Paas, F. (2016). Gesturing during mental problem solving reduces eye movements, especially for individuals with lower visual working memory capacity. *Cognitive Processing*, *17*(3), 269-277.

Priesters, M. A., & Mittelberg, I. (2013, June). Individual differences in speakers' gesture spaces: Multi-angle views from a motion-capture study. In *Proceedings of the Tilburg Gesture Research Meeting (TiGeR)* (pp. 19-21).

Pyers, J. E., Perniss, P., & Emmorey, K. (2015). Viewpoint in the visual-spatial modality: The coordination of spatial perspective. *Spatial Cognition & Computation*, *15*(3), 143-169.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, *7*(4), 226-231.

Richmond, V. P., McCroskey, J. C., & Johnson, A. D. (2003). Development of the nonverbal immediacy scale (NIS): Measures of self-and other-perceived nonverbal immediacy. *Communication Quarterly*, *51*(4), 504-517.

Riding, R. J., & Ashmore, J. (1980). Verbaliser-imager learning style and children's recall of information presented in pictorial versus written form. *Educational Studies*, *6*(2), 141-145.

Riding, R., Burton, D., Rees, G., & Sharratt, M. (1995). Cognitive style and personality in 12-year-old children. *British Journal of Educational Psychology*, *65*(1), 113-124.

Roth, W. M. (2001). Gestures: Their role in teaching and learning. *Review of Educational Research*, *71*(3), 365-392.

Roth, W. M., & Welzel, M. (2001). From activity to gestures and scientific language. *Journal of Research in Science Teaching*, *38*(1), 103-136.

Rousseaux, M., Rénier, J., Anicet, L., Pasquier, F., & Mackowiak-Cordoliani, M. A. (2012). Gesture comprehension, knowledge, and production in Alzheimer's disease. *European Journal of Neurology*, *19*(7), 1037-1044.

Rowe, M. L., Özçalışkan, Ş., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language*, *28*(2), 182-199.

Rowe, M. L., & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science*, *12*(1), 182-187.

Rozga, A., Hutman, T., Young, G. S., Rogers, S. J., Ozonoff, S., Dapretto, M., & Sigman, M. (2011). Behavioral profiles of affected and unaffected siblings of children with autism: Contribution of measures of mother–infant interaction and nonverbal communication. *Journal of Autism and Developmental Disorders*, *41*(3), 287-301.

Rueckert, L., Church, R. B., Avila, A., & Trejo, T. (2017). Gesture enhances learning of a complex statistical concept. *Cognitive Research: Principles and Implications*, *2*(1), 2.

Length, R. (2020). emmeans: Estimated Marginal Means. R package version 1.4.8. https://CRAN .R-project.org/package=emmeans.

Sassenberg, U., Foth, M., Wartenburger, I., & van der Meer, E. (2011). Show your hands—Are you really clever? Reasoning, gesture production, and intelligence. *Linguistics*, *49*(1), 105-134.

Sauer, E., Levine, S. C., & Goldin-Meadow, S. (2010). Early gesture predicts language delay in children with pre-or perinatal brain lesions. *Child Development*, *81*(2), 528-539.

Schmalenbach, S. B., Billino, J., Kircher, T., van Kemenade, B. M., & Straube, B. (2017). Links between gestures and multisensory processing: individual differences suggest a compensation mechanism. *Frontiers in Psychology*, *8*, 1828.

Schubotz, L., Özyürek, A., & Holler, J. (2019). Age-related differences in multimodal recipient design: younger, but not older adults, adapt speech and co-speech gestures to common ground. *Language, Cognition and Neuroscience*, *34*(2), 254-271.

Schubotz, L., Holler, J., Drijvers, L., & Özyürek, A. (2020). Aging and working memory modulate the ability to benefit from visible speech and iconic gestures during speech-in-noise comprehension. *Psychological Research*, 1-15.

Sekine, K., & Kita, S. (2015). Development of multimodal discourse comprehension: Cohesive use of space by gestures. *Language, Cognition and Neuroscience*, *30*(10), 1245-1258.

Sekine, K., Sowden, H., & Kita, S. (2015). The development of the ability to semantically integrate information in speech and iconic gesture in comprehension. *Cognitive Science, 39*(8), 1855-1880.

Sekine, K., Schoechl, C., Mulder, K., Holler, J., Kelly, S., Furman, R., & Özyürek, A. (2020). Evidence for children's online integration of simultaneous information from speech and iconic gestures: an ERP study. *Language, Cognition and Neuroscience*, 1-12.

Sheehan, E. A., Namy, L. L., & Mills, D. L. (2007). Developmental changes in neural activity to familiar words and gestures. *Brain and Language, 101*(3), 246-259.

Silverman, L. B., Bennetto, L., Campana, E., & Tanenhaus, M. K. (2010). Speech-and-gesture integration in high functioning autism. *Cognition*, *115*(3), 380-393.

Ska, B. & Croisile, B. (1998). Effects of normal aging on the recognition of gestures. *Brain and Cognition*, 37 (1), 136–138.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain and Language*, *101*(3), 260-277.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2009). Gestures orchestrate brain networks for language understanding. *Current Biology*, *19*(8), 661-667.

Smithson, L., Nicoladis, E., & Marentette, P. (2011). Bilingual children's gesture use. *Gesture*, *11*(3), 330-347.

Smithson, L., & Nicoladis, E. (2013). Verbal memory resources predict iconic gesture use among monolinguals and bilinguals. *Bilingualism: Language and Cognition*, *16*(4), 934-944.

So, W. C., Kita, S., & Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand. *Cognitive Science*, *33*(1), 115-125.

So, W. C. (2010). Cross-cultural transfer in gesture frequency in Chinese–English bilinguals. *Language and Cognitive Processes*, *25*(10), 1335-1353.

So, W. C., Yi-Feng, A. L., Yap, D. F., Kheng, E., & Yap, J. M. M. (2013). Iconic gestures prime words: comparison of priming effects when gestures are presented alone and when they are accompanying speech. *Frontiers in Psychology*, *4*.

So, W. C., Shum, P. L. C., & Wong, M. K. Y. (2015). Gesture is more effective than spatial language in encoding spatial information. *The Quarterly Journal of Experimental Psychology*, *68*(12), 2384-2401.

Sowden, H., Clegg, J., & Perkins, M. (2013). The development of co-speech gesture in the communication of children with autism spectrum disorders. *Clinical Linguistics & Phonetics*, *27*(12), 922-939.

Spence, C., & Driver, J. (Eds.). (2004). *Crossmodal space and crossmodal attention*. Oxford University Press.

Stanfield, C., Williamson, R., & Özçalışkan, Ş. E. Y. D. A. (2014). How early do children understand gesture–speech combinations with iconic gestures? *Journal of Child Language*, *41*(2), 462-471.

Stefanini, S., Caselli, M. C., & Volterra, V. (2007). Spoken and gestural production in a naming task by young children with Down syndrome. *Brain and Language*, *101*(3), 208-221.

Stone, W. L., Ousley, O. Y., Yoder, P. J., Hogan, K. L., & Hepburn, S. L. (1997). Nonverbal communication in two-and three-year-old children with autism. *Journal of Autism and Developmental Disorders*, *27*(6), 677-696.

Straube, B., Green, A., Jansen, A., Chatterjee, A., & Kircher, T. (2010). Social cues, mentalizing and the neural processing of speech accompanied by gestures. *Neuropsychologia*, *48*(2), 382-393.

Straube, B., Green, A., Weis, S., & Kircher, T. (2012). A supramodal neural network for speech and gesture semantics: an fMRI study. *PloS One*, *7*(11), e51207.

Streeck, J. (2009). Depicting gestures: Examples of the analysis of embodied communication in the arts of the West. *Gesture*, *9*(1), 1-34.

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, *55*(4), 661-699.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257-285.

Swisher, M. V., Christie, K., & Miller, S. L. (1989). The reception of signs in peripheral vision by deaf persons. *Sign Language Studies*, *63*(1), 99-125.

Talbott, M. R., Nelson, C. A., & Tager-Flusberg, H. (2015). Maternal gesture use and language development in infant siblings of children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *45*(1), 4-14.

Talmy L. (1983). How language structures space. In *Spatial Orientation: Theory, Research, and Application*, ed. H Pick, L Acredelo, pp. 225–82. New York: Plenum

Tamis-LeMonda, C. S., Song, L., Leavell, A. S., Kahana-Kalman, R., & Yoshikawa, H. (2012). Ethnic differences in mother–infant language and gestural communications are associated with specific skills in infants. *Developmental Science*, *15*(3), 384-397.

Thompson, L. A. (1995). Encoding and memory for visible speech and gestures: A comparison between young and older adults. *Psychology and Aging*, *10*(2), 215.

Thompson, L. A., & Guzman, F. A. (1999). Some limits on encoding visible speech and gestures using a dichotic shadowing task. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *54*(6), 347-349.

Trafton, J. G., Trickett, S. B., Stitzlein, C. A., Saner, L., Schunn, C. D., & Kirschenbaum, S. S. (2006). The relationship between spatial transformations and iconic gestures. *Spatial Cognition and Computation*, *6*(1), 1-29.

Trujillo, J. P., Simanova, I., Bekkering, H., & Özyürek, A. (2018). Communicative intent modulates production and comprehension of actions and gestures: A Kinect study. *Cognition*, *180*, 38-51.

Turan, E., Kobaş, M., & Göksun, T. (in press). Spatial language and mental transformation in preschoolers: Does relational reasoning matter? *Cognitive Development*.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127-154.

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist*, *30*(2), 128.

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology*, *28*(2), 187-204.

Vanetti, E. J., & Allen, G. L. (1988). Communicating environmental knowledge: The impact of verbal and spatial abilities on the production and comprehension of route directions. *Environment and Behavior*, *20*(6), 667-682.

van Wermeskerken, M., Fijan, N., Eielts, C., & Pouw, W. T. (2016). Observation of depictive versus tracing gestures selectively aids verbal versus visual–spatial learning in primary school children. *Applied Cognitive Psychology*, *30*(5), 806-814.

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing, and evolution. *Phil. Trans. R. Soc. B* 369: 20130292.

Vogel, E. K., & Awh, E. (2008). How to exploit diversity for scientific gain: Using individual differences to constrain cognitive theory. *Current Directions in Psychological Science*, *17*(2), 171-176.

Vogt, S., & Kauschke, C. (2017). Observing iconic gestures enhances word learning in typically developing children and children with specific language impairment. *Journal of Child Language*, *44*(6), 1458-1484.

Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language*, *50*(4), 395-407.

Wakefield, E. M., James, T. W., & James, K. H. (2013). Neural correlates of gesture processing across human development. *Cognitive Neuropsychology*, *30*(2), 58-76.

Wakefield, E., Novack, M. A., Congdon, E. L., Franconeri, S., & Goldin-Meadow, S. (2018). Gesture helps learners learn, but not merely by guiding their visual attention. *Developmental Science*, *21*(6), 12664.

Wartenburger, I., Kühn, E., Sassenberg, U., Foth, M., Franz, E. A., & Van der Meer, E. (2010). On the relationship between fluid intelligence, gesture production, and brain structure. *Intelligence*, *38*(1), 193-201.

Watson, L. R., Crais, E. R., Baranek, G. T., Dykstra, J. R., & Wilson, K. P. (2013). Communicative gesture use in infants with and without autism: A retrospective home video study. *American Journal of Speech-Language Pathology*, 22(1), 25-39.

Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore, MD: Williams & Wilkins. doi: 10.2307/1423612.

Wermelinger, S., Gampe, A., Helbling, N., & Daum, M. M. (2020). Do you understand what I want to tell you? Early sensitivity in bilinguals' iconic gesture perception and production. *Developmental Science*, e12943.

Werner, S., & Noppeney, U. (2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *Journal of Neuroscience*, *30*(7), 2662-2675.

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *The American Journal of Psychology*, *114*(4), 591.

Willems, R. M., Özyürek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex*, *17*(10), 2322-2333.

Willems, R. M., Özyürek, A., & Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *NeuroImage*, *47*(4), 1992-2004.

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, *42*(6), 654-667.

Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain and Language*, *101*(3), 234-245.

Wu, Y. C., & Coulson, S. (2010). Gestures modulate speech processing early in utterances. *Neuroreport*, *21*(7), 522.

Wu, Y. C., & Coulson, S. (2014a). Co-speech iconic gestures and visuo-spatial working memory. *Acta Psychologica*, *153*, 39-50.

Wu, Y. C., & Coulson, S. (2014b). A psychometric measure of working memory capacity for configured body movement. *PloS One*, *9*(1).

Yap, D. F., So, W. C., Melvin Yap, J. M., Tan, Y. Q., & Teoh, R. L. S. (2011). Iconic gestures prime words. *Cognitive Science*, *35*(1), 171-183.

Yeo, A., & Alibali, M. W. (2017). Evidence for overt visual attention to hand gestures as a function of redundancy and speech disfluency. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Yeo, A., Ledesma, I., Nathan, M. J., Alibali, M. W., & Church, R. B. (2017). Teachers' gestures and students' learning: sometimes "hands off" is better. *Cognitive Research: Principles and Implications*, *2*(1), 41.

Yeo, A., & Alibali, M. W. (2018). Does visual salience of action affect gesture production? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(5), 826.

Yeo, L. M., & Tzeng, Y. T. (2019). Tracing effect in the worked examples-based learning: An exploration of individual differences in working memory capacity. *EURASIA Journal of Mathematics, Science and Technology Education*, *15*, 10.

Yeung, A. S., Jin, P., & Sweller, J. (1998). Cognitive load and learner expertise: Split-attention and redundancy effects in reading with explanatory notes. *Contemporary Educational Psychology*, *23*(1), 1-21.

APPENDIX

Action primes along with Turkish labels.

Hammering a nail "çak"

Chopping vegetables "doğra"

Twisting the lid of a jar "çevir"

Dialing a phone "tuşla"

Wiping "sil"

Cutting paper "kes"

Stirring soup "karıştır"

Scrubbing a dish "yıka"

Typing "yaz"

Turning a key "döndür"

LIST OF PUBLICATIONS

Özer, D., Karadöller, D., Göksun, T., & Özyürek, A. (in preparation). Informativeness of gestures guide visual attention during spatial language comprehension.

Özer, D. & Göksun, T. (2020). Gesture use and processing**:** a review on individual differences in cognitive resources. *Frontiers in Psychology,* 11: 573555. **https://doi.org/10.3389/fpsyg.2020.573555**

Özer, D. & Göksun, T. (2020). Visual-spatial and verbal abilities differentially affect processing of gestural vs. spoken expressions. *Language, Cognition and Neuroscience, 35(7),* 896-914. **https://doi.org/10.1080/23273798.2019.1703016**

Özer, D., Göksun, T., & Chatterjee, A. (2019). Differential roles of gestures on spatial language in neurotypical elderly adults and individuals with focal brain injury. *Cognitive Neuropsychology*, 36:5-6, 282-299. **https://doi.org/10.1080/02643294.2019.1618255**

Özer, D., Tansan, M., Özer, E. E**.,** Malykhina, K., Chatterjee, A., & Göksun, T. (2017). The effects of gesture restriction on spatial language in young and elderly adults. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1471-1476). Austin, TX: Cognitive Science Society.

Özer, D., Kranjec, A., Balcı, F., & **Göksun, T.** (2017). The effects of duration words and spatial-temporal metaphors on perceived duration In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1477-1482). Austin, TX: Cognitive Science Society.

Göksun, T., **Özer, D.,** & Akbıyık, S. (in press). *Gesture in the Aging Brain.* In A. Morgenstern & S. Goldin-Meadow (Eds.), Gesture in Language. De Gruyter Mouton

Göksun, T. & **Özer, D.** (2016) Nesnelerdeki Değişim ve Yer-Yön Kavramı: Çocuklarda Mekansal Uzamsal Algının Gelişimsel Süreçleri [Children's development on spatial cognition]. *Aklın Çocuk Hali: Zihin Gelişimi Araştırmaları* (Ed. Aydın, Ç., Göksun, T., Küntay, A. & Tahiroğlu, D.) İstanbul: Koç Üniversitesi Yayınları

LIST OF PRESENTATIONS

*Asterisk denotes oral presentations

***Özer, D.,** Karadöller, D.Z., Türkmen, İ., Özyürek, A., & Göksun, T. (2020, September). Informativeness of gestures in speech context guides visual attention during comprehension of spatial language. The talk given by the first author at the 7th Gesture and Speech in Interaction (GESPIN) Conference.

Karadöller, D., Ünal, E., Sümer, B., **Özer, D.,** Göksun, D., & Özyürek, A. (2019, November). Children but not adults use both speech and gesture to produce informative expressions of Left-Right relations. The talk given by the first author at the 44th Annual Boston University Conference on Language Development, Boston, USA.

Özer, D. & Göksun, T. (2019, March). The Role of Working Memory in the Online Speech-Gesture Integration. Poster presented at the 2019 International Convention of Psychological Science, Paris, France.

***Özer, D.** & Göksun, T. (2018, November). The role of working memory in speech-gesture integration. The talk given at the 20th Conference of the National Psychological Association, Ankara, Turkey.

Özer, D., Chatterjee, A. & Göksun, T. (2018, November). The differential role of gestures in people with and without speech impairments. Poster presented at the 20th Conference of the National Psychological Association, Ankara, Turkey.

***Özer, D.,** Chatterjee, A. & Göksun, T. (2018, July). Differential roles of gestures on spatial language in patients with speech deficits and elderly. The talk given at the 8th Conference of the International Society for Gestural Studies, Cape Town, South Africa.

Tansan, M., **Özer, D.,** Ilıca, D., & Göksun, T. (2018, July). Individual differences in gesture production during spatial tasks. Poster presented at the 8th Conference of the International Society for Gestural Studies, Cape Town, South Africa.

Özer, D., Tansan, M., Özer, E. E.**,** Malykhina, K., Chatterjee, A., & Göksun, T. (2017, July). The effects of gesture restriction on spatial language in young and elderly adults. Poster presented at *38th Annual Conference of the Cognitive Science Society, London, UK.*

Özer, D., Kranjec, A., Balcı, F., & Göksun, T. (2017, July). The effects of duration words and spatial-temporal metaphors on perceived Poster presented at the *38th Annual Conference of the Cognitive Science Society, London, UK.*

Özer, D., Chatterjee, A. & Göksun, T. (2017, June) *The Effects of Medium of Encoding on Silent Gestures: Evidence from Young and Elderly Adults.* Poster presented at Language as a Form of Action Conference, DComm Project, Rome, Italy.

Akhavan, N., **Özer, D**., Altınok, N., Ger, E., Lizskowski, U., Küntay, A. & Göksun, T. (2015, January) *Infant point following predicts later language development: Evidence from Turkish.* Poster presented at 2015 Budapest CEU Conference on Cognitive Development, Budapest, Hungary.

Özer, D., Akhavan, N., Altınok, N., Savas, O., Lizskowski, U., Küntay, A. & Göksun, T. (2015, March) *Pointing the right way: The role of infant and caregiver gestures on language development in Turkish*. Poster presented at 2015 Biannual Meeting of Society for Research in Child Development, Philadelphia, PA.

Özer, D., Bostan, I., Chatterjee, A. & Göksun, T. (2015, March) *Describing and gesturing causal events : Evidence from focal brain-injured patients.* Poster presented at 2015 Annual meeting of Cognitive Neuroscience Society, San Fransisco, CA.

Özer, D., Bostan, I., Chatterjee, A. & Göksun, T. (2015, July) *Describing and gesturing causal events: Evidence from focal brain injury*. Poster presentation at 2015 Cognitive Science Conference, Pasadena, CA.