Effects of aging and cognitive abilities on multimodal language production and comprehension in context

Louise Schubotz

Cover design by Bart Versteegen
Artwork cover and p. 200 by Ehrich Ohser / e.o. plauen (2003). *Vater und Sohn. Sämtliche Streiche und Abenteuer*. Südverlag GmbH, Konstanz.
Printed and bound by Ipskamp Drukkers b.v.

The research reported in this thesis was supported by a pre-doctoral scholarship from the Max Planck International Research Network on Aging (MaxNetAging) awarded to Louise Schubotz.

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

More information can be found at www.mpi.nl/imprs

Effects of aging and cognitive abilities on multimodal language production and comprehension in context

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van de decanen
in het openbaar te verdedigen op maandag 17 mei 2021
om 16:30 uur precies

door

Louise Marlene Rosa Schubotz
geboren op 24 juli 1985
te Berlijn, Duitsland

Promotor

Prof. dr. H.A. Özyürek

Copromotor

Dr. J. Holler

Manuscriptcommissie

Prof. dr. A.S. Meyer

Dr. E. Janse

Prof. dr. S. Kita

University of Warwick, Verenigd Koninkrijk

Für Anne und Winni

Table of Contents

Chapter 1

General Introduction

Human communication is inherently multimodal. In our daily interactions, we use our entire body to convey our messages to others: In addition to speech, we use manual gestures, facial expressions, body posture, or gaze direction in order to enhance the meaning of our utterances, or to help us coordinate our interactions more generally (e.g., de Ruiter, 2007; Kendon, 2004; McNeill, 1992). However, how do everyday communicative interactions change as we grow older? Are there age-related differences in how we use multimodal language? How successful are older adults in communicating with and comprehending others? Recent years have seen a growing interest in the study of language production and comprehension in normal human aging. Yet, we still know little about how aging affects language use in everyday, face-to-face interactions (e.g., Abrams & Farrell, 2011; Thornton & Light, 2006). This lack of knowledge extends to older adults' use of the gestural modality, a core component of language in interactive settings (Bavelas & Chovil, 2000; Clark, 1996; Kendon, 2004; McNeill, 1992). Considering the importance of everyday interactions for social relations, we are thus faced with a serious gap in our understanding of the multimodal communicative practices and competences of older adults, in language production and in language comprehension.

Back in 2011, when I wrote the initial proposal on "Gesture use for social interaction in older adults" that would eventually lead to this doctoral thesis, there were exactly five studies on gesture use in aging that informed my hypotheses (i.e., Cohen & Borsoi, 1996 and Feyereisen & Havard, 1999 on gesture production and Cocks, Morgan, & Kita, 2011, Thompson, 1995, and Thompson & Guzman, 1999 on gesture comprehension). In the meantime, a couple more studies on this topic have been published (e.g., Theocharopoulou, Cocks, Pring, & Dipper, 2015; Arslan & Göksun, 2020) and a more elaborate image of gesture use in aging is starting to emerge. However, how the specific communicative context in which an interaction occurs, such as for example face-to-face interaction, affects older adults' multimodal language use still remains largely unknown. Furthermore, the potential consequences of age-associated changes in cognitive functioning, such as for example reduced working memory (WM) capacity and decreased inhibitory control (e.g., Salthouse, 2010; Hasher & Zacks, 1988; but cf. Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014) for these processes and how they interact with situated uses of multimodal language use are similarly unknown. In order to bridge some of these gaps, this thesis investigates whether and how the production and comprehension of speech and co-speech gestures are affected by normal aging, with a focus on the context in which language production and comprehension occurs and the role of age-associated changes in cognitive functioning in these processes.

1.1. Age-related changes in spoken language production and comprehension

Previous research has provided evidence that older adults' ability to produce and comprehend spoken language systematically differs from that of younger adults (e.g., Abrams & Farrell, 2011; Thornton & Light, 2006). This is often attributed to age-related differences in social or communicative goals (e.g., Horton and Spieler, 2007; James, Burke, Austin, & Hulme, 1998; Underwood, 2010), or to age-related changes in basic perceptual or cognitive functions (e.g., Burke & Shafto, 2008; Thornton & Light, 2006; Schneider & Pichora-Fuller, 2000) as well as changes in neuro-cognitive functions (e.g., Marini & Andreetta, 2016; Peelle, 2019, for recent overviews).[1] As we will see, one aspect will prove to be key in understanding older adults' communicative behavior: the context, in which language use occurs. By this, I refer to the specifics of the communicative situation, which may greatly affect not only older adults' communicative goals, but also the perceptual and cognitive challenges associated with the situation. For language production and for language comprehension, different contextual factors may play a role. In the following section, I will thus shortly summarize the main findings on spoken language production and comprehension in aging, focusing on interactive language use and highlighting the role of contextual and cognitive factors.

1.1.1. Spoken language production in aging

During language production, older adults often display significant deficits compared to younger adults. For example, there are age-related difficulties in lexical retrieval, indicated by an increase in tip-of-the-tong states (e.g., Brown & Nix, 1996; Burke, MacKay, Worthley, & Wade, 1991; Maylor, 1990), less accurate and slower picture naming (e.g., Feyereisen, 1997), or an increase in dysfluencies (e.g., Bortfeld, Leon, Bloom, Schober, & Brennan, 2001). However, age-related difficulties in lexical retrieval appear to be less pronounced in the context of connected speech relative to single word production (see Kavé & Goral, 2007), potentially, because grammatical and semantic context aid retrieval. Older adults also have been found to produce fewer complex

[1] Note that while cross-sectional data suggests that the decline of certain cognitive functions starts as early as age 20 (e.g., Salthouse, 2009), longitudinal data on cognitive aging presents a more optimistic picture, suggesting that on an individual level, decline is both less severe and occurs later on in life, starting around age 60 (e.g., Rönnlund, Nyberg, Bäckman, & Nilson, 2005; for a discussion see Nilson, Sternäng, Rönnlund, & Nyberg, 2009).

sentences relative to younger adults (Kemper, Herman, & Liu, 2004; Kemper & Sumner, 2001; Marini, Boewe, Caltagirone, & Carlomagno, 2005) and to produce less coherent discourse than younger adults on measures of global (e.g., Glosser & Deser, 1992) and local cohesion (e.g., Marini et al., 2005). Yet, during socially-driven language use, such as when telling narratives, older adults are actually rated as high or higher than younger adults on features such as story quality, interest, clarity, and informativeness (e.g., James, Burke, Austin, & Hulme, 1998; Kemper, Kynette, Rash, O'Brien, & Sprott, 1989; Pratt & Robins, 1991; see also Thornton & Light, 2006). This suggests that age-related differences in language production greatly depend on the social and communicative context in which it occurs, as I will elaborate on in the following section.

Spoken language production in aging: The role of contextual factors

One main feature of language used in an interactive context is that it is directed at minimally one addressee. Crucially, this requires the adaptation of utterances according to the addressee's communicative needs, a process termed *recipient design* (Sacks, Schegloff, & Jefferson, 1974) or *audience design* (Clark & Murphy, 1983). Previous research suggests that older adults are able to engage in recipient design based general addressee characteristics like an addressee's age or mental ability, which require a global adaptation of speech style: in terms of overall content and complexity, older adults adjusted their spoken utterances when talking to a child versus an adult (Keller-Cohen, 2014), even to a larger extent than younger adults (Adams, Smith, Pasupathi, & Vitolo, 2002). Similarly, older women adapted their speech style according to whether they interacted with a female college student versus a person with mild mental retardation (Gould & Shaleen, 1999). However, older adults have difficulties with more local, moment-by-moment, fine-grained adaptations that involve the *common ground* between a speaker and an addressee. This term refers to the knowledge, beliefs, and assumptions that conversational partners believe to be mutually shared (Clark, 1996). Depending on the source of this mutually shared knowledge, different types of common ground can be distinguished (Clark, 1996): 1) *Communal common ground* (pp. 100-112), which refers to the knowledge shared in cultural or sub-cultural communities; 2) *Personal common ground* (pp. 112-116), which describes the knowledge shared between particular interlocutors as a result of their prior common experience (such as e.g. past conversations) or their current situation (including visual co-presence); and 3) *Incremental common ground* (pp. 38-39, 221-251), which refers to the mutually shared knowledge between interlocutors that accumulates in the course of an interaction via a process termed *grounding* (Clark & Brennan, 1991), i.e., "the moment-by-moment exchanges that establish information as being in common ground within a conversation" (Holler & Bavelas, 2017, p. 214), thus comprising the information exchanged and successfully understood during the interaction. Generally, the larger the common ground between conversational partners, the more efficient the interaction becomes. In younger adults, this is often characterized by shorter utterances, less complex syntax, or less informational content (e.g. Fussel & Krauss, 1992; Galati & Brennan, 2010; Isaacs & Clark, 1987). Older adults appear to be less efficient at these adaptations. For example, in referential communication tasks which involve the establishment of mutual reference to a limited set of objects over the course of several trials (a form of incremental common ground), older adults produced longer utterances, more errors, and had longer task completion times relative to younger adults (Horton & Spieler, 2007; Hupet, Chartraine, & Nef, 1993; Lysander & Horton, 2012). In tasks requiring the unambiguous identification of referents based on visual scenes, older adults were similarly less efficient than younger adults, indicated by unnecessary over-informativeness (Saryazdi, Bannon, & Chambers, 2019; however, see also Long, Rohde, & Rubio-Fernandez, 2020). Thus, older adults' ability to adjust their speech based on common ground, either established in the course of an interaction, or based on shared visual information (a form of personal common ground), appears to be reduced relative to younger adults.

In part, these age-related differences in recipient design may be due to age-related differences in social and communicative goals: For example, providing information to younger generations may be an important goal for older adults, therefore older adults may be particularly motivated to adapt their utterances for children and less motivated to adapt their utterances for other adults (e.g., Adams et al., 2002). In addition, older adults may also be less motivated to interact with strangers relative to younger adults, instead favoring existing relationships (e.g., Fung, Carstensen, & Lutz, 1999), which may in turn affect the extent of their addressee-based adaptations (Horton & Spieler, 2007). However, these different communicative contexts may also be associated with differences in cognitive demands, which may in turn affect older adults behavior, as we will see in the following section.

Spoken language production in aging: The role of cognitive factors

In addition to the contextual factors that may influence older adults' interactive communicative behavior, age-related changes in cognitive functioning are also likely to modulate interactive language use in general and the addressee-based adaptations of verbal utterances in particular. Specifically, the patterns described above may be attributable to age-related changes in memory functions, including working memory

(WM), and to deficits in inhibitory control, that might interfere with interactive aspects of language use in context.

WM, which is assumed to have a verbal and a visual component, allowing for the temporary maintenance and manipulation of verbal and visual information respectively (Baddeley & Hitch, 1974), plays a significant role when speakers are required to take an addressee's perspective into account while formulating their utterances (Wardlow, 2013; Healey & Grossmann, 2016). Recipient design in conversation crucially requires the ability to incorporate the addressee's perspective during online language processing (see e.g. Brennan, Galati, & Kuhlen, 2010). Local adaptations to aspects like the common ground between interlocutors, particularly incremental common ground, presumably relies heavily on (working) memory resources, as it requires constant updating and incorporating of what knowledge is mutually shared or not, suggesting that age-related changes in WM capacity may affect older adults' ability to engage in common ground-based recipient design (Horton & Spieler, 2007). Global adaptions to aspects like an addressee's age, on the other hand, may be less dependent on (working) memory resources, but rather build on an initial assessment of the addressee as being for example a child and then applying an internalized set of adaptations to one's speech, deemed to be appropriate to address this type of addressee (Keller-Cohen, 2014).

The second cognitive function, inhibitory control, has been related to the ability to inhibit irrelevant, egocentric information from entering memory (Hupet et al., 1993; see also Hasher and Zacks, 1988) and to perspective-taking abilities (Long et al., 2018; Wardlow, 2013). The ability to inhibit one's own, egocentric perspective in favor of the addressee's is another crucial component of successful verbal recipient design (Brennan et al., 2010; Keysar, Barr, & Horton, 1998). Older adults' difficulties with engaging in recipient design based on personal or incremental common ground may therefore also be connected to age-related difficulties in adopting the addressee's perspective at any given moment in the interaction. General addressee characteristics like age or mental retardation, on the other hand, may serve as a constant reminder and strong motivation to adopt the addressee's perspective, thus making this type of addressee-based adaptation less dependent on inhibitory control.

To summarize, in addition to age-related differences in social or communicative goals, age-related differences in WM and inhibitory control likely also contribute to older adults being able to adapt their speech to addressee characteristics, like age, but having difficulties adapting to fine-grained aspects, like personal and incremental common ground.

1.1.2. Spoken language comprehension in aging

While there are clear deficits in language production, spoken language comprehension appears to be relatively preserved in aging, although there are certain deficits attributable to age-related hearing loss (presbycusis) which may affect as much as one third of adults aged 65 or older (e.g., Pichora-Fuller & Singh, 2006). Consequences of age-related hearing loss include for example difficulties in discourse comprehension in quiet and in noise (Schneider, Daneman, Murphy, & See, 2000; Schneider, Daneman, & Pichora-Fuller, 2002). Even mild to moderate hearing loss can have significant negative effects on older adults' ability to comprehend language, particularly in challenging settings, and put them at a greater risk for social isolation (Pichora-Fuller, Alain, & Schneider, 2017). As for language production in aging, the context in which speech is perceived plays a crucial role in older adults' language comprehension.

Spoken language comprehension in aging: The role of contextual factors

In fact, contextual factors that frequently accompany every-day listening situations may influence speech comprehension in older adults to a greater extent than in younger adults. For example, older adults' ability to understand speech in noisy surroundings, such as other conversations or sounds in the background, is more strongly impaired than that of younger adults (e.g., Dubno, Dirks, & Morgan, 1984; Pichora-Fuller, Schneider, & Daneman, 1995). This can be attributed in part to age-related hearing loss, but also to age-related changes in cognitive functioning (e.g. Anderson, White-Schwoch, Parbery-Clark, & Kraus, 2013; CHABA, 1988; Humes, 2002, 2007; Humes, Watson, Christensen, Cokely, Halling, & Lee, 1994; Pichora-Fuller et al., 2017, see also next section). At the same time, older adults' speech comprehension improves significantly if the speech signal is presented in a context that aids speech processing or interpretation: for example, both younger and older adults benefit if visual phonological information, such as the speaker's articulatory lip movements, or *visible speech*, is available (e.g., Sommers, Tye-Murray, & Spehar, 2005; Stevenson, Nelms, Baum, Zurkovsky, Barense, Newhouse, & Wallace, 2015; Tye-Murray, Sommers, Spehar, Myerson, & Hale, 2010; Tye-Murray, Spehar, Myerson, Hale, & Sommers, 2016). Also, older adults' speech comprehension in noise has been shown to benefit as much or even more from additional semantic information, such as sentence context, than younger adults' (e.g. Pichora-Fuller et al., 1995; Smayda, Van Engen, Maddox, & Chandrasekaran, 2016). The fact that in most everyday occurrences of face-to-face, interactive language use both the speaker's articulatory lip movements and a semantic and/or sentence context are available, suggests that older adults may often be able to compensate for age-related deficits in speech comprehension during their daily interactions, at least to some extent.

Spoken language comprehension in aging: The role of cognitive factors

The effects of these contextual factors on older adults' speech comprehension are likely to be modulated by cognitive factors. As for language production, two cognitive abilities in particular have been associated with the comprehension of speech-in-noise (SiN): WM and inhibitory control. Verbal WM capacity predicts comprehension and/or recall of SiN in older adults (Baum & Stevenson, 2017; Koeritzer, Rogers, Van Engen, & Peelle, 2018; Rudner, Mishra, Stenfelt, Lunner, & Rönnberg, 2016), potentially, because additional WM resources are recruited for the auditory processing of the speech signal, leaving fewer resources for subsequent processes related to language comprehension. Inhibitory control, or the ability to selectively focus attention or to suppress irrelevant information, has been connected to the comprehension of single talker speech presented against the background of several other talkers (i.e., multitalker babble, e.g. Janse, 2012; Jesse & Janse 2012; Tun, O'Kane, & Wingfield, 2002). These cognitive effects may in turn be modulated or alleviated by contextual factors. The phonological and semantic information provided by visible speech and sentence context respectively may reduce the processing demands of speech, help focus on the target speaker, and thereby facilitate perception and comprehension.

1.1.3. Interim summary

Based on the literature summarized in the previous paragraphs, I would like to propose that older adults' interactive spoken language production and comprehension is determined by an interplay of cognitive and contextual factors. It is currently unclear whether the age-related behavioral differences also extend to the use of the gestural modality, a core component of interactive language use (e.g., Kendon, 2004; McNeill, 1992), and whether the use of this additional modality is guided by similar principles. The present thesis' focus is on investigating this issue, with the aim to advance our understanding of multimodal communication in older adults and the influence of cognitive abilities on these processes.

1.2. Age-related changes in multimodal language production and comprehension

The remainder of this introductory chapter is dedicated to multimodal language use. Unless otherwise specified, this refers to the production and comprehension of speech and accompanying manual co-speech gestures. Based on the tight relationship between speech and co-speech gestures that I will elaborate on below, gestures are generally considered to be an integral part of the language system (Kendon, 2004; McNeill, 1992; Kelly, Özyürek, & Maris, 2010). In the following, I will first provide a short definition of co-speech gestures and the different gesture types relevant to the work presented in this thesis. Then, I will summarize the main findings on co-speech gesture production and comprehension in younger and in older adults, with a special focus on contextual factors and cognitive factors. This will lead me to the research questions and hypotheses that motivated the empirical chapters presented in this thesis.

1.2.1. Co-speech gestures – definition and types

Broadly speaking, co-speech gestures are the meaningful movements we make with our hands and arms while we speak (McNeill, 1992). Out of the several visual signals that accompany speech in face-to-face settings, including (but not limited to) manual gestures, articulatory lip movements, facial expressions, body posture, or gaze direction, manual co-speech gestures in particular contribute substantially to the semantic and pragmatic aspects of a speaker's message, and are tightly coordinated semantically and temporally with the speech they accompany (e.g., Kendon, 2004; McNeill, 1992).

While all co-speech gestures are related to the speech that they accompany in one way or another, several distinct gesture types can be identified, based on the nature of this speech-gesture relationship. For example, gestures can be used to depict the shape or size of concrete referents or to represent physical movements or actions (*iconic gestures*), to metaphorically express abstract concepts (*metaphoric gestures*), to single out referents in the environment or in fictive space through deixis (*deictic* or *pointing gestures*), to add emphasis to certain elements of speech (*beat gestures*), or to coordinate communicative interactions more generally (*interactional* and *pragmatic gestures*; e.g., Alibali, Heath, & Myers, 2001; Bavelas, Chovil, Lawrie, & Wade, 1992; McNeill, 1992; Kendon, 2004). As I was mainly concerned with the semantic relationship between speech and co-speech gestures, the research presented in this thesis focuses on manual iconic (and to some extent also on metaphoric and pointing) gestures. In the remainder of this thesis, I will also use the term representational gestures to refer these gesture types collectively (see also Alibali et al., 2001).

1.2.2. Co-speech gesture production in younger adults

Turning to the co-speech gesture production in younger adults, we may confidently state that when people speak, they gesture. Speakers use their hands to talk most of the time, although they might not always be aware of it. The tendency to produce co-speech gestures is so high that speakers often gesture on the telephone, despite the fact that the person they are speaking with cannot see them, as we have probably all observed on multiple occasions (see also Bavelas, Gerwing, Sutton, & Provost, 2008). However, as with spoken language used interactively, gesture production is greatly affected by the communicative context and by cognitive factors, which I will elaborate on in the following sections.

Co-speech gesture production in younger adults: The role of contextual factors

The communicative context in which language is used clearly affects the production of co-speech gestures in younger adults, which can be aptly described as forms of audience or recipient design. For example, speakers gesture more frequently when they know that these gestures can be seen by their addressee as opposed to when the addressee cannot see the gestures (e.g. Alibali et al., 2001; Bavelas, Kenwood, Johnson, & Phillips, 2002; Mol, Krahmer, Maes, & Swerts, 2011). Beyond mutual visibility, relative gesture frequency is furthermore affected by dialogic interaction (e.g. Bavelas, et al., 2008) and addressee feedback (Jacobs & Garnham, 2007). In addition, addressee location and addressee feedback also influence how gestures are used to represent semantic information (e.g., Holler & Wilkin, 2011a; Kuhlen, Galati, & Brennan, 2012; Özyürek, 2002), and interactants may even engage in gestural mimicry in order to establish mutual reference to certain objects (Holler & Wilkin, 2011b). When the amount of personal common ground is manipulated, speakers often produce fewer and less informative gestures when they talk about content that their addressee is already familiar with as opposed to content that is new to the addressee (e.g., Gerwing & Bavelas, 2004; Hilliard & Cook, 2015; Holler & Stevens, 2007; Holler & Wilkin, 2009; Parrill, 2010). Hence, common ground-based effects on gestures often resemble those observed for speech (see section 1.1.1.). However, whether this reduction in gesture frequency is proportional to reductions in speech varies across studies, thus leading to different effects on gesture rate (e.g., Campisi & Özyürek, 2013; de Ruiter, Bangerter, &

Dings, 2012, Galati & Brennan, 2014; Hilliard & Cook, 2015; Hoetjes, Koolen, Goudbeek, Krahmer, & Swerts, 2015; see Holler and Bavelas, 2017, for an overview).

Co-speech gesture production in younger adults: The role of cognitive factors

Apart these contextual factors, cognitive factors also affect gesture production. Previous literature suggests that gesturing may aid the speech planning process, e.g. by activating relevant spatial imagery which may also aid in lexical retrieval (Kita, Alibali, & Chu, 2017; Krauss, Chen, & Gottesman, 2000). Additionally, gesturing may facilitate the organization or packaging of spatial information during utterance planning, such that speakers gesture more when they describe spatial information that is difficult to conceptualize or more complex (Hostetter, Alibali, & Kita, 2007; Kita, 2000; Kita et al., 2017; Kita & Davies, 2009; Melinger & Kita, 2007). Gesturing may also lighten the cognitive load more generally, by “off-loading” information that otherwise taxes cognitive resources onto visual space (e.g., Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Wagner, Yip, & Goldin-Meadow, 2012). For example, when explaining math problems while at the same time remembering a string of letters, speakers recall more letters when they are allowed to gesture than when they are not (Goldin-Meadow et al., 2001), an effect not obtained when meaningless hand movements instead of gestures are performed (Wagner et al., 2012). Added support for the cognitively beneficial effects of gesturing come from individual differences studies which report that limited cognitive abilities, in particular verbal and visual WM, lead to an increase in gesture frequency in a number of tasks (e.g. Chu, Meyer, Foulkes, & Kita, 2014; Gillespie, James, Federmeier, & Watson, 2014; Hostetter & Alibali, 2007).

Co-speech gesture production in younger adults: The interplay of contextual and cognitive factors

In recent years, a number of studies investigated the interplay of communicative context *and* cognitive factors on gesture production (e.g., Arslan & Göksun, 2020; Galati & Brennan, 2014; Hoetjes et al., 2015; Masson-Carro, Goudbeek, & Krahmer, 2016). For example, in a story narration task, Galati and Brennan (2014) manipulated communicative and cognitive factors simultaneously. Speakers had to retell the same story three times, twice to the same addressee and once to a new addressee. Speakers used fewer, smaller, and less precise gestures for knowing vs. unknowing addressees. At the same time, gesture size and precision were additionally affected by whether the speaker told the story for the first time (higher cognitive load for speaker) vs. for a second or third time (lower cognitive load for speaker). Speakers’ gestures were smaller and less precise in the lower cognitive load conditions, corroborating the idea that gesticulation can help manage cognitive load (see previous section). However, these speaker-oriented effects were less pronounced when the narration was targeted at an unknowing addressee, that is, the addressee-oriented goal to be more informative modulated the speaker-oriented effects on gesticulation. In another study, where the speaker’s cognitive load was manipulated by either having to retell a full seven-minute cartoon at once (higher cognitive load for speaker) vs. shorter episodes of the same cartoon one at a time (lower cognitive load for speaker), Mol, Krahmer, Maes, and Swerts (2009) found that participants produced fewer communicative gestures for an addressee under higher cognitive load, suggesting that the production of communicative gestures is actually cognitively costly. However, this effect was only present if speaker and addressee could see each other. While the results of these two studies may seem somewhat contradictory at first sight, the different ways in which cognitive load as well as the communicative situation, or context, were manipulated make direct comparisons difficult. What becomes clear, however, is that both cognitive and contextual factors influence how speakers use co-speech gestures in communication with others and that importantly, these factors are intertwined.

Theoretical accounts of co-speech gesture production

These two main functions of co-speech gestures summarized above – addressee-oriented, communicative functions on the one hand, and speaker-oriented, cognitive functions on the other hand – are also reflected in theoretical accounts of co-speech gesture production (for recent overviews see e.g. Galati & Brennan, 2014; Hoetjes et al., 2015; Özyürek, 2017, 2018). Although there is general agreement that co-speech gestures originate from visuo-spatial or motoric representations accessed from WM during the speech planning process (e.g., Hostetter & Alibali, 2008), accounts differ in their assumptions about whether co-speech gestures are communicatively intended by the speaker or whether they are simply a by-product of the speech planning process, and in the role that is attributed to cognitive factors.

From an addressee-oriented, communicative perspective, speakers use gestures with the intention to communicate relevant information to an addressee. This is acknowledged in models like the Growth point theory (McNeill, 1992; 2005; McNeill & Duncan, 2000), the Sketch model (de Ruiter, 2000), or the Interface hypothesis (Kita & Özyürek, 2003), which claim that speech and gesture originate from a shared conceptual level, i.e. a preverbal message or communicative intent. Growth point theory (McNeill, 1992; 2005; McNeill & Duncan, 2000) is a non-modular account which assumes that speech and gestures are inseparable aspects of one communicative intention and cannot be considered independently. In contrast, the Sketch model (de Ruiter, 2000) and the Interface hypothesis (Kita & Özyürek, 2003) are both modular accounts, building on Levelt's (1989) model of speech production and assuming an additional, separate production path for gestures. However, while the Sketch model presupposes no further interaction between the two modalities after the initial conceptualization stage, the Interface hypothesis assumes a bidirectional interaction also at later production stages, based on the finding that speakers' gestures are also shaped by language-specific linguistic features (e.g., Kita & Özyürek, 2003). Regardless of whether one assumes a bidirectional interaction of the two modalities or not, findings which suggest that the speaker's communicative intentions shape gesture production and execution, like the addressee-based adaptations of gesture frequency and rate, gesture size, or gesture position and orientation (e.g., Galati & Brennan, 2014; Holler, Turner & Varcianna, 2013; Jacobs & Garnham, 2007; Özyürek, 2002) generally support these accounts.

From a speaker-oriented, cognitive perspective, gesturing is assumed to facilitate speaking and/or to provide the speaker with a cognitive benefit at the level of memory or conceptual planning. This cognitive relationship is captured in models like the Lexical gesture process model (Krauss et al., 2000), the Gesture as simulated action (GSA) framework (Hostetter & Alibali, 2008; 2018) or the Gesture-for-Conceptualization Hypothesis (Kita et al., 2017). According to the Lexical gesture process model (Krauss et al., 2000), gestures may aid in lexical retrieval via a process of cross-modal priming, i.e., visual imagery activates verbal concepts. Evidence for this hypothesis comes from research showing that speakers gesture more when they have word finding difficulties (e.g., Morsella & Krauss, 1994), and that prohibiting gestures makes speech less fluent (e.g., Rauscher, Krauss, & Chen, 1996). In contrast to the Lexical gesture process model, the scope of the GSA framework and the Gesture-for-Conceptualization Hypothesis goes beyond the single word level. Within the GSA framework (Hostetter & Alibali, 2008; 2018), it is assumed that gestures originate from visual and motor simulations that accompany the speech planning process. Whether a gesture is executed depends on the strength of the underlying activation, and on the speaker's gesture threshold, i.e. the likelihood of overtly producing a gesture. Individual differences in cognitive abilities as well differences in task demands may affect this gesture threshold, with higher intrinsic or extrinsic load leading to a lowering of the gesture threshold and thus higher gesture rates. The Gesture-for-Conceptualization Hypothesis (Kita et al., 2017) assumes that gestures aid speech production by activating, manipulating, packaging, and exploring spatio-motoric information, and that higher cognitive or conceptualization load leads to increased gesticulation. Findings like increased gesture rates under higher cognitive load (Melinger & Kita, 2007), or a relationship between lower cognitive abilities and higher gesture rates (Chu et al., 2014; Gillespie et al., 2014) thus support these latter two accounts. However, all three “cognitive” accounts conceive of gestures as a by-product of thinking-for-speaking and attribute no communicative intention to gesture production itself.[2]

1.2.3. Co-speech gesture production in aging

To date, only a small number of studies has addressed the effects of aging on co-speech gesture production and none has taken the context, that is, the communicative setting in which gestures were produced, into account. Older adults have been found to produce significantly fewer representational gestures than younger adults during monologue object (Cohen & Borsoi, 1996), action (Feyereisen & Havard, 1999; Theocharopoulou et al., 2015), and spatial descriptions (Arslan & Göksun, 2020) as well as when discussing abstract topics (Feyereisen & Havard, 1999). However, no age-related differences were found during descriptions involving motor imagery (Feyereisen & Havard, 1999) or during other, more narrative tasks (daily activity descriptions, story completion; Arslan & Göksun, 2020). Thus, there appears to be an age-related change in representational co-speech gesture production, however, the contexts in which these differences surface and the underlying reasons remain to be investigated further, as we will see in the following sections.

Co-speech gesture production in aging: The role of contextual factors

As stated above, the role of communicative context was hardly investigated in these previous studies: in none of these studies a communicative paradigm was used in which older adults interacted with a co-present, non-confederate addressee. As the previous summary of the communicative functions of gestures indicates, factors like mutual visibility between speaker and addressee, addressee feedback and the amount of shared knowledge between speaker and addressee greatly influence gesture production, at least in younger adults. Thus, the lack of a naïve addressee may have greatly affected

[2] Note however that although Hostetter & Alibali’s 2018 GSA framework does not assume a communicative intention to be underlying gesture production itself, the model assumes that communicative intentions modulate how events are simulated as well as the speaker’s likelihood to produce overt gestures, thereby affecting gesture production indirectly.

older participants' communicative intention or motivation and therefore also their co-speech gesture use.

Co-speech gesture production in aging: The role of cognitive factors

In contrast to contextual factors, the potential effects of cognitive aging on gesture production have been recognized in those previous studies. For example, Arslan and Göksun (2020) could relate the observed differences in representational gesture frequency in the spatial task to individual differences in mental imagery skills.[3] An age-related decrease in the use of mental imagery thus appears to cause fewer images to be expressed in gestures (see also Cohen & Borsoi, 1996). Other proposals, for example, that older adults encode information verbally rather than visually and therefore produce fewer gestures (Feyereisen & Havard, 1999; Theocharopoulou et al., 2015) do not make the underlying mechanisms explicit. However, they also hint at a cognitive cause, suggesting that simultaneous speech and gesture production may be too demanding for older adults, causing them to focus primarily on the spoken modality.

Cognitive abilities which may be hypothesized to be involved in the concurrent production of speech and co-speech gesture are verbal and visual WM. These two abilities have been shown to affect gesture production in younger adults, mainly such that lower verbal and visual WM capacities are associated with higher gesture frequencies (e.g., Chu et al., 2014; Gillespie et al., 2014; see section 1.2.2.). Yet, Arslan and Göksun (2020) could find no association between age-related gestural differences on the spatial task and age-related differences in visual WM scores. Hence, while the literature on younger adults suggests that certain cognitive abilities affect co-speech gesture production in a certain way, the relationship between older adults' gesture production and age-related cognitive changes remains unclear.

Co-speech gesture production in aging: Outstanding questions

Thus, although previous studies on co-speech gesture production in aging suggest that there may be systematic differences between older and younger adults at least in certain domains, several questions remain: Do the previous findings on age-related differences in gesture production generalize to more communicative settings? How is older adults' co-speech gesture production affected by the communicative needs of an addressee? How does the nature of the communicative task, for example, a narrative vs. a spatial

[3] Interestingly, the use of verbal spatial language remained unaffected by aging or mental imagery skills (Arslan & Göksun, 2020).

task, influence older adults' multimodal language production? Finally, whether and how do age-related cognitive changes affect gesture production? By exploring these questions, I aim to not only gain a more ecologically grounded understanding of older adults' communicative competences in interactive settings, but also to contribute to theoretical accounts of co-speech gesture production with respect to the communicative and cognitive functions of gestures and ultimately also the relationship between speech and co-speech gestures during language production itself. In chapters 2 and 3 of this thesis, I present two experimental studies designed to address these questions (see section 1.3. for details).

1.2.4. Co-speech gesture comprehension in younger adults

Let us now turn to co-speech gesture comprehension. So far, I have assumed that one of the functions of co-speech gestures is to communicate. That is, speakers use gestures in addition to speech with the intention to convey relevant information to their addressee. However, the communicative value of co-speech gestures obviously depends not only on the speaker's communicative intentions, but crucially also on the addressee's ability to perceive and process gestures and to integrate the meaning conveyed by these gestures with that conveyed by speech in order to gain a full understanding of the speaker's intended message. And indeed, ample behavioral research with younger adults convincingly shows that addressees process and integrate the meaning conveyed via the two modalities during language comprehension, and that they interpret speech in the context of co-speech gestures and vice versa (e.g. Kelly, Healey, Özyürek, & Holler 2015; Kelly et al., 2010; see Kelly, 2017, for a recent review of the integration of speech and co-speech gesture). For example, listeners pick up important pragmatic and semantic information that is not present in the speech signal and combine it with speech to form an integral interpretation of the speaker's utterance (Kelly, Barr, Church, & Lynch, 1999). Furthermore, neurocognitive studies indicate that speech and co-speech gestures are processed and integrated simultaneously during language comprehension (e.g. Drijvers & Özyürek, 2018; Holle & Gunter, 2007; Obermeier, Holle, & Gunter, 2011; Kelly, Kravitz, & Hopkins, 2004; Özyürek, Willems, Kita, & Hagoort, 2007; Wu & Coulson, 2007, 2010), and the processing appears to occur in overlapping brain regions (Straube et al., 2012; Willems, Özyürek, & Hagoort, 2007; see Özyürek, 2014 for a review).

Co-speech gesture comprehension in younger adults: The role of contextual factors

However, as for co-speech gesture production, the context in which gestures are perceived plays an important role for gesture comprehension and gesture-speech integration. Especially when speech comprehension is hard, gestures can provide the listener/observer with a clear communicative benefit, reminiscent of the beneficial effects of visible speech on older (and younger) adults' speech comprehension in noise (see section 1.1.2.). For example, Drijvers and Özyürek (2017) showed that seeing co-speech gestures significantly improves the comprehension of a degraded speech signal in younger listeners. Importantly, listeners also integrate the semantic information expressed by gestures with the phonological information available from visible speech. In fact, these two visual signals combined provide listeners with a significantly larger benefit than either signal on its own. Furthermore, adverse listening conditions may also boost the reliance on gestures as a valuable source of visual information: In an ERP study, Obermeier, Dolk, and Gunter (2012) found significant effects of speech-gesture integration only under adverse listening conditions, that is, either when speech was presented in babble noise to normal hearing younger adults, or when speech was presented to hearing-impaired younger adults. To summarize, younger adults' comprehension and integration of co-speech gestures is affected by contextual factors like adverse listening conditions or the availability of visible speech.

Co-speech gesture comprehension in younger adults: The role of cognitive factors

Surprisingly, unlike for gesture production, to date little research has been dedicated to the role of cognitive abilities in the speech-gesture integration and comprehension process, potentially because it appears to happen effortlessly. Although a considerable amount of studies has addressed the neural correlates of speech – co-speech gesture processing and integration (e.g., Özyürek, 2014, 2018; Kelly, 2017 for overviews), individual cognitive abilities involved in these processes have received limited attention. It is reasonable to assume that the simultaneous perception, processing, integration and interpretation of auditory and visual information requires cognitive resources. Obvious candidates would be verbal and/or visual WM. Wu and Coulson (2014) formulated the *verbal resources hypothesis*, according to which speech – gesture integration primarily recruits the verbal WM system and the *visual resources hypothesis*, which states that the integration process depends largely on the visuo-spatial WM system. Limited evidence has been found for the visual resources hypothesis, such that individuals with greater visual (but not verbal) WM capacity show greater sensitivity to gesture-speech congruence than individuals with lower visual WM capacity, suggesting that visuo-spatial resources are indeed relevant to some aspect of gestural processing (Wu & Coulson, 2014, however, see also Coulson & Wu, 2019). Yet, more research is needed to further explore the relationship between individual cognitive abilities and co-speech gesture – speech processing and integration.

Theoretical accounts of co-speech gesture comprehension

Based on the apparently effortless perception and integration of speech and co-speech gestures and the finding that in addition to a bidirectional influence between speech and gestures during language comprehension, listeners cannot ignore gestural information even if this is detrimental to their ability to fulfill a given task, Kelly et al. (2010) formulated the *Integrated systems hypothesis*. This hypothesis assumes that speech and gesture mutually interact during language comprehension, and this interaction is obligatory and automatic, i.e. not subject to conscious control or dependent on additional cognitive resources (see also Kelly, Creigh, and Bartolotti, 2010). However, in recent years, several factors have been identified that modulate the semantic integration of gestures with speech, e.g., the perceived intentionality underlying the coupling of speech and gesture (Kelly, Ward, Creigh, & Bartolotti, 2007), the temporal synchrony of speech – gesture onset (Habets, Kita, Shao, Özyürek, & Hagoort, 2011), the presence of background noise (Obermeier et al., 2012), addressee status (Holler, Schubotz, Kelly, Hagoort, Schütze, & Özyürek, 2014; Holler, Kokal, Toni, Hagoort, Kelly, & Özyürek, 2015), or visual WM capacity (Wu & Coulson, 2014), thereby challenging Kelly et al.'s (2010) claim about the obligatory and automatic nature of this process (see Kelly, 2017, for a recent review of the integration of speech and co-speech gesture). A comprehensive account, that incorporates the influence of these different factors on speech – gesture integration and comprehension is yet to be formulated. Similarly, as already stated in the previous section, an explicit role for different cognitive abilities involved in speech – co-speech gesture perception, integration, and comprehension remains to be investigated more fully.

1.2.5. Co-speech gesture comprehension in aging

As summarized in section 1.1.2., older adults are often faced with increasing speech comprehension difficulties due to cognitive and sensory changes, especially under adverse listening conditions (e.g., Thornton & Light, 2006; Sommers & Phelps, 2016). It appears that these difficulties extend to the gestural modality: older adults were found to benefit less than younger adults from co-speech gestures in addition to visible speech under adverse listening conditions (speeded speech, Thompson, 1995; dichotic shadowing, Thompson & Guzman, 1999). Moreover, even under ideal listening conditions, older adults were less likely to integrate the meaning expressed in co-speech gestures with that expressed in speech, even though there were no age-related differences either in speech-only comprehension or gesture-only interpretation (Cocks et al., 2011).

Co-speech gesture comprehension in aging: The role of contextual factors

Due to the relative scarcity of research on the subject, the role that the context in which gestures are perceived plays for older adults' gesture comprehension is difficult to assess. Yet, there are two issues which may have affected older adults' co-speech gesture integration/comprehension in previous research. In the clear speech task used by Cocks et al. (2011), one obvious drawback was that the speaker's face was covered. This somewhat artificial presentation of stimulus materials (though common practice in the gesture literature) may have affected older adults' inclination to integrate the spoken with the gestural message. In the studies by Thompson (1995) and Thompson and Guzman (1999), the speaker's face was visible, and hence also the speaker's articulatory lip movements (visible speech). However, the conditions used to test older adults' ability to benefit from co-speech gestures may have been too challenging (very fast speech rates, dichotic shadowing) such that they might not have captured older adults' true ability to comprehend or benefit from gestures.

Co-speech gesture comprehension in aging: The role of cognitive factors

Rather than to the context in which gestures were perceived, the observed age-related differences in speech-gesture integration have been attributed to age-related WM limitations. Cocks et al. (2011) proposed that the integration of speech and co-speech gestures requires WM capacity in order to store and update intermediate results of the interpretation process. Older adults' WM resources may have been consumed with speech processing operations, leaving insufficient resources for gesture comprehension and integration, an interpretation also advanced by Thompson (1995). However, it is worth pointing out that no direct relationship between age-related differences in cognitive functioning, either verbal or visual WM, and gesture comprehension or the ability to benefit from gestures were established, or even investigated.

Co-speech gesture comprehension in aging: Outstanding questions

As summarized above, previous research on co-speech gesture comprehension in aging is relatively unanimous in the conclusion that age-related WM deficits prevent older adults from processing and exploiting co-speech gestures to the same extent as younger adults do. Yet, a number of questions remain: How is older adults' ability to comprehend and benefit from co-speech gestures affected when the speech signal is embedded in noise, a context in which younger adults have been shown to benefit greatly from additional gestural information? Are older adults able to integrate the semantic information conveyed by gestures with the phonological information conveyed by visible speech to maximally enhance their speech comprehension, like younger adults do? Recall that visible speech is a visual signal that older adults have been shown to benefit from when presented with SiN. Finally, how do age-related changes in cognitive functioning affect the comprehension of communicative co-speech gestures? By addressing these questions, I aim to gain a more ecologically grounded understanding of older adults' language comprehension and at the same time to contribute to accounts of speech – co-speech gesture processing and integration, in particular the involvement of cognitive abilities in these processes. In chapter 4 of this thesis, I present an experimental study designed to address these questions (see the following section for details).

1.3. The present thesis

The aim of the research presented in this thesis was to investigate how aging and age-associated changes in cognitive functioning modulate the production and comprehension of speech and co-speech gestures in different communicative contexts. As the literature summary in this chapter has shown, we are currently faced with several gaps in our understanding of the multimodal communicative competences of older adults as well as the potential role that cognitive aging may play in this respect. In particular, it is presently unclear whether and how older adults use co-speech gestures in face-to-face communication, and whether they adapt their gesture use according to an addressee's communicative needs. Similarly, it is unclear whether older adults can exploit the information conveyed in the gestural modality to improve their language comprehension, in particular when comprehension is difficult due to background noise. While previous research suggests that there are systematic differences between younger and older adults in both gesture production and comprehension, it remains unclear what the effects of the communicative context on the one hand and of age-related cognitive changes on the other hand are on the production and comprehension of multimodal utterances.

I will address these issues – and try to fill some of the current gaps – by placing older and younger adults into different communicative contexts which have been employed in previous investigations of younger adults' co-speech gesture use and/or previous investigations of older adults' spoken language comprehension, as summarized above. In doing so, I will apply two concepts of "communicative context": personal and incremental common ground in language production, and the presence of background noise in language comprehension. Evidently, common ground and background noise are two fundamentally different types of conversational context. Yet, both are highly relevant in successful language use and therefore provide suitable environments for investigating older adults' communicative use of co-speech gestures.[4] The novelty of the approach used in this thesis is that it combines these contextual factors with the assessment of individual differences in cognitive abilities in trying to understand age-related changes in multimodal language use. In this way, I aim to contribute not only to our understanding of multimodal language use in older versus younger adults, but also to theoretical models on the influence of contextual and cognitive factors on multimodal language production and comprehension in general.

1.3.1. General research questions and hypotheses

The questions that guided the research presented here were the following: Is older adults' co-speech gesture production and comprehension affected by the context in which language is produced and perceived and if so, how? Specifically, how does the presence or absence of common ground with an addressee affect language production? How does background noise affect the ability to benefit from co-speech gestures during language comprehension? Finally, how do age-related changes in cognitive functions affect the production and the comprehension of multimodal utterances?

In order to address these questions, I designed three experimental studies, reported in the empirical chapters of this thesis. In the first two studies, I investigated older adults'

[4] Common ground-based recipient design is a pervasive feature of daily language use. As pointed out above, it is currently unclear whether older adults can and do adapt their multimodal utterances to the same extent as younger adults. It is similarly unclear whether and how common ground-based recipient design affects older adults' language comprehension. While there is some previous work on the role of common ground and recipient design for language comprehension in younger adults (e.g., Brown-Schmidt, 2009; Clark & Carlson, 1981; Fussell & Krauss, 1989; Keysar, Barr, Balin, & Paek, 1998), this is to date limited to the spoken modality. Older adults' comprehension of multimodal common ground-based recipient design is certainly worthy of future investigations. However, for the present thesis, I decided to focus on background noise as a contextual factor, primarily, because background noise has been shown to severely affect older adults' ability to comprehend language.

communicative co-speech gesture production using a narrative and a spatial task, manipulating the amount and the type of mutually shared knowledge, or common ground, between participants. In the third study, I investigated older adults' communicative gesture comprehension, specifically their ability to benefit from co-speech gestures in addition to visible speech when trying to understand speech embedded in background noise.

For the studies investigating co-speech gesture production, I expected older adults to show less evidence of verbal and importantly also gestural common ground-based recipient design than younger adults, based on previous findings in the spoken modality (Horton & Spieler, 2007; Hupet et al., 1993; Lysander & Horton, 2012; Saryazdi et al., 2019). In terms of the effects of (cognitive) aging on gesture production more generally, I considered two possible outcomes: Due to age-related cognitive limitations, older adults might rely relatively *more* on gestures. As the literature summary in this chapter has illustrated, previous research with younger adults suggests that producing co-speech gestures supports language production (e.g., Kita et al., 2017; Krauss et al., 2000; Melinger & Kita, 2007), therefore, older adults might *gesture more* relative to younger adults to compensate for age-related language deficits (for a similar view of gesture as a compensatory tool for limited cognitive abilities, see also Özer & Göksun, 2020). Alternatively, older adults might rely relatively *less* on gestures, due to age-related cognitive limitations. Previous literature suggests that the production of communicative gestures may be cognitively costly (e.g., Mol et al., 2009), or that age-related changes in the use of mental imagery negatively affect older adults' gesture production (Arslan & Göksun, 2020; Cohen & Borsoi, 1996). Therefore, older adults might focus on spoken language and avoid the extra production demands of an additional modality, i.e., they may *gesture less* relative to younger adults during language production (see also Theocharopoulou et al., 2015), at least in tasks involving the use of mental imagery (Arslan & Göksun, 2020, but cf. Feyereisen & Havard, 1999).

For the study on gesture comprehension, I similarly considered two possible outcomes for the effects of (cognitive) aging: Older adults may rely *more* on gestures relative to younger adults, as previous research has shown that perceiving co-speech gestures can support language comprehension (e.g., Drijvers & Özyürek, 2017). Particularly in contexts which are known to put older adults at a disadvantage, like speech embedded in background noise (e.g., Thornton & Light, 2006; Sommers & Phelps, 2016), older adults may therefore show a *greater benefit* from additional gestural information relative to younger adults during language comprehension. Alternatively, older adults might rely relatively *less* on gestures during language comprehension. Previous research suggests that the processing and/or integration of an additional, gestural modality may actually be cognitively costly (e.g., Cocks et al., 2011), therefore, older adults may receive a *smaller benefit* from additional gestural information relative to younger adults during language comprehension (see also Thompson, 1995; Thompson & Guzman, 1999).

The novelty of the approach used here is that I systematically investigate the interplay of contextual and cognitive factors in older adults' multimodal language use. Considering and identifying the contextual factors which potentially modulate the effects of cognitive factors is crucial, since aging and age-related changes in cognitive functioning can be hypothesized to affect co-speech gesture production and comprehension in such opposite ways.

When investigating age-related behavioral changes, there is the methodological risk of attributing any differences in gesture production or comprehension to age-related cognitive changes. I aim to avoid this risk by assessing cognitive abilities independently of the gesture production/comprehension tasks and using the resulting cognitive measures as predictors for the behavioral measures I obtain. In this way, for each individual cognitive construct, it is possible to test whether it has an effect on a certain outcome variable, and what the direction of this effect is. Moreover, effects of cognitive and contextual factors will not be confounded, as would be the case if I used a secondary task, such as an additional memory task, or adjusted the communicative task in order to manipulate cognitive load.

1.3.2. Overview of chapters

In the first part of this thesis, the focus is on speech and co-speech gesture production in face-to-face communication (Chapters 2 and 3). Here, I collected video data from 32 younger adults (aged 21 to 30 years) and 32 older adults (aged 64 to 73 years), using one narrative and one spatial task designed to test whether and how aging and cognitive factors influence a speaker's ability to adapt multimodal utterances according to the communicative needs of a naïve addressee. The same participants took part in both production experiments, the order in which the two tasks were administered was counterbalanced across the participants. The second part of this thesis focuses on speech and co-speech gesture comprehension (Chapter 4). Here, I report a study in which I investigated the effects of aging and cognitive factors on the ability to benefit from co-speech gestures in a word recognition task, recording response accuracies and response latencies. Twenty-eight younger adults (aged 20 to 26 years) and 28 older adults (aged 60 to 80 years) took part in this study. None of the participants had previously participated in the production experiments. Below, the objectives and predictions for each experimental chapter are summarized.

Chapter 2 investigates the effects of aging and cognitive factors on the ability to adapt speech and co-speech gestures according to mutually shared knowledge with an addressee in a narrative task that involved retelling short comic strips. Common ground was established at the outset of the interaction by showing both participants one half of the story, while only one participant (the speaker) would also see the other half of the story. I thus employed a form of personal common ground (Clark, 1996), in which some knowledge with respect to the stories was mutually shared between both participants. Previous research shows that younger adults typically adapt their usage of both modalities in similar settings, by reducing the amount of information they express in either one or both of the two modalities in the presence of shared knowledge (e.g., Galati & Brennan, 2010; 2014; Gerwing & Bavelas, 2004; Hilliard & Cook, 2015; Holler & Stevens, 2007; Parrill, 2010). However, this process of addressee-based adaptation may be cognitively costly and therefore affected by cognitive aging (e.g., Horton & Spieler, 2007; Mol et al., 2009). I expected that compared to younger adults, older adults would show less evidence of addressee-based adaptations in their use of speech and crucially also co-speech gestures (e.g., Horten & Spieler, 2007; Hupet et al., 1993; Lysander & Horton, 2012). Furthermore, I hypothesized that the ability to adapt speech and gesture to mutually shared knowledge might be predicted by age-related differences in verbal WM and inhibitory control (e.g., Wardlow, 2013; Healey & Grossmann, 2016; Long et al., 2018). Furthermore, gesture production more generally might be influenced by age-related differences in cognitive abilities, specifically verbal and visual WM and semantic fluency, such that lower cognitive abilities might lead to higher gesture rates (e.g., Chu et al., 2014; Gillespie et al., 2014).

Chapter 3 similarly investigates the effects of aging and cognitive factors on addressee-based adaptations of speech and co-speech gestures, using a spatial task that involved giving instructions on how to assemble 3D-models from building blocks in order to examine whether the findings from the narrative task extend to other communicative contexts. The spatial task presumably relied more strongly on visual and motor imagery than the narrative task, which may affect older adults' speech and crucially co-speech gesture use differently (e.g, Arslan & Göksun, 2020; Feyereisen & Havard, 1999). Common ground was established at the outset of the interaction and additionally accrued in the course of the experiment (i.e., incremental common ground, Clark, 1996).

As in the narration experiment described in Chapter 2, I expected older adults to show less evidence of addressee-based adaptations than younger adults, in speech and in gestures. Furthermore, I tested the possibility that both the ability to adapt speech and gesture to mutually shared knowledge, as well as gesture production more generally, might be influenced by age-related differences in cognitive abilities, specifically verbal and visual WM, inhibitory control, and semantic fluency.

In the study presented in **Chapter 4**, I investigated the effects of aging and cognitive abilities on the comprehension of SiN perceived in the presence of visible speech and iconic co-speech gestures. Previous research shows that older adults' comprehension of SiN benefits from visible speech, while younger adults' comprehension of degraded speech benefits most when both visible speech and iconic co-speech gestures are present (Drijvers & Özyürek, 2017; Drijvers, Özyürek, & Jensen, 2018). My aim was therefore to test whether older adults, too, could benefit from the visual semantic information conveyed by co-speech gestures in addition to the phonological information conveyed by visible speech. I used a single word recognition task in which the speech signal was presented either in clear conditions or against a background of multi-talker babble noise, and the acoustic signal was accompanied by neither visible speech nor co-speech gestures, by visible speech, or by both visible and co-speech gestures. Two possible outcomes were considered: a greater reliance on semantic context in older adults (e.g. Pichora-Fuller et al., 1995) might result in a larger gestural benefit as compared to younger adults; conversely, difficulties with speech-gesture integration, potentially caused by age-related WM limitations (Cocks et al., 2011; Thompson, 1995) might result in a smaller gestural benefit. In my analyses, I controlled for the possibility that the capacity to benefit from co-speech gestures was modulated by individual differences in hearing acuity, verbal and visual WM, and inhibitory control.

Chapter 5 of this thesis consists of a summary of the individual results of each empirical chapter, followed by a general discussion and suggestions for further research.

Finally, I would like to remark that the chapters presented in this thesis are based on articles as submitted to peer-reviewed journals and underwent only minor editing prior to inclusion in this thesis. Each chapter presents a self-contained text. I ask the reader to kindly excuse inevitable repetitions of key concepts and literature that occur across the individual chapters of this thesis.

Chapter 2

Age-related differences in multimodal recipient design: Younger, but not older adults, adapt speech and co-speech gestures to common ground in a narrative task

Abstract

Speakers can adapt their speech and co-speech gestures based on knowledge shared with an addressee (common ground-based recipient design). Here, we investigate whether these adaptations are modulated by the speaker's age and cognitive abilities. Younger and older participants narrated six short comic stories to a same-aged addressee. Half of each story was known to both participants, the other half only to the speaker. The two age groups did not differ in terms of the number of words and narrative events mentioned per narration, or in terms of gesture frequency, gesture rate, or percentage of events expressed multimodally. However, only the younger participants reduced the amount of verbal and gestural information when narrating mutually known as opposed to novel story content. Age-related differences in cognitive abilities did not predict these differences in common ground-based recipient design. The older participants' communicative behavior may therefore also reflect differences in social or pragmatic goals.

This chapter is based on: Schubotz, L., Özyürek, A., and 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. doi: 10.1080/23273798.2018.1527377.

2.1. Introduction

In spite of a growing literature on language and aging, little is known about the language use of older adults in face-to-face interactions (for comprehensive overviews see e.g. Abrams & Farrell, 2011; Thornton & Light, 2006). This lack of knowledge extends to older adults' use of the gestural modality, a core component of language use in face-to-face settings (Bavelas & Chovil, 2000; Clark, 1996; Kendon, 2004; McNeill, 1992). Considering the prominence of face-to-face interaction in every-day language use, we are thus faced with a serious gap in our understanding of the communicative competencies of older adults as well as the potential role that age-related cognitive changes may play in this respect.

Language used in interaction is produced and tailored for an addressee, shaped by a process called *recipient design* (Sacks, Schegloff, & Jefferson, 1974) or *audience design* (Clark & Murphy, 1983). Recipient design is based on an addressee's communicative needs and affects the way in which language users both speak and gesture for others (e.g. Campisi & Özyürek, 2013; de Ruiter, Bangerter, & Dings, 2012; Galati & Brennan, 2014; Hoetjes, Koolen, Goudbeek, Krahmer, & Swerts, 2015; Holler & Stevens, 2007; Holler & Wilkin, 2009). Taking an addressee's perspective into account and designing one's utterances accordingly may be a cognitively demanding process (e.g. Horton & Gerrig, 2005; Horton & Spieler, 2007; Long, Horton, Rohde, & Sorace, 2018; Wardlow, 2013). Considering that healthy human aging is frequently associated with changes in cognitive functioning (Salthouse, 1991), systematic age-related changes in multimodal recipient design may be expected. However, although previous studies have investigated older adults' recipient design in speech, as well as their gesture production in general, these two issues have not yet been brought together. It is currently unclear whether, and if so, how older adults use their multiple communicative channels when designing utterances for others and which role general cognitive abilities play in this process. In order to address these issues, we compared younger and older adults' speech and gesture use in a narrative task that required the addressee-based adaptation of utterances, taking cognitive abilities as a potential modulating factor into account.

2.1.1. Multimodal recipient design in younger and older adults

Verbal recipient design

The ability to engage in recipient design is frequently investigated by manipulating the amount of *common ground* between conversational partners, defined as the knowledge, beliefs and assumptions that conversational partners believe to be mutually shared and that require the appropriate adaptation of utterances (Clark, 1996). Generally, the larger the common ground, i.e. the more information conversational partners mutually share, the less they put into words. This is characterized, for example, by shorter utterances, less complex syntax, or less informational content (e.g. Fussel & Krauss, 1992; Galati & Brennan, 2010; Isaacs & Clark, 1987). Older adults' ability to engage in recipient design based on common ground has previously been compared to that of younger adults using referential communication tasks. Here, participants are required to establish mutual reference to a limited set of objects over the course of several trials, thereby gradually increasing the amount of common ground (Horton & Spieler, 2007; Hupet, Chantraine, & Nef, 1993; Lysander & Horton, 2012). The results of these studies have shown that younger adults' interactions become increasingly more efficient, indicated by shorter utterances and task-completion times on later compared to earlier trials. Older adults, on the other hand, are less efficient than younger adults, indicated by longer utterances, longer task-completion times, and more errors. It thus appears that compared to younger adults, older adults are less successful at interactively designing their utterances for others.

The role of cognitive abilities in verbal recipient design

Horton and Spieler (2007) suggest that older adults' inferior performance on these referential communication tasks may be due to age-related cognitive limitations, specifically difficulties in retrieving partner-specific information from memory (see also Horton & Gerrig, 2005). Additionally, there are indications that working memory may play a role in recipient design: Work on visual perspective-taking abilities in younger (Wardlow, 2013) and older adults (Healey & Grossmann, 2016) suggests that working memory plays a significant role when speakers are required to take an addressee's visual perspective into account while formulating their utterances. Older adults perform more poorly on these tasks. Recipient design in conversation similarly requires the awareness that the addressee's perspective may differ from one's own, as well as the ability to incorporate this knowledge during online language processing (see e.g. Brennan, Galati, & Kuhlen, 2010), and should therefore also rely on working memory.

In addition to memory functions, inhibitory control has also been proposed to play a role in verbal recipient design. Hupet et al. (1993) speculate that deficits in inhibitory control could cause older adults to have difficulties inhibiting irrelevant, egocentric information from entering memory (see also Hasher and Zacks, 1988), which may explain why they have difficulties with partner-specific adaptations in dialogue. Furthermore, inhibitory control or executive function has also been related to perspective-taking abilities in younger (Wardlow, 2013) and older adults (Long et al., 2018). Thus, inhibitory control may be underlying the ability to inhibit one's own, egocentric perspective in favour of the addressee's, another crucial component of successful verbal recipient design (Brennan et al., 2010; Keysar, Barr, & Horton, 1998; see also Brown-Schmidt, 2009, for the role of executive function in perspective-taking during language comprehension).

Both working memory and inhibitory control are assumed to decline in healthy aging (Hasher, Lustig, & Zacks, 2007; Hasher & Zacks, 1988; Salthouse, 1991; but see Verhaeghen, 2011, for a more critical examination of the role of executive functions in age-related cognitive change). One of the aims of the current study was therefore to establish whether these factors contribute to the behavioral differences in verbal recipient design previously observed in younger vs. older adults.

Multimodal recipient design

Most of the studies described above do not consider the multimodal character of face-to-face language use.[5] Yet, information conveyed visually is essential to face-to-face interaction. Especially representational co-speech gestures, i.e. "gestures that represent some aspect of the content of speech" (Alibali, Heath, & Myers, 2001, p. 172), contribute crucially to the meaning of a message. For example, speakers can use their hands to indicate the size or shape of an object, to depict specific aspects of an action, or to spatially locate referents that they mention in their speech by pointing. There is a close semantic and temporal alignment between representational co-speech gestures and the speech they accompany (McNeill, 1992; Kendon, 2004; see Özyürek, 2017, for a recent review). However, rather than being fully redundant, gestures often depict information that semantically adds to and complements what is being said (Holler & Beattie, 2003a, 2003b; Rowbotham, Holler, Wearden, & Lloyd, 2016). Moreover, like spoken utterances, co-speech gesture use is sensitive to social context variables. For example, representational gesture rate (e.g. the number of gestures produced per 100 words) is modulated by the visibility between speaker and addressee (e.g. Alibali et al., 2001; Bavelas, Kenwood, Johnson, & Phillips, 2002; Mol, Krahmer, Maes, & Swerts, 2011), as well as by dialogic interaction (e.g. Bavelas, Gerwing, Sutton, & Provost, 2008). Addressee location and feedback influence how gestures represent semantic information (Holler & Wilkin, 2011; Kuhlen, Galati, & Brennan, 2012; Özyürek, 2002) and how frequently gestures occur in relation to speech (Jacobs & Garnham, 2007). Hence,

[5] With the exception of Lysander and Horton (2012), who take eye-gaze into consideration.

for a fuller understanding of older adults' ability to communicate with others, it is necessary to take information conveyed in the gestural modality into account.

Research with younger adults shows that common ground appears to affect speech and gesture in similar ways. In the presence of mutually shared knowledge, when common ground is assumed, gestures often become less informative (e.g. Gerwing & Bavelas, 2004; Hilliard & Cook, 2015; Holler & Stevens, 2007; Parrill, 2010), and/or less frequent, at least in absolute terms. In relative terms, this means that, most commonly, speech and gesture reduce to a comparable degree so that gesture rate does not differ in the presence or absence of mutually shared knowledge (e.g., Campisi & Özyürek, 2013; de Ruiter et al., 2012, Galati & Brennan, 2014; Hilliard & Cook, 2015; Hoetjes et al., 2015).[6] This is in line with the notion that the two modalities operate as a single, integrated system (Kita & Özyürek, 2003; McNeill, 1992; So, Kita, & Goldin-Meadow, 2009), and that this speech-gesture system operates in a coordinated and flexible manner, in response to current communicative demands (e.g., Kendon, 1985; 2004). It is currently unclear however, whether the speech-gesture system is equally flexible in older adults, particularly when designing utterances for others. The present study will address this issue. In doing so, we also take into account the role of cognitive abilities, as there are indications that gesture production is closely tied to cognitive functions.

The role of cognitive abilities in multimodal utterances and recipient design

Previous research has shown close ties between general cognitive abilities and gesture production. In order to understand whether and how older adults adapt their multimodal utterances to an addressee's needs, we therefore also have to take the cognitive functions of gestures into account.

Generally speaking, gesturing is assumed to provide the speaker with a cognitive benefit. Co-speech gestures may aid the speaker in the speech planning process, e.g. in conceptual planning (Hostetter, Alibali, & Kita, 2007; Kita & Davies, 2009; Mehlinger & Kita, 2007), or by lightening cognitive load more generally, i.e. freeing up cognitive resources during speaking (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Wagner Cook, Yip, & Goldin-Meadow, 2012). Limited cognitive abilities lead to an increase in

[6] Note, however, that the proportional relation of speech and gesture, expressed in measures of relative frequency, such as gesture rate (e.g. the number of gestures per 100 words), may vary considerably, depending on whether the two modalities are reduced to the same extent, or whether the reduction in one modality is stronger than in the other; for a detailed discussion of this issue see Holler and Bavelas (2017).

gesture frequency, e.g. lower visual working memory (Chu, Meyer, Foulkes, & Kita, 2014), lower verbal working memory (Gillespie, James, Federmeier, & Watson, 2014), or lower phonemic fluency in combination with higher spatial skills (Hostetter & Alibali, 2007). Although differences in the tasks used to assess cognitive functioning and to elicit gestures make the individual studies difficult to compare, the results can be interpreted as further support for gesticulation as a compensatory mechanism for individuals' weaker cognitive abilities.

Based on the supposed cognitive benefit of gesticulation and the generally assumed age-related declines in working memory and other cognitive functions (Salthouse, 1991), one might expect older adults to gesture more than younger adults. However, the general observation is that older adults produce fewer representational co-speech gestures. This has been found for tasks including object (Cohen and Borsoi, 1996) or action descriptions (Feyereisen & Havard, 1999; Theocharopoulou, Cocks, Pring, & Dipper, 2015). Feyereisen and Havard (1999) propose that the observed difference may be due to different speech styles, arguing that there may be a "trade-off between richness of verbal and gestural responses" (p. 169) causing older adults to produce fewer representational gestures when facing the task of speaking and gesturing concurrently. Similarly, Theocharopoulou et al. (2015) suggest that older participants encode information verbally rather than visually, resulting in more verbal elaboration and fewer gestures. These findings suggest an age-related shift in the speech-gesture system, with older adults relying relatively more on speech than on gestures.

However, none of these studies used a communicative paradigm in which older speakers interacted with co-present, non-confederate addressees, a factor that can significantly affect communicative behavior (e.g. Kuhlen & Brennan, 2013). Thus, whether older adults' decrease in gesture production also manifests itself in contexts where there is a real addressee present and to what extent older adults can then adapt their gestures to the needs of their addressees – given that recipient design itself might be a cognitively demanding task – remains unknown.

2.1.2. The present study

The main goals of our research are therefore to find out whether, and if so how, younger and older adults differ in their use of speech and co-speech gestures when interacting with an addressee, i.e., whether they adapt their utterances to mutually shared knowledge between speaker and addressee, and whether differences in addressee-based adaptations may be related to differences in cognitive abilities.

In order to address these issues, we designed a narration task in which a primary participant (the speaker) narrated six short comic strips to a secondary participant (the addressee), manipulating whether story content was shared (common ground or CG) or not (no common ground or no-CG) between participants. We thus induced a form of personal common ground (Clark, 1996), in which the mutually shared knowledge existed from the outset of the interactions rather than building up incrementally (as in e.g. Horton and Spieler, 2007; Hupet et al., 1993, or Lysander and Horton, 2012).

As for cognitive abilities, we assessed speakers' verbal and visual working memory (verbal and visual WM) as well as inhibitory control and semantic fluency. As summarized above, verbal WM and inhibitory control have previously been related to verbal recipient design (Long et al., 2018; Wardlow, 2013; Hupet et al., 1993). Furthermore, verbal and visual WM have been found to be related to gesticulation in general (e.g. Chu et al., 2014 for visual WM; Gillespie et al., 2014 for verbal WM). Finally, we assessed semantic fluency as an indicator of word finding difficulties, which are thought to increase with increasing age (e.g. Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Burke, MacKay, Worthley, & Wade, 1991), and may be related to gesticulation (Rauscher, Krauss, & Chen, 1996).

Our main dependent variables were the speech-based measures 'number of words' and 'number of narrative events per narration', and the gesture-based measures 'gesture rate per 100 words' as well as the 'percentage of narrative events accompanied by a gesture' (or multimodal events). We included both speech-based measures in our analysis, as word counts are a global measure of narration length, while number of narrative events serves as a better approximation of the amount of information contained in the narration. Similarly, gesture rate per 100 words globally captures a speaker's relative weighting of gestures to speech, normalizing for differences in narration length (e.g. Alibali et al., 2001), whereas the percentage of multimodal events is a closer approximation of the amount of semantic information contained in gesture relative to that contained in speech.

In addition, we coded speakers' explicit references to common ground, as this can provide a further indication of their awareness of mutually shared knowledge. Also, we coded the addressees' verbal and non-verbal feedback in order to control for the possibility that any age-related differences in the speakers' behavior might be attributable to systematic age-related differences in addressee behavior.

In line with previous findings, we expected an effect of our common ground manipulation on speech production such that younger adults would use fewer words and include fewer narrative events when relating shared as opposed to novel information (e.g. Campisi & Özyürek, 2013; Fussel & Krauss, 1992; Galati & Brennan, 2010, 2014; Holler & Wilkin, 2009; Isaacs & Clark, 1987). Based on the results obtained by Horton and Spieler (2007), Hupet et al. (1993) and Lysander and Horton (2012), we expected this effect to be significantly smaller in older adults. We additionally aimed to investigate the impact of cognitive abilities on recipient design in speech, expecting that older adults' lower verbal working memory and lower inhibitory control would be associated with a smaller reduction in words and narrative elements (based on the work by e.g. Healey & Grossmann, 2016; Horton & Gerrig, 2005; Horton & Spieler, 2009; Hupet et al., 1993; Long et al., 2018; Wardlow, 2013).

Regarding the effect of the common ground manipulation on gesture production in younger adults, we expected an overall reduction in gesture frequency and semantic content, in line with the studies cited above. Note that we refrain from making directed predictions for the effect of common ground on gesture rate and multimodal utterances specifically, though, since previous findings vary with respect to the proportional reduction of gesture in relation to speech (see Holler and Bavelas, 2017, for an overview). Instead, our focus is the direct comparison between younger and older adults in how they adapt their multimodal utterances to the addressee's knowledge state. Due to the previously found age-related differences in verbal behavior in relation to common ground (Horton & Spieler, 2007; Hupet et al., 1993; Lysander & Horton, 2012) and due to speech and gesture functioning as one integrated system (Kita & Özyürek, 2003; McNeill, 1992), we predict older adults to be less adaptive to common ground than younger adults, not only in their speech but also in the way they draw on gesture when designing utterances for their recipients.

For a general effect of age on representational gesture production, two possible hypotheses can be formulated considering the literature summarized in the previous section. Based on the findings by Cohen and Borsoi (1996), Feyereisen and Havard (1999), and Theocharopoulou et al. (2015), we might expect older adults to gesture at a lower rate than younger adults. On the other hand, due to potential age-related cognitive limitations, older adults may actually gesture more than younger adults in order to free up cognitive resources (Goldin-Meadow et al., 2001; Wagner Cook et al., 2012) or compensate for weaker cognitive abilities (Chu et al., 2014; Gillespie et al., 2014; Hostetter & Alibali, 2007).

2.2. Method

2.2.1. Participants

Thirty-two younger adults (16 women) between 21 and 30 years old (M_{age} = 24.31, SD = 2.91) and 32 older adults (16 women) between 64 and 73 years old (M_{age} = 67.69, SD = 2.43) participated in the study. All participants were native Dutch speakers with self-reported normal or corrected-to-normal vision and hearing and no known history of neurological impairment. Each participant was allocated to a same-age and same-sex pairing. The role of speaker or addressee was randomly assigned and kept constant across the entire experiment. Only the speaker data were analyzed here. All participants in the role of speaker had minimally secondary school education, except for one older participant who only had primary school education. Participants were recruited from the participant pool of the Max Planck Institute for Psycholinguistics and received between € 8 and € 16 for their participation, depending on the duration of the session. The experiment was approved by the Ethics Commission for Behavioural Research from the Radboud University Nijmegen.

2.2.2. Materials

Six black-and-white comic strips from the series "Vater und Sohn" (by cartoonist e.o. plauen, for an example see Appendix A) were used to elicit narratives. Each strip consisted of a self-contained story, which centered on the activities of a father and a son. Half of the strips consisted of four frames, the other half of six frames. The strips did not contain any writing but consisted of black and white drawings only and were not known to the participants beforehand. Four experimental lists determined the order in which the different strips were presented. Initially, we created two orders of presentation for the six stories, one being the reverse of the other. In doing this, we alternated between four- and six-frame stories. In a second step, we assigned the condition in which the stories occurred. For each story, either the first or the second half (corresponding to two or three frames, depending on story length) could be presented in common ground. We alternated between which half of each story would be presented in common ground (e.g. first story – first half, second story – second half, third story – first half, etc.). Counterbalancing the order of common ground presentation across lists ultimately resulted in four experimental lists. Each list was tested eight times, distributed evenly across age groups and sexes.

2.2.3. Procedure and common ground manipulation

Upon arrival, the speaker and the addressee were asked to sit in designated chairs at a table at 90° from each other. Two video cameras were set up on tripods at a small distance from the table, one of them getting a frontal view of the speaker, the other one positioned such that it captured both speaker and addressee (see Figure 1 for stills from the two cameras). Sound was recorded with an additional microphone suspended from the ceiling over the table and connected to the speaker camera.

Figure 1. Example of the lateral (left panel) and frontal (right panel) views of the speaker in the experimental set-up. In this frame, the speaker refers to "a really big fish", both in her speech and in her gesture.

Participants were introduced to each other and received a description of the experiment. This and all subsequent instructions were given both in writing and verbally to ensure that all participants received and understood the information necessary to successfully participate in the experiment. Signed consent was acquired from all participants.

For the narration task, all participants completed one practice trial and six experimental trials, narrating a total of seven stories. At the beginning of each trial, both participants were presented with either the first or the second half of the comic strip and were instructed to look at it together for a limited amount of time without talking, with the aim to experimentally induce common ground about this part of the story. Hence, in each trial there was both CG and no-CG content. Subsequently, the drawings were removed and a screen was put up on the table between speaker and addressee. The speaker then received the full story to look at, with no time limit imposed. Once the speaker signaled that she had understood and memorized the story, drawings and screen were removed again and the speaker narrated the entire story to the addressee. She was instructed to narrate the full story, keeping in mind that the addressee had already seen part of it. Addressees were instructed to listen to the narrations and ask all clarification questions at the end. Then the screen was put back up and the addressee answered a question about the story in writing.[7] Participants received no feedback about the accuracy of these answers so as to not influence speakers' communicative behavior. Depending on the pair, the task took about 20 to 30 minutes. After the experimental tasks were completed, the addressee was allowed to leave, while the speaker performed the cognitive tests.

2.2.4. Transcription and coding

Speech coding

All recordings from the two cameras were synchronized and subsequently segmented into trials. Transcription of speech and annotation of gestures was done in Elan (Version 4.9.4; Wittenburg, Brugman, Russel, Klassmann, Sloetjes, 2006). For all segments, the speaker's initial narration, i.e. the first retelling of the full story without potential subsequent repetitions, was identified. All analyses reported here are based on these initial narrations only, discarding repetitions or clarifications elicited by the addressee following the initial narration. This is motivated by the fact that the focus of our study was the effect of our experimental manipulations on the speakers' behavior rather than the impact of speaker-addressee interaction (for a similar argument see Horton and Gerrig, 2005). Speech from the speaker was transcribed verbatim, including disfluencies such as filled pauses and word fragments. However, disfluencies were excluded from the word counts presented in the results section, as we were mainly interested in speech content and did not want potential age-related differences in the number of disfluencies to influence the word count (e.g. Mortensen, Meyer, & Humphreys, 2006). For this reason, we also distinguished between speech belonging to the narrative proper (i.e. relating to story content) and non-narrative speech such as statements about the task or comments relating to the speaker or the addressee (for this distinction see McNeill, 1992).

Explicit references to common ground

Among the non-narrative speech, we identified explicit references to common ground, i.e. statements such as "this time we saw the first half together". These explicit references to common ground give additional insight into whether participants were

[7] Note that the questions did not target common ground vs. no common ground information systematically and can therefore unfortunately not give any insights into the addressee's information uptake as based on the speaker's narration.

aware of the shared knowledge or not and will be reported separately in the results section.

Narrative event coding

For the narrative event coding, we roughly followed the procedure described in Galati and Brennan (2010). We devised a narrative event script for each of the six stories, containing all elements that we deemed necessary in order to narrate the story accurately and fully (for an example see Appendix B). For the largest part, these were observable events that advanced the plot, with the exception of a few inferences on the intentions of the stories' characters. One event roughly consisted of one "idea unit" (Butterworth, 1975) and frequently corresponded to one syntactic clause. We then checked these scripts against the actual narrations, including additional events in the script if they were included by a substantial number of participants across both age groups. On average, the 4-frame stories contained a total of 18.67 (*SD* = .6) events and the 6-frame stories contained a total of 27.67 events (*SD* = .6). Collapsed across both story types, each story contained 4.63 events per frame (*SD* = .11), with the actual number of events per frame ranging from 1 to 7.

In a subsequent step, we scored each participant's narration based on these fixed scripts for whether the scripted event was contained in the narration or not (note that we only took into consideration the spoken part of the narrations here). In cases where only part of the event was included in the narration, the participant received half a score. A second coder blind to the experimental hypothesis coded 10 % of the trials (N = 20). Inter-rater agreement on narrative event scoring was 94 % overall.

Gesture coding

For the gesture coding, we first identified all co-speech gestures produced by the speaker during narrative speech, disregarding non-gesture movements as well as gestures accompanying non-narrative speech. Our unit of analyses was the gestural stroke, i.e. the most meaningful part of the gesture determined according to criteria established in previous co-speech gesture research (Kendon, 2004; Kita, van Gijn, & van der Hulst, 1998; McNeill, 1992). We then categorized these strokes as representational and non-representational gestures (see Alibali et al., 2001). For our purposes, representational gestures include iconic gestures, which iconically depict shape or size of concrete referents or represent physical movements or actions;[8] metaphoric gestures, which resemble iconic gestures but relate to speech in a metaphorical manner (e.g. a rotating movement of the hand to indicate the passing of time); and pointing gestures or deictics, i.e. finger points to a specific location in imaginary space, e.g. that of a story character (McNeill, 1992).

All other gestures were considered non-representational and include what are frequently called beat gestures, i.e. biphasic movements of the hand, for example to add emphasis, as well as pragmatic gestures (Kendon, 2004), i.e. gestures which have pragmatic functions, for example to convey information about how an utterance should be interpreted, or relating to managing the interaction more generally (Bavelas et al., 1992, 1995).

A second coder blind to the experimental hypotheses coded 10% of the trials for stroke identification, and another 10% of the trials for gesture categorization. Inter-rater agreement on stroke identification, based on stroke onsets and offsets, was 92.3%. Inter-rater agreement on gesture categorization was 97.9%, Cohen's Kappa = .95.

Gesture rates

As we were mainly interested in the semantic content of the narratives and the accompanying gestures, in our analyses we focus exclusively on the representational gestures (i.e., iconic, metaphoric, and abstract deictic gestures). In addition to reporting the raw representational gesture frequency as a descriptive measure, we used two different measures of gesture production in relation to speech in our main analyses.

Representational gesture rate (gestures per 100 words). We computed a gesture rate per 100 words (see above for criteria on word count) by dividing the number of gestures by the number of words a given participant produced for each condition within each trial separately and multiplied this by 100.

Percentage of multimodal events. We computed a percentage of multimodal events for each participant by dividing the number of narrative events accompanied by a gesture by the total number of narrative events per condition within each trial and multiplied this by 100.

---

[8] "Re-enactments", i.e. movements of the body that represented specific actions of the stories' characters, were also coded as iconic gestures, even if they did not include manual movements.

In Appendix C, we additionally report the analyses of gesture frequencies in order to be able to draw direct comparisons between our study and previous studies on gesticulation in older adults (Cohen & Borsoi, 1996; Feyereisen & Havard, 1999; Theocharopoulou et al., 2015), as well as the analyses of gesture rate per narrative event, as used e.g. by Galati and Brennan (2014).

Addressee feedback

As stated in the introduction, gesture production has been found to be sensitive to addressee feedback (e.g. Holler & Wilkin, 2011; Jacobs & Garnham, 2007; Kuhlen, Galati, & Brennan, 2012). In order to ensure that any potential difference in gesture production between younger and older adults would not be due to systematic differences in addressee behavior, we coded the addressees' verbal (backchannels, questions, other verbal remarks) and non-verbal feedback (head movements, manual gestures) for two of the six stories. An analysis of this addressee behavior is reported in the results section.

2.2.5. Cognitive measures

Participants performed the Operation Span Task (Ospan) as a measure of verbal WM, the Corsi Block Task (CBT) as a measure of visuo-sequential WM, the Visual Patterns Test (VPT) as a measure of visuo-spatial WM, the Trail Making Test (TMT) as a measure of inhibitory control, and the animal naming task to assess semantic fluency. Detailed descriptions of these cognitive tasks, how they were administered, and how the scores were computed can be found in Appendix D.

2.2.6. Statistical methods

To investigate the influence of age and the common ground manipulation on the main speech- and gesture-based measures (word and narrative event count, gesture rate and percentage of multimodal events), as well as on explicit reference to common ground and addressee feedback, we fitted linear mixed-effect models in R version 3.2.1 (R Development Core Team, 2015), using the package lme4 (Bates, Maechler, and Bolker, 2017). We only report best-fitting models established via likelihood ratio tests for model comparisons, eliminating all non-significant predictors in the model comparison process. All the models reported contain random intercepts for participants and items (story), as well as by-participant random slopes for the common ground manipulation unless explicitly stated otherwise. Reported p-values were obtained via the package lmerTest (Kuznetsova, Brockhoff, and Christensen, 2016). The function lsmeans from the package emmeans (Lenth, 2018) was used to test linear contrasts among predictors for the individual models.

To investigate the influence of cognitive abilities on our main dependent measures, and to test whether potential age-related differences in verbal and gestural behavior could be attributed to age-related differences in cognitive abilities, we applied the same basic procedure as described above. We built on the best-fitting models established in the previous analyses and created separate models for each cognitive predictor. As the analyses were exploratory, we performed a backwards-model-stripping procedure, starting out with a full model including the cognitive predictor of interest, age, and the common ground manipulation, as well as all their interaction terms, eliminating non-significant interactions and predictors in the model comparison process.

2.3. Results

2.3.1. Gesture frequency and gesture types per age group

Younger adults produced 849 gestures accompanying narrative speech, out of which 542 were iconic gestures (63.84%), 7 metaphoric gestures (0.82%), 104 deictic gestures (12.25%), and 196 non-representational gestures (23.09%). Older adults produced 673 gestures accompanying narrative speech, out of which 479 were iconic gestures (71.17%), 13 metaphoric gestures (1.93%), 60 deictic gestures (8.92%), and 121 non-representational gestures (17.98%). Note again that only representational gestures were included to compute the dependent measures gesture frequency, gesture rate, and percentage of multimodal events reported in the following sections.

2.3.2. Effects of age and common ground on speech and co-speech gesture

Table 1. Means (and SD) for the speech- and gesture-based dependent measures for each age group and condition. CG = common ground condition; no-CG = no common ground condition.

	Younger		Older	
	CG	No-CG	CG	No-CG
Number of words	44.63 (21.35)	65.09 (19.93)	52.39 (12.45)	54.59 (12.47)
No. of narrative events	4.58 (2.07)	5.75 (1.07)	5.37 (.94)	5.35 (1.1)
Gesture frequency	2.02 (1.39)	4.78 (2.39)	3.01 (2.09)	2.74 (2.18)
Gestures/100 words	5.89 (4.5)	7.73 (3.94)	5.96 (3.94)	4.88 (3.86)
% Multimodal events	34.57 (21.56)	54.63 (19.95)	39.59 (24.13)	31.56 (23.64)

Mean values and standard deviations for the various dependent measures by age group and common ground condition are listed in Table 1. The distribution of observations for word count, narrative event count, gesture rate, and percentage of multimodal events is displayed in Figure 2.

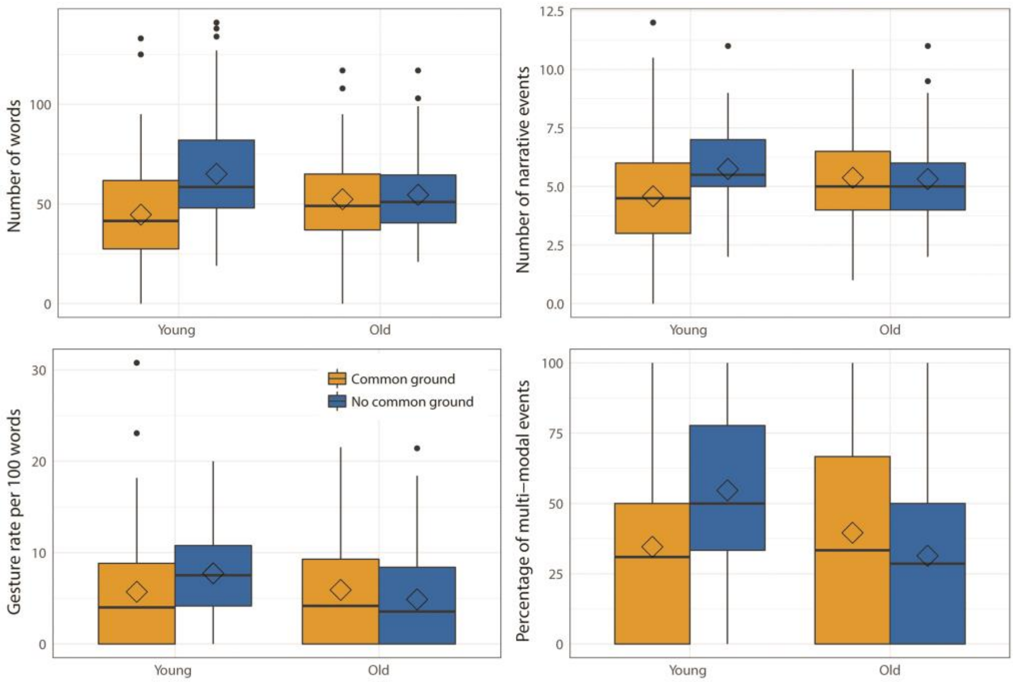


Figure 2. Distribution for the speech- and gesture-based dependent measures summarized by age group and condition (boxplots display six [story] * two [condition manipulation] data points per participant). The black line represents the median; the diamond represents the mean; the two hinges represent the 1st and 3rd quartile; the whiskers capture the largest and smallest observation but extend no further than 1.5 * IQR (data points outside 1.5 * IQR are represented by dots).

Words and narrative events

As described in section 2.2.6, we fitted linear mixed effects models to evaluate the effects of age and common ground manipulation, as well as their interaction, on word count and narrative event production. The models are summarized in Table 2.

In order to obtain the simple main effects of the two predictors we compared nested models to the omnibus models via likelihood ratio tests, excluding only the predictor variable of interest, one at a time, but keeping the respective other predictor as well as the interaction term. There was no main effect for age, such that younger and older adults did not differ in the overall number of words and narrative events they produced

Table 2. Linear mixed-effects models for the effects of age and common ground manipulation on word count and number of narrative events mentioned. Age group = young and Condition = CG[a] are on the intercept. N = 32.[b]

	Words				Narrative events			
	β	SE	t	p	β	SE	t	p
Intercept	44.63	5.37	8.31	< .001	4.58	.53	8.68	< .001
Age group$_{old}$	7.76	6.03	1.29	.207	.78	.55	1.42	.17
Condition$_{no\text{-}CG}$^{a}	20.47	3.57	5.79	<.001	1.17	.41	2.84	.008
Age group$_{old}$: Condition$_{no\text{-}CG}$	-18.26	5.00	-3.65	<.001	-.2	.58	-2.06	.048

[a] CG = common ground; no-CG = no common ground.

[b] Both models contain random intercepts for participants and items and by-participant random slopes for the common ground manipulation.

(both p's > .05). There was an effect of common ground manipulation, significant for word count ($\chi^2(1) = 15.88$, $p < .001$) but not for narrative event count ($\chi^2(1) = 3.59$, $p = .06$), such that participants produced fewer words in the CG as opposed to the no-CG condition. However, this effect was modulated by age, as there were significant interactions between age group and common ground manipulation.

Individual contrasts revealed that only younger adults produced significantly more words and narrative events in the no-CG as opposed to the CG condition (β = 20.47, $SE = 3.65$, $t(34.13) = 5.60$, $p < .001$ and β = 1.17, $SE = .43$, $t(34.10) = 2.75$, $p = .01$ respectively), whereas this difference was not significant for older adults (both p's > .05). Younger adults did not differ from older adults in the number of words and narrative events produced in the CG and no-CG conditions (all p's > .05).

To summarize, younger and older adults did not differ in the overall number of words and narrative events they produced. However, a significant effect of our common ground manipulation was only present in the younger adults, i.e. they used more words and more narrative events when talking about novel as opposed to shared story content.

Representational gesture rate and percentage of multimodal events

As for the speech-based measures, we fitted linear mixed effects models to evaluate the impact of age and common ground manipulation on gesture rate per 100 words and percentage of multimodal events. Note that we did not include a by-participant random slope in the model predicting gesture rate, as this yielded a perfect correlation for the random effects. The final models are summarized in Table 3.

Table 3. Linear mixed-effects models for the effects of age and common ground manipulation on gesture rate per 100 words and percentage of multimodal events. Age group = young and Condition = CG[a] are on the intercept. N = 32.[b]

	Gesture rate per 100 words				Percentage multimodal events			
	β	SE	t	p	β	SE	t	p
Intercept	5.87	1.18	5.00	< .001	34.57	7.19	4.81	< .001
Age group$_{old}$	-.04	1.40	-.03	.98	5.25	7.91	.66	.51
Condition$_{no\text{-}CG}$[a]	1.86	.54	3.46	< .001	20.06	3.46	5.8	< .001
Age group$_{old}$: Condition$_{no\text{-}CG}$	-2.9	.75	-3.84	< .001	-28.25	4.91	-5.76	< .001

[a] CG = common ground; no-CG = no common ground.

[b] Both models contain random intercepts for participants and items. The model predicting the percentage of multimodal events includes by-participant random slopes for the common ground manipulation.

Again, we used likelihood ratio tests to compare nested models in order to obtain the simple main effect of age and common ground manipulation. This yielded no main effects for age or common ground manipulation for both measures (all p's > .05). However, the model summaries (Table 3) show that for the reference group of the younger adults, the effect of common ground was significant, such that participants gestured at a higher rate and produced more multimodal events in the no-CG as opposed to the CG condition. This effect was modulated by age, as the significant interactions between age group and common ground manipulation show.

Individual contrasts confirmed that younger adults gestured at a significantly higher rate in the no-CG as opposed to the CG condition ($\beta = 1.86$, $SE = .54$, $t(343.70) = 3.46$, $p < .001$), whereas older adults showed the reverse trend ($\beta = -1.04$, $SE = .54$, $t(342.69) = -1.96$, $p = .051$). Younger adults also produced significantly more multimodal events in the no-CG as compared to the CG condition ($\beta = 20.06$, $SE = 3.58$, $t(33.86) = 5.61$, $p < .001$), whereas older adults showed the reverse pattern ($\beta = -8.18$, $SE = 3.59$, $t(34.36) = 2.28$, $p = .029$). Contrasts further revealed that younger and older adults did not differ in the rate at which they gestured and in the percentage of multimodal events in the CG condition (both p's > .05). However, there was an age-related difference in the no-CG condition that approached significance for gesture rate ($\beta = 2.86$, $SE = 1.43$, $t(38.28) = 1.97$, $p = .053$) and was significant for percentage of multimodal events ($\beta = 22.99$, $SE = 7.75$, $t(32.95) = 2.97$, $p = .006$). That is, younger adults trended towards gesturing at a higher rate and produced a larger percentage of multimodal events than older adults in the no-CG condition.

To summarize, older and younger adults did not differ in their gesture rate and the percentage of multimodal events overall. However, we found different effects of our common ground manipulation for younger versus older adults. While younger adults gestured at a higher rate and produced more multimodal events when narrating novel as opposed to known story content for their addressees, the opposite was the case for the older adults.

Explicit reference to common ground and addressee feedback

In addition to the main analyses reported above, we explored the influence of age and common ground manipulation on the frequency of speakers' explicit references to common ground, and on the frequency of addressee feedback. Explicit references to common ground can serve as an additional indicator of whether speakers were aware of their addressees' knowledge state. Controlling for addressee feedback is necessary in order to preclude the possibility that younger and older speakers' verbal and gestural behavior differs due to differences in addressee behavior. We fitted linear mixed effect models as described in section 2.2.6. Note that we did not include by-participant random slopes in the models, as this yielded a perfect correlation for the random effects. Full model summaries are provided in Appendix E.

Explicit reference to common ground. Per story, younger adults made on average .72 explicit references to common ground in the CG condition (*SD* = .59) and .03 (*SD* = .09) in the no-CG condition. Older adults made on average .11 explicit references in the CG condition (*SD* = .23) and zero in the no-CG condition per story. With age group = young and common ground condition = CG mapped onto the intercept, the best fitting model contained effects for age (β = -.41, *SE* = .1, $t(50.8) = -3.87$, $p < .001$), common ground condition (β = -.67, *SE* = .07, $t(352) = -9.84$, $p < .001$), as well as the significant interaction term (β = .51, *SE* = .1, $t(352) = 5.33$, $p < .001$). Likelihood ratio tests showed that there was no overall main effect for age ($\chi^2(1) = 2.52$, $p = .11$), but only for common ground condition ($\chi^2(1) = 66.95$, $p < .001$). Thus, the two age groups did not differ significantly from each other in the overall number of explicit references to common ground they made. However, in the CG condition, younger adults produced significantly more explicit references than older adults. Hence, younger adults provided stronger indications of their awareness of the addressee's knowledge state than older adults.

Addressee feedback. We divided the amount of addressee feedback by the number of words per narration to account for differences in narration length. Both younger and older addressees produced numerically more feedback in the CG condition (M_{young} = .07, *SD* = .04; M_{old} = .06, *SD* = .04) than in the no-CG condition (M_{young} = .05, *SD* = .03;

M_{old} = .04, *SD* = .03). The best fitting model contained a significant main effect for the common ground condition (β = -.02, *SE* = .006, $t(93.52) = -2.96$, $p = .004$), confirming the significance of this difference. The main effect for age approached significance (β = -.02, *SE* = .009, $t(31.01) = -1.99$, $p = .06$) such that older adults produced marginally less feedback overall than younger adults. Importantly, the interaction term of age and common ground condition did not improve the model fit, indicating that there was no systematic difference in the amount of feedback that younger and older addressees gave based on common ground condition. Hence, the observed age-related differences in common ground-based adaptation of speech and gesture reported above are unlikely to be due to differences in addressee feedback.

Effects of addressee feedback on verbal and gestural behavior. We followed this analysis up by entering addressee feedback as a predictor into the previously reported models on word and narrative event count, gesture rate, and percentage of multimodal events, drawing on the subset of data for which feedback was coded. This was done in order to test whether accounting for feedback would modulate the effect of common ground for the younger adults that we established in the main analyses. We found that including feedback did not improve the models predicting word count or percentage of multimodal events. For narrative event count, there was no effect of the common ground manipulation in this subset, but addressee feedback had a significant effect such that more feedback predicted a reduction in narrative events (β = -9.91, *SE* = 4.32, $t(119.58) = -2.29$, $p = .02$). This effect appears to be driven more by the younger than by the older adults, but the interaction was not statistically significant. Finally, for gesture rate, feedback had a significant effect (β = 36.16, *SE* = 9.9, $t(113.41) = 3.65$, $p < .001$) such that more feedback predicted a higher gesture rate. However, crucially for our study, the effect of feedback did not influence the effect of common ground or its interaction with age. Overall then, taking addressee feedback into consideration did not eliminate the effect of the common ground manipulation observed in the speech- and gesture-based measures.

2.3.3. Effects of cognitive abilities on verbal recipient design and co-speech gesture

As we were also interested in the influence of cognitive abilities on verbal recipient design and on gesture production, we next turned to these factors. Particularly, we wanted to test whether the age-related differences in verbal and gestural behavior could be attributed to age-related differences in cognitive functioning. As a group, younger adults significantly outperformed older adults on all cognitive tests with the exception of the semantic fluency task, see Table 4. For subsequent analyses, we standardized each task's scores by z-scoring. Correlations between cognitive predictors and dependent measures are reported in Appendix F.

Table 4. Mean scores (and SD) per age group on cognitive tests, plus statistical comparisons (independent t-tests and Mann-Whitney tests where appropriate).

	Younger	Older	Test statistic
Verbal WM (Operation span task)	44.06 (8.39)	34.73 (8.92)	t(29) = 2.99**
Semantic Fluency (Animal naming test)	31.5 (9.4)	27.75 (5.99)	t(30) = 1.35
Inhibitory control (Trail Making Test, TMT)[a b]	14.5 (27)	21.5 (62)	W = 65.5*
Visuo-spatial WM (Visual Patterns Test, VPT)[a]	13 (4)	10 (8)	W = 187.5***
Visuo-sequential WM (Corsi Block Task, CBT)[a]	54 (42)	37.5 (34)	W = 215***

*** = p < .001, ** = p < .01, * = p < .05

[a] Owing to the non-normality of the data, the figures represent Median (and Range).

[b] Note that smaller numbers indicate better performance on this task.

Words and narrative events

First, we tested our hypothesis that verbal WM and inhibitory control influence verbal recipient design, by including these cognitive variables in the models predicting the number of words and narrative events produced per narration. We fitted linear mixed effects models, applying a backwards model-stripping procedure as described in section 2.2.6. Both cognitive measures did not significantly improve the models fit for word and narrative event count, either as main effects or in interaction with age and common ground.

Representational gesture rate and percentage of multimodal events

Next, we tested the hypothesis that lower visuo-spatial or visuo-sequential WM, verbal WM, or semantic fluency are associated with an increase in gesticulation, and whether this affects the two age groups differently, by including these cognitive variables in the models predicting gesture rate and percentage of multimodal events. As for the previous analysis, none of the cognitive measures significantly contributed to models predicting either of the two gesture-based measures.

To summarize, we could not find any evidence that the observed age-related differences in cognitive abilities were predictive of the age-related differences in verbal and gestural behavior reported in the main analyses. Furthermore, individual differences in cognitive abilities also could not predict verbal or gestural behavior more generally, regardless of age group or common ground condition.

2.3.4. Summary of results

Overall, there were no age-related differences in how much participants spoke and gestured. However, in the presence of common ground, only younger adults used fewer words, fewer narrative events, gestured at a lower rate, and produced fewer multimodal events as compared to when there was no common ground. Older adults, on the other hand, did not adapt their speech to common ground. Also, unlike younger adults, they produced fewer gestures in relation to speech in the no-CG condition than in the CG condition. They also gestured less than younger adults in the no-CG condition.

Furthermore, younger adults made more explicit references to common ground than older adults in the CG condition, overtly indicating their awareness of the mutually shared knowledge.

Crucially there were no age-related differences in the amount of addressee feedback, making this an unlikely explanation for differences in verbal and gestural behavior between the two age groups. Additionally, we found that more addressee feedback was predictive of a reduction in narrative events and an increase in gesture rate, regardless of age and common ground.

Finally, although we found significant age-related differences in cognitive abilities, these did not explain the age-related differences in verbal and gestural adaptation to the common ground manipulation.

2.4. Discussion

The present study provides a first insight into how younger and older adults adapt their speech and co-speech gestures to an addressee's knowledge state when narrating short stories, and whether this verbal and gestural behavior is affected by cognitive abilities. We found that younger and older adults did not differ in the number of words and narrative events they used, or in their representational gesture rate and percentage of multimodal utterances overall. However, adaptations of both speech and co-speech gestures based on mutually shared knowledge between speaker and addressee occurred only in the younger, but not in the older adults. Age-related differences in cognitive abilities did not predict these differences in behavior, nor did addressee feedback behavior modulate the observed effects. The individual results will be discussed in more detail below.

2.4.1. Effects of age and common ground on verbal recipient design

Overall, there were no age-related differences in the number of words and narrative events produced per narration. This suggests that younger and older adults were able to remember and reproduce approximately the same amount of information.

Crucially, with respect to our hypotheses concerning the adaptation to mutually shared knowledge, we found that younger adults showed a stronger effect of common ground on speech than older adults. That is, younger adults used fewer words and narrative events to narrate known story content compared to novel content. This is in line with previous findings for younger adults in similar narration tasks (e.g. Galati & Brennan, 2010; Holler & Wilkin, 2009). It shows that the more knowledge speakers assume to be mutually shared, the less verbal information is conveyed. The fact that younger adults frequently referred to common ground explicitly when relating familiar content, e.g. by stating "you've already seen the first half so I'll go through it quickly" similarly shows that they were aware of their addressee's knowledge state.

Furthermore, we found indications that younger adults were not only aware of the addressee's knowledge state as a function of the common ground manipulation, but that they were also sensitive to the addressees' verbal and visual backchannel signals. In the present study, addressees provided more backchannel signals in the presence of shared knowledge. Previous research, for example Galati and Brennan (2014), found shared knowledge to be associated with a reduction in addressee feedback. However, in their task, addressees listened to the retelling of the same story twice, which may have caused the addressee to be less involved and less responsive during the second retelling. In the present task, on the other hand, common ground was manipulated within each story, and even though the addressee had seen part of the story already (thus constituting common ground), they had not spoken about it or heard the speaker narrate the content previously. The purpose of the increased feedback during common ground content may have been to actively indicate to the speaker that the addressee recognized the content and to affirm that it was mutually shared. Furthermore, the increase in addressee feedback predicted a decrease in narrative events, demonstrating speakers' sensitivity to this addressee behavior. This additional finding highlights the important influence of the addressee's behavior on the speaker's language use (Bavelas, Coates, & Johnson, 2000).

In contrast to the younger adults, older adults hardly differed in the number of words and narrative events they used to talk about known versus novel story content. Also, they made fewer explicit references to common ground in the CG condition than the younger adults, meaning they were less likely to verbally mark mutually shared knowledge for their addressee. We had expected this effect of age on verbal recipient design based on earlier studies showing that older adults are less good at establishing conversational common ground incrementally than younger adults (Horton & Spieler, 2007; Hupet et al., 1993; Lysander & Horton, 2012). In principle, two kinds of explanations for these behavioral differences are conceivable: Older adults may not be able to engage in common ground-related recipient design as we induced it here due to age-related cognitive limitations, but they may also respond differently to the communicative situation than younger adults due to other factors. We explore both options in the following paragraphs.

Based on previous research (e.g. Healey & Grossmann, 2016; Horton & Spieler, 2007; Hupet et al., 1993; Wardlow, 2013; Long et al., 2018), we had hypothesized that verbal WM and inhibitory control influence the ability to engage in recipient design. Deficits in verbal WM may limit the extent to which speakers can focus their resources on considering which information is or is not mutually shared when designing their utterances, and on adapting the utterances accordingly. Deficits in inhibitory control may be related to difficulties in inhibiting the speaker's own, egocentric perspective or suppressing irrelevant information, both of which are necessary for recipient design to occur (Brennan et al., 2010; Keysar et al., 1998). As older adults in the present study had significantly lower verbal WM and inhibitory control than younger adults, this might have contributed to their lack of verbal addressee-based adaptations. However, we could find no support for this hypothesis, as neither of the two cognitive abilities could predict differences in verbal behavior. Of course, the small sample size employed in the present study limits our ability to interpret this absence of an effect. Additionally, it is possible that the particular tasks that we used to assess verbal WM and inhibitory control do not tap into the actual processes involved in verbal recipient design.

Nevertheless, as we found no support for the cognitive account, it is necessary to consider alternative explanations for the older adults' behavior. Previous research suggests that age-related differences in communicative behavior may also be related to differences in social or pragmatic goals (e.g. Adams, Smith, Pasupathi, & Vitolo, 2002; Horton & Spieler, 2007; James, Burke, Austin, & Hulme, 1998). For example, older adults may have had the primary goal of narrating the story "well", therefore giving equal weight to both known and unknown story content in their narrations, whereas younger adults may have focused primarily on being concise and providing information that the addressee did not yet have (see e.g. James et al., 1999, who found that older adults are judged to be better at story telling than younger adults). Another possibility is that older adults may have wished to demonstrate that they remembered all parts of the story well and thus could perform well on the story telling task in general, as beliefs about age-related memory decline are widespread, also among older adults (e.g. Lineweaver & Hertzog, 1998). This desire may have overruled any common ground-based adaptations of their speech. Finally, the fact that older speakers always narrated the stories for older addressees may also have influenced their verbal behavior. Potentially, older speakers may have thought that their addressees could not remember all of the mutually shared content due to memory limitations and therefore refrained from reducing verbal content in the CG condition. Previous research shows that older adults adapt their verbal utterances based on addressee characteristics such as age (Adams, Smith, Pasupathi, & Vitolo, 2002; Keller-Cohen, 2014) or mental retardation (Gould & Shaleen, 1999). Future research could address this possibility by testing mixed age pairs in order to see whether older speakers adapt their speech differently for younger addressees (and younger speakers differently for older addressees).

2.4.2. Effects of age and common ground on multimodal recipient design

As in verbal recipient design, younger and older adults also differed in how they adapted their representational gesture use to their addressee's knowledge state. Younger adults gestured at a higher rate and produced more multimodal utterances when communicating novel as opposed to mutually shared content, similar to the findings by Jacobs and Garnham (2007) (but see Holler & Bavelas, 2017, for a summary of the range of different effects common ground can have on gesture). This increase in multimodal information appears to be a direct effect of speakers adapting to the addressee's knowledge state, providing the addressee with a comprehensive verbal and visual representation of the novel part of the story, and a verbally and especially visually reduced representation when talking about familiar content. It is additionally interesting to note that even though the CG condition was associated with an increased amount of addressee feedback, which in turn predicted an increase in gesture rate, this did not eliminate the effect of common ground on gesture rate. Taken together, these findings illustrate that younger adults could flexibly adapt not only their speech, but also their gestures to the communicative requirements of the situation (Kendon, 1985; 2004).

For older adults, we observed a pattern opposite to that of the younger adults: They tended to gesture at a lower rate and produced fewer multimodal events when talking about novel content, both compared to their own production for shared content and compared to younger adults' production for novel content. We had expected that older adults would show a smaller common ground effect on gesture production than younger adults, based on our predictions for verbal audience design and on the hypothesis that speech and gesture function as one integrated system (McNeill, 1992; Kita & Özyürek, 2003). Therefore, it is surprising that we found common ground to influence older adults' gesture production in this opposite direction, also considering the absence of an effect on their speech. One possible explanation for this finding is that relating novel story content required more cognitive effort than relating mutually shared content. Older adults may have been aware that they should provide more information, yet failed to do so verbally, potentially due to memory limitations. This presumed increase in cognitive load associated with the novel content condition may have led to a reduction in multimodal events, as gestures produced primarily for the benefit of an addressee may actually be cognitively costly to the speaker (Mol, Krahmer, Maes, & Swerts, 2009). However, this speculation rests on the assumptions that the gestures produced during this narrative task were primarily intended to illustrate the story for the addressee, and that older adults failed to engage in verbal recipient design due to cognitive limitations, which we could not find evidence for (but due to our sample size, this needs to be followed up with future research, see previous section).

The present study shows that younger and older adults differ in how they adapt speech and gestures to the common ground shared with an addressee. Ultimately, it seems likely this behavior is determined by a combination of cognitive and social or pragmatic factors (see also Horton & Spieler, 2007). Based on the design of the present study, however, we cannot tease the individual contributions of these two factors apart. First of all, our ability to interpret the absence of cognitive effects is limited by the small sample used in our study. Additionally, it might be that the cognitive tests we used did not capture the abilities that are involved in recipient design. Also, in this study, we did not assess what the speakers' goals and intentions were, and whether there were systematic differences between younger and older adults with respect to this. Thus, while the present study provides clear evidence of age-related differences in multimodal recipient design, we currently can only provide some preliminary ideas on what causes these. Future studies are needed which include larger samples and a broader range of interactive tasks and measures.

2.4.3. General effects of age and cognitive abilities on gesticulation

Despite the age-related difference in how speakers adapted multimodally to common ground, younger and older adults did not differ in terms of representational gesture rate or the percentage of multimodal narrative events they produced overall. The analyses of gesture frequency and gesture rate per narrative event yielded identical results (see

Appendix C). Thus, our results are not in line with the earlier finding that older adults gesture less than younger adults overall (Cohen & Borsoi, 1996; Feyereisen & Havard, 1999; Theocharopoulou et al., 2015). We would like to propose that the difference in findings is due to the communicative paradigm we employed. Whereas participants in the previous studies on gesture production in aging either had no addressee at all or an experimenter-addressee, in the present study we used co-present, non-confederate addressees. Previous research with younger adults indicates that the presence of a visible, attentive addressee increases the relative frequency of representational gestures (e.g. Jacobs & Garnham, 2007; Kuhlen, Galati, & Brennan, 2012). In the current study, older and younger addressees differed only marginally with respect to the amount of feedback they gave, and in both age groups, an increase in addressee feedback was predictive of an increase in gesture rate. Certainly, this should be considered gestural recipient design (as has been argued for effects of addressee feedback on gesture form, Holler & Wilkin, 2011), albeit not the kind of common ground-based recipient design that we intended to investigate through our experimental manipulation.

As younger and older adults did not differ in how much they gestured in relation to speech, there was also no support for the hypothesis that older adults produce more gestures than younger adults in order to compensate for their relative deficit in cognitive abilities, based on accounts of gestures being cognitively beneficial (Chu et al., 2014; Gillespie et al., 2014; Goldin-Meadow et al., 2001; Hostetter & Alibali, 2007; Wagner et al., 2012). Additionally, we found no associations between verbal WM, visuo-sequential WM, or semantic fluency and gesticulation as we assessed them. The field would benefit from a broader investigation of the relationship between cognitive abilities and gesticulation in older adults, using a wider range of gesture elicitation tasks and of cognitive measures (as well as the large sample required for investigating individual differences), similar to previous work with younger adults (Chu et al., 2014; Gillespie et al., 2014).

Nevertheless, the fact that in the absence of shared knowledge, older adults gestured less than younger adults, might be an indication that older adults reduce their gesture production in contexts that induce a higher cognitive load. Future work is needed to test this possibility.

2.5. Conclusion

The present study offers a first glimpse of how aging affects multimodal recipient design in the context of common ground. In an interactive setting, older adults spoke as much and gestured as frequently in relation to speech as younger adults, and were similarly sensitive to addressee feedback on the whole. However, only younger adults adapted both their speech and gesture use for their addressee based on the mutually shared knowledge established at the outset of the interaction, such that they provided relatively less multimodal information when there was shared knowledge, and relatively more multimodal information when there was not. Older adults did not adapt their speech based on the addressee's knowledge state and conveyed less, rather than more, multimodal information in the absence of shared knowledge.

If we take younger adults' behavior in this task as the baseline against which to compare the older adults, we must conclude that older adults failed to engage in successful common ground-based recipient design. That is, while younger adults flexibly adapted both their speech and their gestures to the communicative requirements of the situation, older adults appeared less flexible in the way they drew on their different communicative modalities. We attribute these behavioral differences at least in part to age-related changes in social or pragmatic goals, as they were not reliably predicted by the significant age-related differences in cognitive abilities. Yet, we acknowledge our limited sample size and do not want to exclude the possibility of a cognitive explanation for some findings, such as that older adults produced fewer multimodal events in the absence of shared knowledge.

Our findings raise the question of whether the age-related differences in verbal and gestural patterns found here persist in other types of communicative tasks where common ground builds up incrementally, and whether they have an impact on how older adults are comprehended by others, both young and old.

Acknowledgements

We would like to thank Nick Wood for his help with video editing and Renske Schilte for her assistance with transcribing the data. We would also like to thank Alexia Galati and one anonymous reviewer for their helpful comments and suggestions on earlier versions of this manuscript.

Chapter 3

Working memory and semantic fluency predict younger and older adults' multimodal recipient design in a spatial task

Abstract

Aging appears to impair the ability to adapt speech and gestures based on knowledge shared with an addressee (common ground-based recipient design) in narrative settings. Here, we test whether this extends to spatial settings and is modulated by cognitive abilities. Younger and older adults gave instructions on how to assemble 3D-models from building blocks on six consecutive trials. We induced mutually shared knowledge by either showing speaker and addressee the model beforehand, or not. Additionally, shared knowledge accumulated across the trials. Younger and crucially also older adults provided recipient-designed utterances, indicated by a significant reduction in the number of words and of gestures when common ground was present. Additionally, we observed a reduction in semantic content and a shift in cross-modal distribution of information across trials. Rather than age, individual differences in verbal and visual working memory and semantic fluency predicted the extent of addressee-based adaptations. Thus, in this spatial task, individual cognitive abilities modulate the interactive language use of both younger and older adults.

This chapter is based on: Schubotz, L., Özyürek, A., and Holler, J. Working memory and semantic fluency predict younger and older adults' multimodal recipient design in an interactive spatial task (under review).

3.1. Introduction

Mutually shared knowledge between a speaker and an addressee (their *common ground*, Clark, 1996) affects how speakers speak and gesture in interaction with others. Previous research suggests that this addressee-based adaptation of utterances, or *recipient design* (Sacks, Schegloff, & Jefferson, 1974), is modulated by normal human aging, such that older adults are less capable of engaging in successful recipient design than younger adults (Horton & Spieler, 2007; Hupet, Chartraine, & Nef, 1993; Lysander & Horton, 2012; Saryazdi, Bannon, & Chambers, 2019). Recent work employing a face-to-face, narrative setting suggests that this extends to the gestures accompanying speech (Schubotz, Özyürek, & Holler, 2019). However, it remains unclear whether these behavioral differences in speech also manifest in other communicative settings. Therefore, the aim of the present study was to investigate the extent of older and younger adults' verbal and gestural recipient design in a face-to-face interactive, spatial task, and to determine whether the communicative behavior in this context is modulated by age-related differences in cognitive abilities.

3.1.1. Age-related differences in verbal and gestural recipient design

Previous research suggests that there are systematic differences in how younger and older adults adapt their speech and their co-speech gestures based on knowledge shared with an addressee. Younger and older adults' addressee-related adaptations in the spoken modality (verbal recipient design) have been investigated using referential communication tasks, in which common ground builds up gradually as a result of repeatedly referring to the same set of referents over the course of several trials (incremental common ground, e.g. Clark & Wilkes-Gibbs, 1986; Horton & Spieler, 2007; Hupet et al., 1993; Lysander & Horton, 2012). Younger adults are generally found to interact increasingly more efficiently, indicated by shorter utterances and fewer dialogue turns on later compared to earlier trials, as common ground accumulates. Although older adults' interactions follow the same general pattern of reduction, their interactions are characterized by longer utterances, more dialogue turns, and/or more errors relative to younger adults (Horton & Spieler, 2007; Hupet et al., 1993; Lysander & Horton, 2012). Additionally, unlike younger adults, older adults failed to produce appropriate common ground-based utterances in a subsequent task which involved familiar and new addressees (Horton & Spieler, 2007). In summary, older adults have been found to be overall less efficient and require greater effort to establish common ground than younger adults in the spoken modality.

Younger and older adults' addressee-related adaptations in the gestures accompanying their speech (gestural recipient design) have so far only been compared in one recent study (Schubotz et al., 2019). As interactive, face-to-face language use comprises of speech and co-speech gestures, taking this additional visual modality into account is crucial. Co-speech gestures are meaningful hand movements that accompany speech and contribute to the meaning of utterances in important ways via their semantic or pragmatic content (Kendon, 2004; McNeill, 1992). Schubotz et al. (2019) investigated younger and older adults' verbal and gestural adaptation to common ground, by asking younger and older participants to narrate short comic stories to a same-aged addressee. Mutually shared knowledge was induced by showing both participants part of the story at the start of each trial, while only the designated narrator saw the full story subsequently (a form of personal common ground, see Clark, 1996). Only younger, but not older adults, provided longer and more informative narrations and gestured at a higher rate when relating unknown as opposed to mutually shared story content. Older adults showed no evidence of common ground-based recipient design either in their speech or in their gestures, and even produced fewer rather than more gestures in relation to speech when relating novel story content in this task.

The role of cognitive abilities in recipient design

Such changes in communicative behavior, which may be taken to reflect the absence of appropriate recipient design, or the failure to take the addressee's knowledge state into account, have previously been speculated to be caused by cognitive aging, such as age-related deficits in working memory (WM) or inhibitory control (e.g., Healey & Grossman, 2016; Hupet et al., 1993; Horton & Spieler, 2007; Long, Horton, Rohde, & Sorace, 2018; Wardlow, 2013). WM may be involved in the ability to establish an addressee's perspective and to incorporate this perspective during online language processing, while inhibitory control may be involved in the ability to inhibit one's own perspective in favor of the addressee's (e.g., Brennan, Galati, & Kuhlen, 2010; Keysar, Barr, & Horton, 1998; see also Brown-Schmidt, 2009, for the role of executive function in perspective-taking during language comprehension).

Yet, Schubotz et al. (2019), could not establish a relationship between measures of verbal WM or inhibitory control and older adults' lack of verbal and gestural recipient design. Although it is possible that the measures employed did not capture the abilities involved in the task at hand, they also considered the possibility that, beyond changes in cognitive abilities, differences in communicative goals may have determined how older adults design their utterances for others (e.g., Adams, Smith, Pasupathi, & Vitolo, 2002; Horton & Spieler, 2007; Underwood, 2010; see also Long, Rohde, & Rubio-Fernandez, 2020, for an account of how differences in communicative strategies affect older adults' language use). While the younger participants in Schubotz et al. (2019) presumably focused mainly on information transfer, i.e. providing the addressee with information she did not yet have, older adults may have interpreted it as a task where it mattered to be a 'good story teller', in the sense of providing an easy-to-follow narrative and being clear and exhaustive in terms of story events. That is, aspects like the wish to narrate a nice and complete story may have overruled common ground-based adaptations of speech and gesture.

3.1.2. Spatial vs. narrative task demands during multimodal utterance design

One aspect which likely affects speakers' multimodal language use is the type of communicative task they wish to accomplish, e.g., whether this task is predominantly narrative (such as a story-telling task) or spatial (such as providing spatial descriptions or instructions). Associated task demands may not only modulate the use of the different modalities during utterance production, but also the extent to which WM and other cognitive resources are taxed. For example, observations of younger (e.g. Alibali, 2005) and older speakers (Feyereisen & Havard, 1999) show that people gesture more frequently when they talk about spatial topics, including visual and motor imagery, as opposed to verbal, abstract ones. This suggests that in spatial tasks, information is distributed differently across the two modalities. Gestures carry relatively more communicative weight and might therefore also be more relevant for successful recipient design. It is currently unclear how this distribution of information is organized, particularly for older adults, and how it is affected by pragmatic factors, such as common ground.

Furthermore, in spite of this higher gesture frequency observed during spatial descriptions, previous research involving visual and motor imagery in a monolog setting (Cohen & Borsoi, 1996; Feyereisen & Havard, 1999; Theocharopoulou, Cocks, Pring, & Dipper, 2015) found older adults to produce relatively fewer depictive gestures than younger adults overall. Schubotz et al. (2019), on the other hand, found older adults to gesture as frequently as younger adults in the narrative task. It remains to be seen whether, given an interactive, face-to-face setting, older adults' gesture frequency is comparable to that of younger adults, also in a spatial task.

In addition to these direct effects on multimodal utterance production, visuo-spatial tasks may also differ from narrative tasks in terms of the involvement of cognitive abilities. Visuo-spatial tasks presumably rely more strongly on visuo-spatial cognitive abilities. While it appears that certain abilities such as visuo-spatial perception or mental imagery maintenance undergo only minor age-related decline, spatial WM is more strongly affected by age-related changes (for a review see Klencklen, Després, & Dufour, 2012). Whether potential age-related differences in spatial cognition affect the use of spatial language remains unclear (see Markostamou, Coventry, Fox, & McInnes, 2015).

Finally, previous research on younger adults suggests that gesticulation allows speakers to "off-load" information onto visual space, thereby freeing up cognitive resources more generally (e.g., Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Wagner Cook, Yip, & Goldin-Meadow, 2012). This suggests that the potentially higher gesture rates associated with a spatial task, relative to a narrative task, might allow older adults to engage more cognitive resources towards recipient design.

In summary, it remains an open question how older adults' interactive use of speech and co-speech gestures in a visuo-spatial task compares to that of younger adults, and whether and to what extent communicative behavior, including multimodal recipient design, is modulated by cognitive abilities.

3.1.3. The present study

In the present study, we employed an interactive, spatial task in order to investigate whether the previously observed age-related differences in recipient design, spoken as well as gestured, extend to the spatial domain, and whether potential age-related differences in behavior can be attributed to differences in cognitive abilities.

In order to address these issues, we designed an interactive task in which a primary participant (the speaker) assembled little wooden castles from a set of building blocks and subsequently instructed a secondary participant (the addressee) on how to assemble the same castles. Mutually shared knowledge between speaker and addressee was manipulated per trial, by either showing both participants a picture of the to-be constructed model shortly at the beginning (common ground [CG]), or not (no common ground [no-CG]). We thus induced a form of personal common ground (Clark, 1996), in which the mutually shared knowledge existed from the outset of the interaction. Additionally, as speaker and addressee interacted over six consecutive trials, and the speaker referred to the same entities repeatedly, we also expected common ground relating to the individual building blocks and the steps of how to assemble them to build up incrementally (incremental common ground, Clark, 1996).

Apart from this additional possibility for incremental common ground to develop in the course of the experiment, the task employed here differed from the narrative task employed in Schubotz et al. (2019) in several other ways. First of all, as argued above, a visuo-spatial task differs substantially from a narrative task, both in terms of how information can be distributed across the two modalities, and in terms of the cognitive functions involved in the task. Furthermore, in the present task, the goal (i.e., give addressee instructions on how to assemble the model) allowed less room for individual interpretation and therefore minimized the likelihood that potential age-related differences in behavior could be attributed to age-related differences in task interpretation or communicative goals.

For our manipulation of personal common ground, we expected younger adults to produce fewer words and fewer gestures and to convey less information in the two modalities when providing instructions for previously seen models as compared to unseen models (Galati & Brennan, 2010; 2014; Holler & Wilkin, 2009; Schubotz et al., 2019). For the effects of incremental common ground, we expected the repeated references to the same entities and assembly steps over the course of the experiment to result in increasingly shorter utterances (e.g. Clark & Wilkes-Gibbs, 1986; Fussell & Krauss, 1989) and fewer gestures (Galati & Brennan, 2014), in concordance with a reduction in information content in both modalities. We made no directional predictions relating to gesture rate (i.e., number of gestures per 100 words), since this relation has been differently affected in previous studies (see Holler & Bavelas, 2017).

Due to the previously observed absence of verbal and gestural adaptations to personal common ground in older adults (Schubotz et al., 2019; also Horton & Spieler, 2007), we expected older adults to be less adaptive than younger adults to mutually shared knowledge induced on individual trials, not only in their speech but also in the way they draw on gesture when designing utterances for knowing vs. unknowing recipients. However, due to differences in task demands/design, particularly the spatial nature of the present task and the way in which the task goal was formulated, these age-related effects may be less pronounced than those obtained in Schubotz et al. (2019). In terms of verbal and gestural adaptations to incremental common ground, we expected older adults to show an overall pattern of reduction similar to that of younger adults, although this may be less pronounced than in younger adults (e.g., Hupet et al., 1993).

We additionally assessed how information was distributed across the two modalities, i.e. whether information was expressed uniquely in speech, uniquely in gesture, or whether it was expressed in both modalities. This provides an indication of the relative communicative weight that gestures carry and can additionally been seen as an indicator of recipient design: encoding the same piece of information twice, in both modalities, is arguably more informative than encoding information in only a single modality (see also de Ruiter, Bangerter, & Dings, 2012). Therefore, we expected that younger participants would encode more information in both modalities for unknowing as opposed to knowing addressees. Similarly, we expected younger adults to encode increasingly fewer pieces of information in both modalities across the experiment, as common ground incrementally accumulates. Again, these effects may be smaller or absent in the older adults.

In order to test for the role of cognitive abilities in speech and co-speech gesture use and their adaptation to common ground, we assessed speakers' verbal and visual WM as well as inhibitory control and semantic fluency. As summarized above, verbal WM and inhibitory control have previously been related to verbal recipient design, as well as visual perspective taking (e.g. Long et al., 2018; Wardlow, 2013; Hupet et al., 1993). Our expectation was that higher verbal WM and inhibitory control might be associated with more pronounced verbal and/or gestural recipient design and that potential age-related differences in verbal and gestural behavior may be attributable to age-related differences in these cognitive functions.

Furthermore, verbal and visual WM have previously been related to gesticulation in general (e.g. Chu, Meyer, Foulkes, & Kita, 2014 for visual WM; Gillespie, James, Federmeier, & Watson, 2014 for verbal WM), such that lower cognitive abilities lead to higher gesture frequencies (for the cognitively beneficial effects of gesticulation see also e.g. Goldin-Meadow et al., 2001; Wagner et al., 2012). Similarly, lower semantic fluency is an indicator of word finding difficulties, which in turn may be associated with an increase in gesticulation (Rauscher, Krauss, & Chen, 1996). We included these measures in order to be able to control for the possibility that potential age-related differences in the interactive use of co-speech gestures are attributable to age-related differences in these cognitive functions.

3.2. Method

3.2.1. Participants

The same participants as in Schubotz et al. (2019) participated in the present experiment: thirty-two younger adults (16 women) between 21 and 30 years old (M_{age} = 24.31, SD = 2.91) and 32 older adults (16 women) between 64 and 73 years old (M_{age} = 67.69, SD = 2.43). All participants were native Dutch speakers with self-reported normal or corrected-to-normal vision and hearing and no known history of neurological impairment. Each participant was allocated to a same-age and same-sex pairing. The role of speaker or addressee was randomly assigned and kept constant across the entire experiment. Only the speaker data was analyzed here. All participants in the role of speaker had minimally secondary school education, except for one older participant who only had primary school education. Participants were recruited from the participant pool of the Max Planck Institute for Psycholinguistics and received between € 8 and € 16 for their participation, depending on the duration of the session. The experiment was approved by the Ethics Commission for Behavioural Research from Radboud University Nijmegen. Signed consent was acquired from all participants.

3.2.2. Materials

We created six black-and-white line drawings of simple castle-like buildings (for examples see Figure 1). Each castle could be assembled from seven wooden building blocks, all of the same color: two cubes, two rectangular prisms, two triangular shapes (right triangles), and one arc-shaped block, ranging in size from 4 x 4 x 4 cm (cubes) to 4 x 4 x 12 cm (arc). The buildings were constructed such that the two triangular shapes always formed the top of the building, the position of the remaining building blocks varied. All models were fully symmetrical. We intentionally kept the models simple in order to ensure that older adults would be able to memorize them correctly in spite of potential age-related memory deficits.

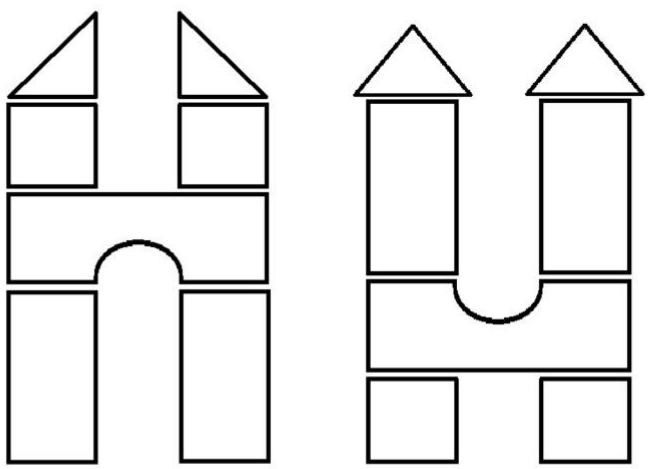


Figure 1. Two of the six stimuli used in the experiment.

Four experimental lists determined the order in which the different models were presented. Initially, we created two orders of presentation for the six models, one being the reverse of the other. Counterbalancing the order of common ground presentation across lists resulted in four experimental lists. Each list was tested eight times, distributed evenly across age groups and sexes.

3.2.3. Procedure

Upon arrival, the speaker and the addressee were asked to sit in designated chairs at a table at 90° from each other. Two video cameras were set up on tripods at a small distance from the table, one capturing a frontal view of the speaker, the other one positioned such that it captured both speaker and addressee (see Figure 2 for stills from the two cameras). Sound was recorded with an additional microphone suspended from the ceiling over the table and connected to the speaker camera.

Figure 2. Example of the lateral (left panel) and frontal (right panel) views of the speaker in the experimental set-up. In this frame, the speaker indicates the size and position of the two small cubes.

All participants completed one practice trial and six experimental trials. At the beginning of the practice trial and of half of the experimental trials (the CG trials), both participants were presented with a line drawing of a model and instructed to look at it carefully for five seconds without talking, with the aim to experimentally induce common ground about the composition of the model. Subsequently, the drawing was removed and a screen was put up on the table between speaker and addressee. The speaker then received the drawing and the seven building blocks and assembled the model according to the drawing. Once the speaker indicated that he/she was done, the experimenter checked the model for accuracy and then took the blocks and the drawing away. The screen was taken off the table and the speaker described to the addressee how to assemble the model, without using the building blocks. Addressees were instructed to listen to the descriptions and ask all clarification questions at the end. Once speaker and addressee had discussed potential questions, the screen was put back up and the addressee received the building blocks in order to assemble the model according to the speaker's instructions. Addressees built the model behind the screen in order to avoid any engagement of the speaker during this process. Additionally, the experimenter took away the construction built by the addressee before removing the screen, and feedback on the addressee's performance was given only at the end of the entire task. This was done in order to avoid any adaptation of the speaker's instructions based on the addressee's performance.

For the other half of the experimental trials (the no-CG trials), the procedure was identical, except for the first step: participants did not see the model picture beforehand, rather, the screen was put up at the beginning of the trial and the speaker received the model picture and the building blocks immediately. Depending on the pair, the task took about 20 to 30 minutes. After the experimental tasks were completed, the addressee was allowed to leave, while the speaker performed the cognitive tests.

3.2.4. Transcription and coding

Speech coding

All recordings from the two cameras were synchronized and subsequently segmented into trials. Transcription of speech and annotation of gestures was conducted in Elan (Version 4.9.4; Wittenburg, Brugman, Russel, Klassmann, Sloetjes, 2006). For all segments, the speaker's initial instruction, i.e. the first complete instruction on how to assemble the model without potential subsequent repetitions, was identified. All analyses reported here are based on these initial instructions only, discarding repetitions or clarifications elicited by the addressee following the initial instruction, as the focus of our study was the effect of our experimental manipulations on the speakers' behavior rather than the impact of speaker-addressee interaction (for a similar argument see Horton and Gerrig, 2005). Speech from the speaker was transcribed verbatim, including disfluencies such as filled pauses and word fragments. However, disfluencies were excluded from the word counts presented in the results section, as we were mainly interested in speech content and did not want potential age-related differences in the number of disfluencies to influence the word count (e.g. Mortensen, Meyer, & Humphreys, 2006). For this reason, we also distinguished between "narrative speech" belonging to the instruction proper (i.e. relating to the model building itself) and "non-narrative speech", such as statements about the task or comments relating to the speaker or the addressee (for the basis of this distinction see McNeill, 1992). Only speech belonging to the instruction proper entered the word count.

Gesture coding

For the gesture coding, we first identified all co-speech gestures produced by the speaker during the instruction proper, disregarding non-gesture movements as well as gestures accompanying non-narrative speech. Our unit of analyses was the gestural stroke, i.e. the most meaningful part of the gesture determined according to criteria established in previous co-speech gesture research (Kendon, 2004; Kita, van Gijn, & van der Hulst, 1998; McNeill, 1992). We then categorized these strokes as representational and non-representational gestures (see Alibali, Heath, & Myers, 2001). For our purposes, representational gestures include iconic gestures, which iconically depict shape or size of concrete referents or represent physical movements or actions; metaphoric gestures, which resemble iconic gestures but relate to speech in a metaphorical manner (e.g. a rotating movement of the hand to indicate the passing of time); and pointing gestures or deictics, i.e. finger or whole-hand points to a specific location in real or imaginary space, e.g. that of a building block (McNeill, 1992).

All other gestures were considered non-representational and include what are frequently called beat gestures, i.e. biphasic movements of the hand, for example to add emphasis, as well as pragmatic gestures (Kendon, 2004), i.e. gestures which have pragmatic functions, for example to convey information about how an utterance should be interpreted, or relating to the interaction with the addressee (Bavelas, Chovil, Coates, & Roe, 1995; Bavelas, Chovil, Lawrie, & Wade, 1992).

A second coder blind to the experimental hypotheses coded 10% of the trials randomly selected from across all participants for stroke identification, and another 10% of the trials for gesture categorization. Inter-rater agreement on stroke identification was 90.99%. Inter-rater agreement on gesture categorization was 96.43%, Cohen's Kappa = .86.

Representational gesture frequency and gesture rate (gestures per 100 words)

As we were mainly interested in the semantic content of the descriptions and the accompanying gestures, in our analyses we focus exclusively on representational gestures (i.e., iconic, metaphoric, and deictic gestures). In addition to reporting the raw representational gesture frequency, we computed a gesture rate per 100 words (see above for criteria on word count) by dividing the number of gestures by the number of words a given participant produced for each trial and multiplied this by 100. This gesture rate normalizes for differences in instruction length (e.g. Alibali et al., 2001).

Information content coding

In order to assess whether age and common ground (both personal and incremental) affected the information content of the speakers' utterances, we additionally coded for semantic features expressed in speech and in gesture. Per block, several pieces of information could theoretically be encoded: extrinsic features like the block's location and its orientation, and intrinsic features like its shape and its size. The actual scoring of individual features depended on the modality (see below, see also Holler & Wilkin, 2009, for a similar approach to scoring semantic features in speech and co-speech gesture).

Coding of semantic features encoded in speech. For speech, we scored whether the verbal description contained information with respect to three categories, namely each block's location (e.g., "at the bottom", "on top"), its orientation (e.g., "upside down", "vertically"), and the intrinsic features shape and size (e.g., "triangle", "square", "long", "short"). Note that more metaphorical descriptions like "bridge" or "roof" were not counted as conveying orientation or shape information, since these terms refer to objects that may take a variety of shapes and may therefore elicit different visual imagery in different people. For each feature, we scored "1" if the information was present (the maximum score per feature was always "1", even if the information was repeated or rephrased) or "0" if the information was absent. For the small cubes (see Fig. 1), encoding its orientation was not possible, yielding a maximum score of eleven features per description (four blocks à three features, minus one). A second coder blind to the experimental hypotheses recoded 10% of the trials. Inter-rater agreement on scoring of location and of orientation in speech was 100% each, inter-rater agreement on scoring of shape/size in speech was 97.5%, Cohen's Kappa = .92.

Coding of semantic features encoded in gesture. For gestures, we scored whether manual movements contained information with respect to the same semantic aspects but used just two categories, namely a block's location (e.g., pointing to a certain point in space, performing the gesture in the appropriate area in space; gestures had to be spatially coherent with respect to the actual model and with respect to each other) and its orientation, shape, or size (e.g., moving two fingers up and down to indicate a block's vertical orientation, tracing a triangle shape, using two fingers to indicate the size of a block). Unlike for speech, for gesture we collapsed orientation and shape/size, because gestures consistently expressed several aspects at the same time due to their holistic nature, making it difficult to score these aspects separately (e.g., tracing an arc indicates the shape, the size, and the orientation of the arc-shaped block all at the same time). As for speech, we scored “1” if the information was present (“1” was the maximal score, even if several gestures were used to convey different aspects of one feature, e.g. a block’s shape) and “0” if the information was absent. Additionally, in the coding of location, we introduced half a score (“.5”). This was used if the gesture itself encoded the correct location information, but could not be related to a previous gesture, e.g. because there was no previous gesture or because the previous gesture was not performed in the correct location. By introducing this penalty, we aimed to account for whether the descriptions were spatially coherent or not. The maximum score for gesture information was eight (four blocks à two features). The same coder who coded the information in speech also recoded the same 10% of the trials for information content in gesture. Inter-rater agreement on location coding was 92.50%, Cohen’s Kappa = .87, inter-rater agreement on orientation/shape/size coding was 98.75%, Cohen’s Kappa = .97.

Number of semantic features encoded in speech and in gestures

For each description, we computed the sum of semantic features encoded in speech and the sum of semantic features encoded in gesture. These provide an index of the total information that a speaker provided in each modality for each description, independently of what was represented in the respective other modality. In Appendix G, we additionally present an analysis of the normalized counts, i.e. the sums of semantic features encoded in speech and in gesture divided by the number of words and gestures respectively, which provides a measure of “information density” and an index of how efficiently the two modalities are used.

Distribution of information across speech and gestures

Finally, we also computed how many semantic features were expressed in a single modality, i.e. only in speech or only in gestures, and how many semantic features were expressed in both modalities, e.g. by referring to the triangles in speech while at the same time tracing their shape with the fingers. Based on these counts we computed the percentages of information encoded uniquely in speech, uniquely in gesture, or in both modalities, which provides an index of how information is distributed across the two modalities.

3.2.5. Cognitive measures

Participants performed the Operation Span Task (Ospan) as a measure of verbal WM, the animal naming task to assess semantic fluency, the Trail Making Test (TMT) as a measure of inhibitory control, the Visual Patterns Test (VPT) as a measure of visuo-spatial WM, and the Corsi Block Task (CBT) as a measure of visuo-sequential WM. Detailed descriptions of these cognitive tasks, how they were administered, and how the scores were computed can be found in Appendix D. The summary of test scores provided in Table 1 indicates that younger adults outperformed older adults on all measures, except for the semantic fluency test.

Table 1. Mean scores (and SD) per age group on cognitive tests, plus statistical comparisons (independent t-tests and Mann-Whitney tests where appropriate).

	Younger	Older	Test statistic
Verbal WM (Operation span task)	44.06 (8.39)	34.73 (8.92)	$t(29) = 2.99$**
Semantic Fluency (Animal naming test)	31.5 (9.4)	27.75 (5.99)	$t(30) = 1.35$
Inhibitory control (Trail Making Test, TMT)[a b]	14.5 (27)	21.5 (62)	W = 65.5*
Visuo-spatial WM (Visual Patterns Test, VPT)[a]	13 (4)	10 (8)	W = 187.5***
Visuo-sequential WM (Corsi Block Task, CBT)[a]	54 (42)	37.5 (34)	W = 215***

*** = $p < .001$, ** = $p < .01$, * = $p < .05$

[a] Owing to the non-normality of the data, the figures represent Median (and Range).

[b] Note that smaller numbers indicate better performance on this task.

3.2.6. Statistical methods

To investigate the influence of age, personal common ground (CG vs. no-CG trials), and incremental common ground (operationalized as trial number), as well as their interaction effects on the main speech- and gesture-based measures (word and gesture count, gesture rate, information contained in speech, gesture, and speech and gesture combined), we fitted linear mixed-effect models in R version 3.2.1 (R Development Core Team, 2015), using the package lme4 (Bates, Maechler, & Bolker, 2017). We used likelihood ratio tests for model comparisons, eliminating all non-significant interactions in the model comparison process. For each dependent measure, we only report the estimates, SEs, t-values and p-values for the main experimental predictors, as well as other significant predictors and interactions (if applicable). All the models reported contain random intercepts for participants and items, but no by-participant random slopes for the personal common ground manipulation or for incremental common ground (trial number), as this led to perfect correlations of random factors throughout. Reported p-values were obtained via the package lmerTest (Kuznetsova, Brockhoff, and Christensen, 2016).

To investigate the influence of cognitive abilities on our main dependent measures, and to test whether potential age-related differences in verbal and gestural behavior could be attributed to age-related differences in cognitive abilities, we applied the same basic procedure as described above, creating separate models for each cognitive predictor. As the analyses were exploratory, we performed a backwards-model-stripping procedure, starting out with a full model including the z-scored cognitive predictor of interest, age, the common ground manipulation, and trial number, as well as all their interaction terms, eliminating non-significant interactions and predictors based on likelihood ratio tests.

3.3. Results

3.3.1. Descriptive statistics

Out of the 96 instructions participants gave, six contained an error. In all cases, participants confused the position of the cubes with that of the rectangles. Exactly half of the errors occurred in descriptions by older adults, suggesting that overall, younger and older adults were able to memorize the castles equally well.

Younger adults produced a total of 692 gestures, out of which 618 were iconic gestures (89.31%), 3 metaphoric gestures (.43%), 70 abstract deictic gestures (10.12%), 1 concrete deictic gesture (.14%), and 61 non-representational gestures (8.82%). Older adults produced a total of 722 gestures, out of which 651 were iconic gestures (90.17%), 1 metaphoric gesture (.14%), 67 abstract deictic gestures (9.28%), 3 concrete deictic gestures (.42%), and 19 non-representational gestures (2.63%).

Table 2. Means and SDs for dependent measures by age group and personal common ground condition.

	Younger		Older	
	CG	No-CG	CG	No-CG
Number of words	42.4 (18.29)	50.71 (23.14)	39.79 (21.19)	48.94 (32.59)
Number of gestures	6.63 (3.87)	7.79 (5.69)	6.71 (5.21)	8.33 (8.54)
Gestures/100 words	17.24 (11.00)	15.04 (7.56)	18.69 (13.28)	19.14 (12.05)
Speech info total	7.29 (1.75)	7.54 (1.71)	6.44 (2.06)	6.6 (2.33)
Gesture info total	4.59 (2.54)	5.04 (2.67)	4.49 (3.16)	4.71 (3.06)
% Speech unique info	45.66 (29.84)	43.04 (29.36)	46.42 (35.5)	44.07 (34.3)
% Gesture unique info	15.37 (14.92)	14.94 (12.34)	18.18 (18.85)	19.41 (20.07)
% Info speech & gesture	38.98 (22.74)	42.01 (24.95)	35.4 (25.3)	36.51 (26.23)

Mean values and standard deviations for the various dependent measures by age group and personal common ground condition are listed in Table 2 (for means and *SD*s by age group and incremental common ground, see Appendix H). It is interesting to note that for both age groups, the gesture rate (i.e., the gesture frequency normalized by the number of words) was considerably higher in the present task than in Schubotz et al. (2019), where younger adults produced on average 5.89 gestures per 100 words in the CG condition (7.73 in no-CG), and older adults produced on average 5.96 gestures per 100 words in the CG condition (4.88 in no-CG). This difference could be expected, seeing that gestures generally play a more prominent role when talking about spatial topics as opposed abstract or verbal ones (e.g., Feyereisen & Havard, 1999; Lavergne & Kimura, 1987).

3.3.2. Effects of experimental predictors and cognitive abilities on word count, gesture frequency, and gesture rate

Word count

Word count was predicted by personal common ground, such that fewer words were used in CG than in no-CG trials (β = -8.73, *SE* = 2.06, *t*(154.86) = -4.23, *p* < .001). Also, there was a significant effect of incremental common ground, such that fewer words were used on later as compared to earlier trials (β = -4.15, *SE* = .60, *t*(154.86) = -6.87, *p* < .001). There was no main effect of age (β = -2.19, *SE* = 6.49, *t*(31.87) = -.34, *p* = .74), and no interaction between any of the predictors.

Including verbal WM yielded a significant interaction with incremental common ground, such that participants with higher verbal WM showed a stronger reduction in number of words across trials (β = -1.42, *SE* = .60, *t*(150.47) = -2.37, *p* =.02). Note that participants with lower WM did not fail to reduce but rather started out with a lower number of words on early trials which remained constant across the experiment, while participants with higher WM used a higher number of words on early trials and reduced on later trials (see Figure 3). None of the other cognitive predictors contributed significantly to the original model.

Gesture count (gesture frequency)

As for word count, the only significant predictors for gesture count were personal common ground (β = -1.40, *SE* = .58, *t*(154.77) = 2.42, *p* = .02) and incremental common ground (β = -.81, *SE* = .17, *t*(154.77) = -4.79, *p* < .001), with fewer gestures in CG as compared to no-CG trials, and fewer gestures on later trials as compared to earlier trials.

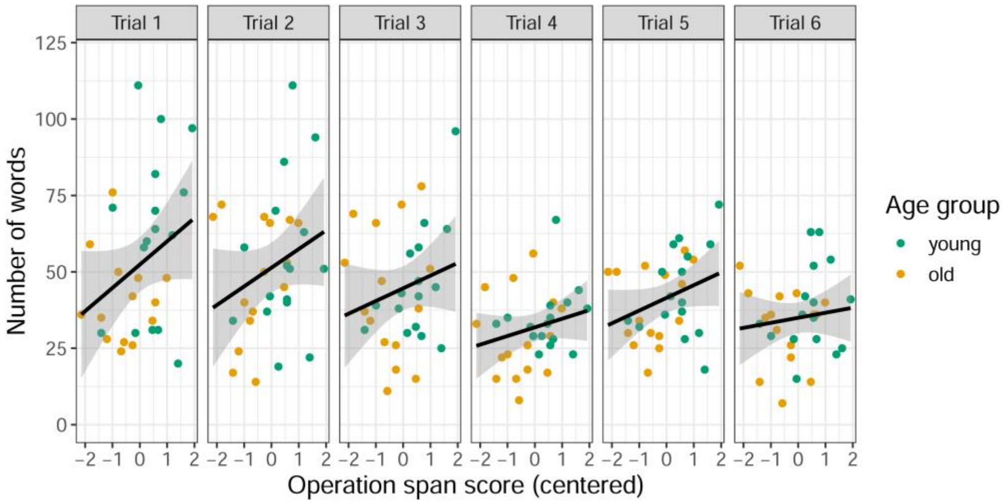


Figure 3. Interaction effect of verbal WM (z-scored) and incremental common ground on word count, collapsed across CG and no-CG conditions. Note that there was no significant effect for age group.

There was no main effect for age (β = .31, *SE* = 1.60, *t*(31.97) = .20, *p* = .85), and no interaction with the other predictors.

Including semantic fluency yielded a significant interaction with personal common ground, such that participants with higher semantic fluency showed a stronger reduction in gesture frequency across trials (β = -.41, *SE* = .17, *t*(154.90) = -2.47, *p* = .01). Note that participants with lower fluency did not fail to reduce but rather started out with a lower number of gestures on early trials which remained constant across the experiment, while participants with higher semantic fluency used a higher number of gestures on early trials and reduced on later trials (see Figure 4). None of the other cognitive predictors contributed significantly to the original model.

Gesture rate (gestures/100 words)

Gesture rate was not significantly predicted by any of the experimental predictors, age (β = 2.77, *SE* = 3.27, *t*(32) = .85, *p* = .40), personal common ground (β = -.88, *SE* =.97, *t*(160) = -.91, *p* = .37), or incremental common ground (β = -.12, *SE* = .28, *t*(160) = -.42, *p* = .68). There were also no effects for cognitive predictors.

For the full model summaries of the analyses reported in this section, see Appendix I.

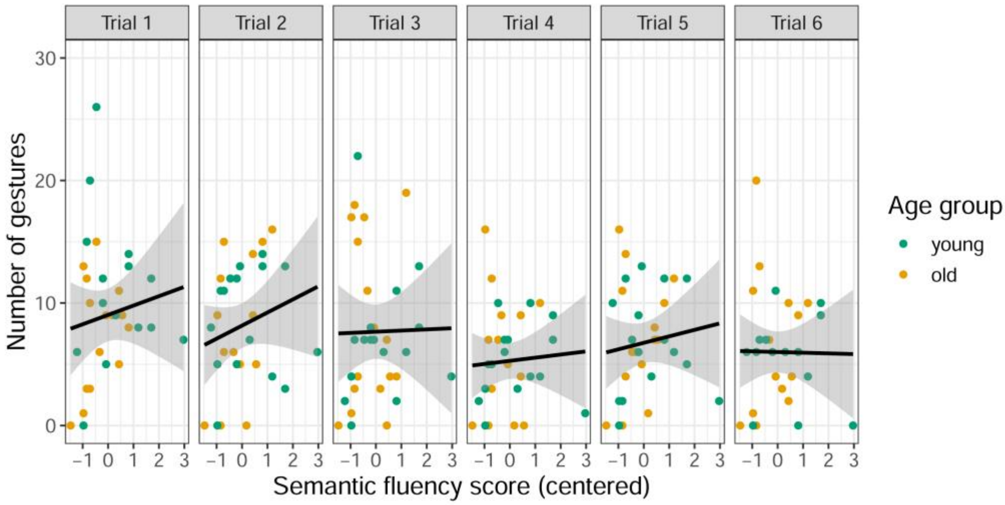


Figure 4. Interaction effect of semantic fluency (z-scored) and incremental common ground on gesture count, collapsed across CG and no-CG conditions. Note that there was no significant effect for age group.

3.3.3. Effects of experimental predictors and cognitive abilities on information encoded in speech and in gestures

Number of semantic features encoded in speech and in gestures

The amount of semantic features expressed in speech was predicted only by incremental common ground, such that later trials contained fewer features (β = -.14, *SE* = .05, *t*(160) = -2.70, *p* =.008). There were no effects for age (β = -.90, *SE* = .56, *t*(32) = -1.17, *p* = .12) or personal common ground (β = .21, *SE* = .18, *t*(160) = 1.17, *p* = .25). Similarly, the amount of semantic features expressed in gesture was predicted only by incremental common ground, such that later trials contained fewer features (β = -.21, *SE* = .06, *t*(154.91) = -3.75, *p* < .001). There were no effects for age (β = -.22, *SE* = .90, *t*(32) = -.24, *p* = .81), or personal common ground (β = .33, *SE* = .19, *t*(154.91) = 1.75, *p* = .08). There were no effects for cognitive factors on either measure.

Distribution of information across speech and gestures

The percentage of information expressed uniquely in speech was predicted by incremental common ground, such that more information was encoded uniquely in speech on later as compared to earlier trials (β = 1.95, *SE* = .64, *t*(154.90) = 3.04, *p* = .003). There were no effects for age (β = .90, *SE* = 10.13, *t*(32) = .09, *p* = .93) or personal common ground (β = -2.48, *SE* = 2.20, *t*(154.90) = -1.13, *p* = .26). Similarly, the percentage of information expressed uniquely in gesture was predicted by incremental common ground, such that more information was encoded uniquely in gestures on later as compared to earlier trials (β = .96, *SE* = .46, $t(154.79) = 2.12$, $p = .04$). There were no effects for age (β = 3.64, *SE* = 4.69, $t(31.99) = .78$, $p = .44$) or personal common ground (β = .41, *SE* = 1.55, $t(154.79) = .26$, $p = .79$). There were no effects for cognitive factors on either measure.

The percentage of semantic features expressed twice, both in speech and in gestures, was predicted by incremental common ground, such that there was a lower percentage of information encoded twice on later as compared to earlier trials (β = -2.92, *SE* = .60, $t(154.85) = -4.88$, $p < .001$). There were no effects for age (β = -4.54, *SE* = 7.13, $t(31.99) = -.64$, $p = .53$) or personal common ground (β = 2.07, *SE* = 2.04, $t(154.85) = 1.01$, $p = .31$).

Including visuo-sequential WM yielded a significant interaction with personal common ground. In CG trials, participants expressed the same percentage of information twice, in both modalities, regardless of visual WM score. However, in no-CG trials, participants with higher visual WM expressed a higher percentage of information twice, in both modalities, than participants with lower visual WM (β = 4.99, *SE* = 2.01, $t(155.68) = 2.48$, $p = .01$).

For the full model summaries of the analyses reported in this section, see Appendix I.

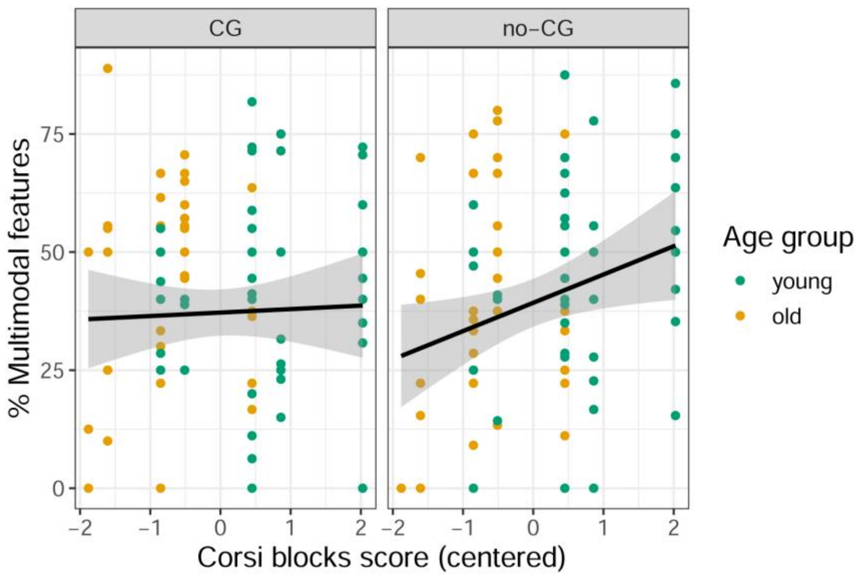


Figure 5. Interaction effect of visual WM (z-scored) and personal common ground manipulation on percentage of information expressed twice, in both modalities, collapsed across trials. Note that there was no significant effect for age group.

3.3.4. Summary of results

Word and gesture frequency and rate

We found no significant age-related differences in the verbal and gestural behavior of younger and older adults. Both younger and older adults' behavior showed significant effects of personal and incremental common ground: there was a significant reduction in word count and gesture frequency in CG as compared to no-CG trials as well as across the experiment, i.e. going from the first to the final trial. For both age groups, a parallel decrease in both modalities from no-CG to CG trials and across the experiment was indicated by a constant gesture rate.

Information encoded in speech and/or gesture

Again, there were no age-related differences in the amount and distribution of information expressed in speech and in gestures.

The number of features (location, size/shape, orientation) expressed in speech and in gestures decreased across the experiment (incremental common ground). However, there was no effect of personal common ground (CG vs. no-CG trials).

With respect to the distribution of information across the two modalities, the percentage of semantic information encoded uniquely in either of the two modalities increased across the experiment, while the information expressed twice, both in speech and in gesture, decreased across the experiment. That is, we saw a shift across the experiment from encoding information in both modalities, to encoding information only in one single modality.

Effects of cognitive predictor variables

Although there were no significant age-related differences in any of these dependent measures and no interaction effects of age group and personal or incremental common ground, we found interaction effects of individual cognitive abilities with the common ground variables. Incremental common ground interacted with verbal WM and with semantic fluency, such that across the experiment, a reduction in word count was more pronounced in individuals with better verbal WM and a reduction in gesture frequency was more pronounced in individuals with higher semantic fluency. Personal common ground interacted with visual WM, such that participants with higher visual WM encoded more information twice, in both modalities, on no-CG trials than participants with lower visual WM.

3.4. Discussion

The present study offers new insights into multimodal recipient design by older and younger adults in a spatial task. Based on previous research, we initially hypothesized that older adults would show less evidence of common ground-based recipient design than younger adults in speech and in gesture in terms of description length and gesture frequency, in terms of information content, and in terms of how the information is distributed across the two modalities. Additionally, we hypothesized that individual differences in cognitive abilities may modulate age-related differences in behavior.

Contrary to our expectations, we found no significant behavioral differences between the two age groups on measures of word and gesture frequency, amount of information expressed in the two modalities or how the information was distributed across the two modalities. Speakers of both age groups adapted their multimodal instructions to our experimentally induced personal and incremental common ground. Rather than by the speakers' age, recipient design in several measures was predicted by individual differences in cognitive abilities. Individual results will be discussed in the following sections.

3.4.1. Effects of age and personal and incremental common ground on multimodal recipient design

As in Schubotz et al. (2019), we found no age-related differences in overall word count or gesture frequency. This suggests that the relatively lower gesture frequency reported previously for older adults in the visuo-spatial domain (Cohen & Borsoi, 1996; Feyereisen & Havard, 1999; Theocharopoulou et al., 2015) was not attributable to task demands but rather to the lack of a truly communicative setting (see also discussion in Schubotz et al., 2019).

Furthermore, and contrary to what we expected based on previous findings (Horton & Spieler, 2007; Hupet et al., 1993; Lysander & Horton, 2012; Schubotz et al., 2019), we also found no age-related differences in verbal and gestural recipient design: both younger and older adults reduced the number of words and of gestures in CG compared to no-CG trials (personal common ground manipulation), as well as across the experiment (incremental common ground manipulation) to the same extent. The parallel decrease in speech and gesture resulted in a constant gesture rate, as has been found in some previous studies (e.g., Campisi & Özyürek, 2013; de Ruiter et al., 2012; Galati & Brennan, 2014; Hilliard & Cook, 2016; see Holler & Bavelas, 2017 for a review). Also, both age groups reduced the amount of information expressed in speech and in gestures across the experiment, i.e., in response to incremental common ground. This reduction in semantic content expressed in speech and in gestures had previously only been observed for manipulations of personal common ground (e.g., Holler & Wilkin, 2009; Schubotz et al., 2019).[9]

Furthermore, we observed a shift in how information was distributed across speech and gesture across the experiment. On earlier trials, speakers encoded more information in both modalities, i.e., the semantic features that were expressed in speech were also expressed in gesture and vice versa. Later on, information was more frequently encoded only in speech or, to a lower percentage, only in gesture. Encoding the same piece of information in both modalities is arguably more informative than encoding it only in one modality. Hence, this pattern mirrors the general observation of speaker-gesturers becoming increasingly efficient in terms of their speech and gesture use across the experiment (i.e. as common ground accrues) and provides an additional example of how well the use of the two modalities is coordinated in recipient-designed messages. The absence of any effects of age on this measure, too, suggests that older adults are as skillful as younger adults with respect to the coordination of information across the two modalities in interactive settings. The analyses of speech and gesture “information density”, reported in Appendix G, further corroborate this observation.

3.4.2. Interaction effects of cognitive abilities with personal and incremental common ground on multimodal recipient design

Rather than age, we found that individual cognitive variables modulated the extent of verbal and gestural recipient design based on personal and incremental common ground. These findings go beyond previous research on the interplay between cognitive and communicative constraints on speech and gesture use (e.g., Galati & Brennan, 2014; Masson-Carro, Goudbeek, & Krahmer, 2016), as they identify individual differences in specific cognitive variables which influence common ground-based adaptations, rather than inducing an external cognitive load by increasing task demands. It is interesting that these associations surfaced in a visuo-spatial task (see also Long et al., 2018; Wardlow,

[9] Interestingly, in the present study, the same amount of verbal and gestural information was expressed in CG as compared to no-CG trials. It appears that regardless of personal common ground condition, speakers always deemed the same amount of information minimally necessary in order to construct the model, which may make this finding specific to our task – after all, the present task was restricted to just four semantic aspects that were relevant for completing the task.

2013), but not in a more verbal, narrative task (Schubotz et al., 2019). Possibly, the cognitive measures employed here were better suited to capture the abilities involved in the present task as compared to the narrative task, due to the different cognitive abilities involved.

Verbal WM influenced how strongly speakers reduced the number of words across the experiment, i.e., in response to incremental common ground: Individuals with higher verbal WM showed a stronger pattern of reduction than those with lower verbal WM. Presumably, WM resources are needed to update the speaker's discourse model on which information is or is not mutually shared, and to access this information while designing and adapting one's utterances accordingly (see also e.g. Brennan et al., 2010; Horton & Gerrig, 2005; Wardlow, 2013).

Semantic fluency modulated the reduction of gesture frequency in response to incremental common ground: Participants with higher semantic fluency showed stronger evidence of gestural adaptations to incremental common ground than those with lower semantic fluency. Potentially, higher semantic fluency, i.e. the efficiency of accessing and retrieving words from existing semantic categories (Martin, Wiggs, Lalonde, & Mack, 1994), allowed speakers to be more flexible in how they used gestures in addition to their verbal message. For example, Hostetter and Alibali (2007) suggest that speakers with high verbal skill may use gestures to make their utterances more communicatively effective, as may also have been the case in the present study.

Finally, we also found that visual WM affected the distribution of information across the two modalities based on personal common ground: individuals with better visual WM encoded more information in both modalities for unknowing addressees than individuals with lower visual WM. This suggests that visual WM, i.e., the ability to store and manipulate visual information, also influences how well speakers can use speech and gesture together for their addressee. We would like to speculate that there might be a mechanism similar to the one proposed for the effects of semantic fluency above: Speakers with higher visual WM may have been more efficient at storing and retrieving the visual information from memory due to their higher spatial skills, and were thus able to use gestures more flexibly in order to tailor their multimodal utterances to their addressees' needs.

Interestingly, our findings are not in line with earlier research suggesting a direct relationship between lower visual or verbal WM (Chu et al., 2014; Gillespie et al., 2014 respectively) and an increase in gesture frequency. It is likely that in other contexts, in which the communicative or interactive function of gestures is less emphasized, the relationship between cognitive abilities and gesticulation manifests itself differently. Yet, note that our findings are based on a relatively small sample. Ideally, future research should replicate these results, using larger sample sizes.

3.5. Conclusion

Taken together, our results indicate that like younger adults, older adults were aware of the presence or absence of shared knowledge induced experimentally, i.e., personal and incremental common ground, and could adapt their multimodal utterances accordingly. Additionally, our findings suggest that younger and older adults' common ground-based adaptations were affected by individual differences in cognitive abilities, with higher cognitive performance in verbal and visual WM and semantic fluency allowing for more strongly pronounced recipient design.

Thus, previous findings of age-related deficits in common ground-based recipient design in the verbal (Horton & Spieler, 2007; Hupet et al., 1993; Lysander & Horton, 2012; Schubotz et al., 2019) and gestural domain (Schubotz et al., 2019) do not seem to generalize to the spatial task employed in the present study. First and foremost, by virtue of being spatial, the present task presumably placed different demands on the speech and co-speech gesture production system. The fact that gestures were very prominent during the spatial descriptions may have given speakers the opportunity to "off-load" information onto visual space, thereby freeing up cognitive resources (see Goldin-Meadow et al., 2001; Wagner et al., 2012), which then became available for other cognitive operations, like the common ground-based adaptation of utterances. Furthermore, the language used in the present task consisted of a fairly restricted vocabulary, consisting mainly of geometric shape and size attributes and spatial prepositions; this may have additionally decreased the demands of verbal utterance planning, thus leaving more capacity for the cognitive operations involved in recipient design. In addition, the straight-forward nature of the present task presumably reduced age-related differences in task interpretation and communicative goals, which may have contributed to the results obtained by Schubotz et al. (2019).

We would like to suggest that this interplay of cognitive and contextual factors determined older adults' communicative behavior, causing the different pattern of results observed in the present task compared to Schubotz et al. (2019). Future research might further explore this possibility, by systematically manipulating the type of cognitive factors involved in a given task, the task difficulty, and the speakers' communicative goals.

To summarize, in the present study, we found no evidence that the ability to engage in common ground-based recipient design, both verbally and gesturally, decreases as a function of age. In the spatial instruction task that we employed, both age groups flexibly adapted their speech and co-speech gesture use and the amount of information they expressed in the two modalities according to their addressee's knowledge state in terms of personal and incremental common ground. Importantly, individual differences in verbal and visual WM and semantic fluency modulated the extent of these addressee-based adaptations, such that higher cognitive abilities predicted more strongly pronounced recipient design. We conclude that a combination of context-specific communicative requirements and of cognitive factors determines how younger and older adults speak and gesture in interaction with others.

Acknowledgements

We would like to thank Nick Woods for his help with video editing, Renske Schilte for her assistance with transcribing the data, and Victoria Poulton and Patrick Hobbelen for their help with the reliability coding.

Chapter 4

Aging and working memory modulate the ability to benefit from visible speech and iconic gestures during speech-in-noise comprehension

Abstract

When comprehending speech-in-noise (SiN), younger and older adults benefit from seeing the speaker's mouth, i.e. visible speech. Younger adults additionally benefit from manual iconic co-speech gestures. Here, we investigate to what extent younger and older adults benefit from perceiving both visual articulators while comprehending SiN, and whether this is modulated by working memory and inhibitory control. Twenty-eight younger and 28 older adults performed a word recognition task in three visual contexts: mouth blurred (speech-only), visible speech, or visible speech + iconic gesture. The speech signal was either clear or embedded in multitalker babble. Additionally, there were two visual-only conditions (visible speech, visible speech + gesture). Accuracy levels for both age groups were higher when both visual articulators were present compared to one or none. However, older adults received a significantly smaller benefit than younger adults, although they performed equally well in speech-only and visual-only word recognition. Individual differences in verbal working memory and inhibitory control partly accounted for age-related performance differences. To conclude, perceiving iconic gestures in addition to visible speech improves younger and older adults' comprehension of SiN. Yet, the ability to benefit from this additional visual information is modulated by age and verbal working memory. Future research will have to show whether these findings extend beyond the single word level.

This chapter is based on: Schubotz, L., Holler, J., Drijvers, L. and 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*. doi:10.1007/s00426-020-01363-8.

4.1. Introduction

In every-day listening situations, we frequently encounter speech embedded in noise, such as the sound of cars, music, or other people talking. Relative to younger adults, older adults' language comprehension is often particularly compromised by such background noises (e.g. Dubno et al., 1984). However, the visual context in which speech sounds are perceived in face-to-face interactions, particularly the speaker's mouth movements and manual gestures, may facilitate the comprehension of speech-in-noise (SiN). Both younger and older adults have been shown to benefit from visible speech, i.e. the articulatory movements of the mouth (including lips, teeth and tongue) (e.g. Sommers et al., 2005; Stevenson et al., 2015; Tye-Murray et al., 2010; 2016). Recent work has also demonstrated that younger adults' perception of a degraded speech signal benefits from manual iconic co-speech gestures in addition to visible speech (Drijvers & Özyürek, 2017; Drijvers et al., 2018). Co-speech gestures are meaningful hand movements which form an integral component of the multimodal language people use in face-to-face settings (e.g. Bavelas & Chovil, 2000; Kendon, 2004; McNeill, 1992). Iconic gestures in particular can be used to indicate the size or shape of an object or to depict specific aspects of an action and thus to communicate relevant semantic information (McNeill, 1992). Whether older adults, too, can benefit from such gestures is currently unknown. The aim of the current study was to find out whether and to what extent older adults are able to make use of iconic co-speech gestures in addition to visible speech during SiN comprehension.

In investigating this question, we also consider whether hearing loss and differences in cognitive abilities play a role in this process. Both factors have been associated with the disproportionate disadvantage older adults experience due to background noises (e.g. Anderson et al., 2013; CHABA, 1988; Humes, 2002, 2007; Humes et al., 1994; Pichora-Fuller et al., 2017; see also Akeroyd, 2008). While age-related hearing loss has direct effects on central auditory processing, it also increases the cognitive resources needed for speech perception (Sommers & Phelps, 2016). Aging is frequently associated with declines in cognitive functioning, e.g. working memory (WM) or inhibitory mechanisms (Hasher, Lustig, & Zacks, 2007; Hasher & Zacks, 1988; Salthouse, 1991). In combination with hearing loss, this may further contribute to an overall decrease in resources available for cognitive operations like language comprehension or recall (e.g. Sommers & Phelps, 2016). Accounting for sensory and cognitive aging is thus crucial in the investigation of older adults' comprehension of SiN and the potential benefit they receive from visual information.

Previous research suggests that perceiving a speaker's articulatory mouth movements can alleviate the disadvantages in SiN comprehension that older adults experience due to sensory and cognitive aging to some extent. The phonological and temporal information provided by visible speech reduces the processing demands of speech and facilitates perception and comprehension (Peelle & Sommers, 2015; Sommers & Phelps, 2016). Accordingly, older and younger adults benefit from visible speech when perceiving SiN, both on a behavioral (e.g. Avivi-Reich et al., 2017; Smayda et al., 2017; Sommers et al., 2005; Stevenson et al., 2015; Tye-Murray et al., 2010; 2016) and on an electrophysiological level (Winneke & Phillips, 2011). The size of the benefit depends on the quality of the acoustic speech signal, or signal-to-noise ratio (SNR), as well as on individual auditory and visual perception and processing abilities (Tye-Murray et al., 2016). Once a certain noise threshold is reached, where individuals can no longer extract meaningful information from the auditory signal, they fail to exhibit any behavioral benefit from visible speech (Ross et al., 2007; Stevenson et al., 2015). As this threshold may be reached earlier in older than in younger adults due to age-related hearing loss, older adults may experience smaller visible speech benefits (e.g. Stevenson et al., 2015; Tye-Murray et al., 2010). Similarly, reduced lip-reading abilities in older adults may also lead to a smaller visible speech benefit (e.g. Sommers et al., 2005, Tye-Murray et al., 2010, 2016).

In addition to visible speech, the semantic information contained in iconic co-speech gestures also enhances speech comprehension and helps in the disambiguation of a lexically ambiguous or degraded speech signal, at least in younger adults. A large body of behavioral and neuroimaging research has shown that under optimal listening conditions, the information conveyed by iconic co-speech is integrated with speech during online language processing (e.g. Holle & Gunter, 2007; Kelly et al., 1999; 2010; Obermeier et al., 2011; for a review see Özyürek, 2014). For speech embedded in multitalker babble noise, word identification is better when sentences are accompanied by an iconic gesture (Holle et al., 2010) and listeners use iconic co-speech gestures to disambiguate lexically ambiguous sentences (Obermeier et al., 2012).

It is important to note that this previous research has investigated the effects of gestures in isolation, by blocking speakers' heads or mouths from view. In every-day language use however, visible speech and co-speech gestures are not isolated phenomena, but naturally co-occur. Therefore, Drijvers and Özyürek (2017) and Drijvers et al. (2018) investigated the joint contribution of both visual articulators on word recognition in younger adults, using different levels of noise-vocoded speech.[10] The combined effect of visible speech and gestures was significantly larger than the effect of either visual articulator individually, at least at a moderate noise vocoding level. At the worst vocoding level, where a phonological coupling of visible speech movements with the auditory signal was no longer possible (see also Ross et al., 2007; Stevenson et al., 2015), gestures provided the only source for a visual benefit.

Considering that iconic gestures provide such valuable semantic information to younger listeners under adverse listening conditions, one might expect their benefit to be comparable or even more pronounced for older adults, since older adults are more severely affected by SiN and have been shown to gain as much or more from additional semantic information (e.g. Pichora-Fuller et al., 1995; Smayda et al., 2017, for effects of sentence context on SiN comprehension).

However, there are indications that older adults may fail to process gestures in addition to speech, and/or to integrate gestures with speech. Cocks et al. (2011) found that older adults were just as good as younger adults in interpreting gestures without speech sound, i.e., visual-only presentation, but had difficulties interpreting co-speech gestures in relation to speech (note that here, the speaker’s face was covered, i.e. no information from visible speech was available). Under highly demanding listening conditions (i.e., very fast speech rates, dichotic shadowing), older adults similarly did not benefit from the semantic information contained in gestures in addition to visible speech, in contrast to younger adults (Thompson, 1995; Thompson & Guzman, 1999). Cocks et al. (2011, p. 34) suggest that it is possible that these findings are due to age-related WM limitations, as “the integration process [of speech and gesture] requires working memory capacity in order to retain and update intermediate results of the interpretation process for speech and gesture.” Older adults’ WM resources may have been consumed with speech processing operations, leaving insufficient resources for gesture comprehension and integration.

Therefore, as the ability to benefit from gestures may depend on an individual’s WM capacity, older adults may benefit less from gestures in addition to visible speech than younger adults, also when perceiving SiN. Furthermore, older adults may focus more strongly on the mouth area as a very reliable source of information, to the potential

[10] Like Drijvers and Özyürek (2017), we use the term “visual articulators” to refer to both the articulatory movements of the mouth and manual co-speech gestures as the media via which visual information is conveyed, as this term is neutral with respect to intentionality.

disadvantage of other sources of visual information (Thompson & Malloy, 2004), such that they might benefit less from gestures in the context of visible speech.

Since the contribution of visible speech and co-speech gestures to older adults' processing of SiN has not been studied in a joint context, it is currently unknown whether older adults can benefit at all from the semantic information contained in co-speech gestures when perceiving SiN, in addition to the benefit derived from visible speech. Similarly, the role that changes in cognitive functioning associated with aging play in the processing of these multiple sources of visual information remains unknown. Given that both visible speech and iconic co-speech gesture form an integral part of human face-to-face communication, these articulators have to be considered jointly in order to gain a comprehensive and ecologically grounded understanding of older adults' comprehension of SiN.

4.1.1. The present study

The primary aim of the present study was therefore to investigate whether aging affects the comprehension of SiN perceived in the presence of visible speech and iconic co-speech gestures, and whether these processes are mediated by differences in sensory and cognitive abilities.

In order to explore this issue, we presented younger and older participants with a word recognition task in three visual contexts: speech-only (mouth blurred), visible speech, and visible speech + gesture. The speech signal was presented without background noise or embedded in two different levels of background multi-speaker babble noise, and participants had to select the written word they heard among a total of four words. These included a phonological as well as a semantic (i.e., gesture-related) distractor and an unrelated answer.

Generally, we expected that both age groups would perform worse at higher noise levels, and that older adults would be affected more strongly than younger adults, potentially mediated by hearing acuity. More importantly, we expected that younger adults' word recognition in noise should improve most when both visual articulators (i.e. mouth movements and gesture) were present, as compared to the benefit from visible speech only, comparable to what has been found for younger adults using noise-vocoded speech (Drijvers & Özyürek, 2017; Drijvers et al., 2018). For the older adults, we refrained from making directed predictions on whether or not they, too, could make use of the semantic information contained in co-speech gesture in addition to visible speech, as the research summarized in the introductory section suggests that either outcome is conceivable (Cocks et al., 2011; Pichora-Fuller et al., 1995; Smayda et al., 2017; Thompson, 1995).

In order to test whether the expected differences between the two age groups in response accuracies and the size of the potential visual benefit is modulated by differences in cognitive abilities, we measured participants' verbal and visual WM and inhibitory control. WM is assumed to be critical for online (language) processing, allowing for the temporary storage and manipulation of perceptual information (Baddeley & Hitch, 1974). Verbal WM capacity predicts comprehension and/or recall of SiN in older adults (Baum & Stevenson, 2017; Koeritzer et al., 2018; Rudner et al., 2016), potentially, because additional WM resources are recruited for the auditory processing of SiN, leaving fewer resources for subsequent language comprehension and recall. Visual WM capacity predicts gesture comprehension in younger adults, presumably playing a role in the ability to conceptually integrate the visuo-spatial information conveyed by gestures with the speech they accompany (Wu & Coulson, 2014). As the ability to process, update and integrate multiple streams of information may likewise depend on sufficient WM resources (Cocks et al., 2011), we expected higher WM capacities to be predictive of better performance overall, as well as a higher benefit of visible speech and gestures.

We additionally included a measure of inhibitory control, as the ability to selectively focus attention or to suppress irrelevant information has been connected to the comprehension of single talker speech presented against the background of several other talkers (i.e., multitalker babble, e.g. Janse, 2012; Jesse & Janse 2012; Tun et al., 2002). Therefore, we also expected better inhibitory control to be predictive of higher performance overall.

Finally, we evaluated the type of errors that participants made in the visible speech + gesture condition, in order to test whether older adults focus more exclusively on the mouth area than younger adults (Thompson & Malloy, 2004). If this were the case, we would expect them to make proportionally fewer gesture-based semantic errors and more visible speech-based phonological errors than younger adults in this condition.

4.2. Method

4.2.1. Participants

30 younger adults (14 women) between 20 and 26 years old (M_{age} = 22.04, *SD* = 1.79) and 28 older adults (14 women) between 60 and 80 years old (M_{age} = 69.36, *SD* = 4.68) took part in the study. The older participants were all community dwelling residents. The

younger participants were students at Nijmegen University or Nijmegen University of Applied Sciences. All participants were recruited from the participant pool of the Max Planck Institute for Psycholinguistics and received between € 8 and € 12 for their participation, depending on the duration of the session. Participants were native Dutch speakers with self-reported normal or corrected-to-normal vision and no known neurological or language-related disorders. Educational level was assessed in terms of highest level of schooling. For the older participants, this ranged from secondary school level (25% of participants) via "technical & vocational training for 16 to 18-year-olds" (50% of participants) to university level (25% of participants). All of the younger participants were enrolled in a university program at the time of testing. The experiment was approved by the Ethics Commission for Behavioral Research from Radboud University Nijmegen. The data of two younger male participants were lost due to technical failure.

4.2.2. Background measures

Hearing acuity

Hearing acuity was assessed with a portable Oscilla© USB-330 audiometer in a sound-attenuated booth. Individual hearing acuity was determined as the participants' pure-tone average (PTA) hearing loss over the frequencies of ½, 1, and 2 kHz and 4 kHz. The data of one older male participant was lost due to technical failure. The average hearing loss in the older group was 24.95 dB (*SD* = 8.04 dB; *Median* = 22.5 dB; *Range* = 13.75 to 37.5 dB) and in the younger group 7.68 dB (*SD* = 3.58 dB; *Median* = 7.5 dB, *Range* = 0 to 15 dB). This difference was significant, Wilcoxon rank sum test, $W = 4$, $p < .001$.

Verbal WM

The backward digit-span task was used as a measure of verbal WM (Wechsler, 1981), which has been used in previous investigations of audiovisual processing and related topics in younger and older adults (e.g., Koch & Janse, 2016; Thompson & Guzman, 1999; Tun & Wingfield; 1999). Unlike word or listening/reading span tasks, the digit span task has the advantage of not being affected by word semantics or frequency (Jones & Macken, 2015). Participants repeated digit sequences of increasing length in reverse order, requiring both item storage and manipulation (Bopp & Verhaeghen, 2005). Scores were computed as the longest correctly recalled sequence. Younger participants scored significantly higher than older participants, M = 5.21 (*SD* = 1.34; *Median* = 5; *Range* = 3 to 8) vs. $M = 4.29$ (*SD* = 1.24; *Median* = 4; *Range* = 0 to 7), $W = 547$, $p = .009$.

Visual WM

The Corsi Block-Tapping Task (CBT, Corsi, 1972) provides a measure of the visuo-sequential component of visual WM. Participants imitated the experimenter in tapping nine black cubes mounted on a black board in sequences of increasing length. Scores were calculated as the length of the last correctly repeated sequence multiplied by the number of correctly repeated sequences. Younger adults performed significantly better than older adults, *M* = 48.71 (*SD* = 19.74; *Median* = 42; *Range* = 30 to 126) vs. *M* = 25.71 (*SD* = 9.28; *Median* = 25; *Range* = 12 to 42), *W* = 721, *p* < .001.

Inhibitory control

Trail Making Test parts A and B (Parkington & Leiter, 1949) were used in order to assess inhibitory control. This test has been used in previous investigations of audiovisual processing in younger and older adults (e.g., Jesse & Janse, 2012; Smayda et al., 2016). In part A, participants connected circled numbers in sequential order. In part B, they alternated between numbers and letters, requiring the continuous shifting of attention. The difference between the times needed to complete both parts (i.e. B-A) provides a measure of inhibition/interference control, as it isolates the switching component of part B from the visual search and speed component of part A (Sanchez-Cubillo et al., 2009). The mean difference between parts B and A was significantly larger for the older adults *M* = 29.54 s (*SD* = 12.88; *Median* = 29; *Range* = 3.7 to 65) than for the younger adults *M* = 16.9 s (*SD* = 8.41; *Median* = 15.65; *Range* = 6 to 47.2), *W* = 142, *p* < .001.

4.2.3. Pretest

We conducted a pretest in order to establish the noise levels at which younger and older adults might benefit most from perceiving gestural information in addition to visible speech (reported in detail in Appendix J). Based on this pretest, we selected SNRs -18 and -24 dB for the main experiment.

4.2.4. Materials

The materials in this experiment were similar to the set of stimuli used in Drijvers & Özyürek (2017) and consisted of 220 videos of an actress uttering a highly frequent Dutch action verb while she was displayed with either having her mouth blurred, visible, or visible and accompanied by a co-speech gesture (see Figure 1, panel A). All verbs were unique and only displayed in one condition. All gestures depicted the action denoted by the verb iconically, e.g. a steering gesture resembling the actress holding a steering wheel for the verb rijden ("to drive"). Gestures were matched on how well they fit with the verb, i.e. their iconicity (see Drijvers & Özyürek, 2017). Each video had a duration of 2 s, with an average speech onset of 680 ms after video onset. Gesture preparation started 120 ms after video onset, and the 'stroke', i.e. the most effortful and meaning-bearing part of the gesture (Kendon, 2004; McNeill, 1992), coincided with the spoken verb.

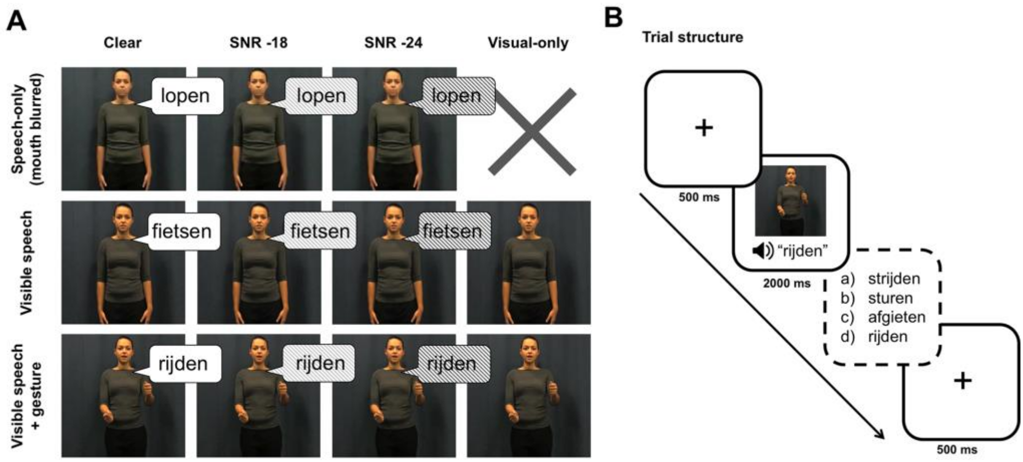


Figure 1. Experimental overview. (A) Overview of conditions. Action words are in Dutch: lopen ("to walk"), fietsen ("to cycle"), rijden ("to drive"). (B) Trial structure. Answer options are in Dutch: strijden ("to fight", phonological competitor), sturen ("to steer", semantic competitor), afgieten ("to drain", unrelated foil), rijden ("to drive", target).

The speech in the videos was either presented as clear speech or embedded in eight-talker babble, with an SNR of -18, or with an SNR of -24. The babble was created by overlaying 20 s fragments of talk of eight speakers (four male and four female) using the software Praat (Boersma & Weenink, 2015). Subsequently, the babble was edited into 2 s fragments and merged with the original sound files using the software Audacity®. The background babble started as soon as the video started and commenced until the video was fully played. The sound of the original videos was intensity scaled to 65 dB. In order to create videos with SNR-18, the original sound file was overlayed with babble at 83 dB, for SNR-24 with babble at 89 dB.

To test for the contribution of gestures in addition to visible speech to the comprehension of SiN, we divided the 220 videos over 11 conditions, with 20 videos per condition (for a schematic representation see Figure 1, panel A). Combining the three visual modalities (speech-only [mouth blurred], visible speech, visible speech + gesture) and three audio conditions (clear speech, SNR -18, SNR -24) yielded nine audiovisual conditions.[11] Two additional conditions without audio were included to test how much information participants could obtain from visual-only information: no-audio + visible mouth movements, which is similar to assessing lip-reading ability, and no-audio + visible mouth movements + gesture, assessing people's ability to grasp the semantic information conveyed by gestures in the presence of visible speech.

We created 28 experimental lists (each list was tested twice, once for a younger and once for an older participant). These lists were created by pseudo-randomizing the order of the 220 videos. Each participant saw each of the 220 videos exactly once in either of the four audio conditions; across the experiment, each video occurred equally often in each audio condition. Per list, the same audio or visual condition could not occur more than five times in a row.

The answer options contained four action verbs: 1) the target verb uttered by the actress; 2) a phonological competitor related to the target verb phonologically; 3) a semantic competitor related to the gesture (if present in the video); and 4) an unrelated foil (see Figure 1, panel B). The semantic competitors were selected on the basis of a pretest (reported in Drijvers & Özyürek, 2017) and consist of action verbs that could plausibly be accompanied by the iconic gesture, i.e., the meaning of the gesture could be mapped to both the target and the competitor. Examples are a "driving" gesture (i.e., moving the hands as if holding a steering wheel) with the target "to drive" (rijden) and the semantic competitor "to steer" (sturen, see Figure 1, panel B), or a "sawing" gesture (i.e., moving hand back and forth as if holding a saw) with the target verb "to saw" (zagen) and the semantic competitor "to cut" (snijden). The four answer options were presented in random order.

Due to a technical error in video presentation, one video had to be removed from the entire dataset, resulting in 219 trials per participant.

4.2.5. Procedure

All participants received a written and verbal introduction to the experiment and gave their signed informed consent. For the main part of the experiment, participants were explicitly instructed to react as accurately and as quickly as possible.

[11] Note that although labelled speech-only (mouth blurred) condition, participants may still glean some information from the speaker's upper face in this condition, which may help identify SiN (Davis & Kim, 2006).

First, hearing acuity was tested as described in section 4.2.2. Subsequently, participants performed the main experiment, seated in a dimly lit sound proof both and supplied with headphones. Videos were presented full screen on a 1650 x 1080 monitor using Presentation software (Neurobehavioral Systems, Inc.) with the participant at approximately 70 cm distance from the monitor. All trials started with a fixation cross of 500 ms, after which the video was played. Then the four answer options were displayed on the screen in writing, numbered a) through d). Participants chose their answer by pushing one of four accordingly numbered buttons on a button box (see Figure 1, panel B for a schematic representation of the trial structure). After every 80 trials, participants could take self-timed breaks. Depending on the participant, this main part of the experiment took approximately 30 to 40 minutes. Afterwards, participants performed the cognitive tests as described above, and filled in a brief self-rating scale to assess their personal attitudes towards gesture production and comprehension (adapted from 'Brief Assessment of Gesture' (BAG) tool, Nagels et al., 2015) as well as a short questionnaire assessing how they made use of the gestures in the current experiment. Older adults agreed significantly less than younger adults with the statement "I like talking to people who gesture a lot while they talk" (W = 584, Bonferroni-adjusted p = .01), but did not significantly differ on any other item. In total, the experimental session lasted between 50 and 75 minutes, depending on the participant.

4.2.6. Statistical methods

We performed three sets of analyses: one for response accuracies, one for the relative benefits of visible speech, of gestures, and of both combined, and one for the proportion of semantic and phonological errors in the visible speech + gesture condition. In line with previous literature on the benefit of visible speech on speech comprehension (e.g., Smayda et al., 2017; Stevenson et al., 2015), we focus our analyses on response accuracies rather than response latencies. However, we report the analyses of the response latencies in Appendix K.

We conducted all analyses in the statistical software R (version 3.3.3, R Development Core Team, 2017), fitting (generalized) linear mixed effects models using the functions glmer and lmer from the package lme4 (Bates et al., 2017).

Analyses were conducted in two steps: first, we evaluated only the experimental predictor variables, their interactions, and the mean-centered pure-tone averages (PTA) as a covariate, applying a backwards model-stripping procedure to arrive at the best-fitting models. We did this by removing interaction terms and predictor variables stepwise based on p-values, using likelihood-ratio tests for model comparisons.

In a second step, we used these best-fitting models as a basis to which we added the mean-centered cognitive variables as covariates in order to test whether additional variation could be explained by differences in cognitive functioning.

All models contained by-participant random intercepts, but no by-item random intercepts, as not all items (i.e., verbs) occurred in all visual modalities. Also, we did not include by-participant random slopes for noise or visual conditions, as this led to convergence failures throughout.

Only the fixed effect estimates, standard errors of the estimates, and estimates of significance of the most parsimonious models are reported. Reported p-values were obtained via the package lmerTest (Kuznetsova et al., 2017). We used the function glht from the package multcomp (Hothorn et al., 2017) in combination with custom-built contrasts to explore individual contrasts where desired, correcting for multiple comparisons.

Response accuracies

We analyzed response accuracies as a binary outcome, scoring 0 for incorrect responses and 1 for correct responses.

Relative benefit

Additionally, we computed each participant's relative benefit scores based on the average response accuracies for each multimodal condition, using the formula (A – B)/(100 – B) (Sumby & Pollack, 1954; Drijvers & Özyürek, 2017). This relative benefit allows for a direct comparison of how much older and younger adults benefitted from the different types of visual information. Additionally, it adjusts for the maximum gain possible and corrects for possible floor effects (see Sumby & Pollack, 1954; see also Ross et al., 2007, for a critical discussion of different benefit scores). The visible speech benefit was thus computed as (visible speech – speech-only)/(100 – speech-only), the gestural benefit was computed as (visible speech+gesture – visible speech)/(100 – visible speech), and the double benefit was computed as (visible speech+gesture – speech-only)/(100 – speech-only).

In fitting the models predicting the relative benefit, we excluded data from "clear" trials, as performance for both age groups was near ceiling and participants often scored at perfect accuracy in the speech-only (mouth blurred) and visible speech conditions, which placed a zero in the denominator of the relative benefit formula.

Proportion of semantic and phonological errors

We computed the proportion of semantic and phonological errors out of all errors made in the visible speech + gestures condition. Rather than using raw error counts or proportion of errors out of all answers, these proportions of errors out of errors account for the possibility that one age group made more errors than the other across the board. Note that we excluded error proportion data for “clear” trials, as performance was frequently at perfect accuracy.

4.3. Results

We first present the analyses of the response accuracies, followed by the analyses of the relative benefit of visible speech, gestures, and both combined, and the analyses of error proportions.

4.3.1. Response accuracies

Figure 2 represents the response accuracies in the audiovisual trials (i.e., with video and sound) and visual-only trials (i.e. with only video, no sound).

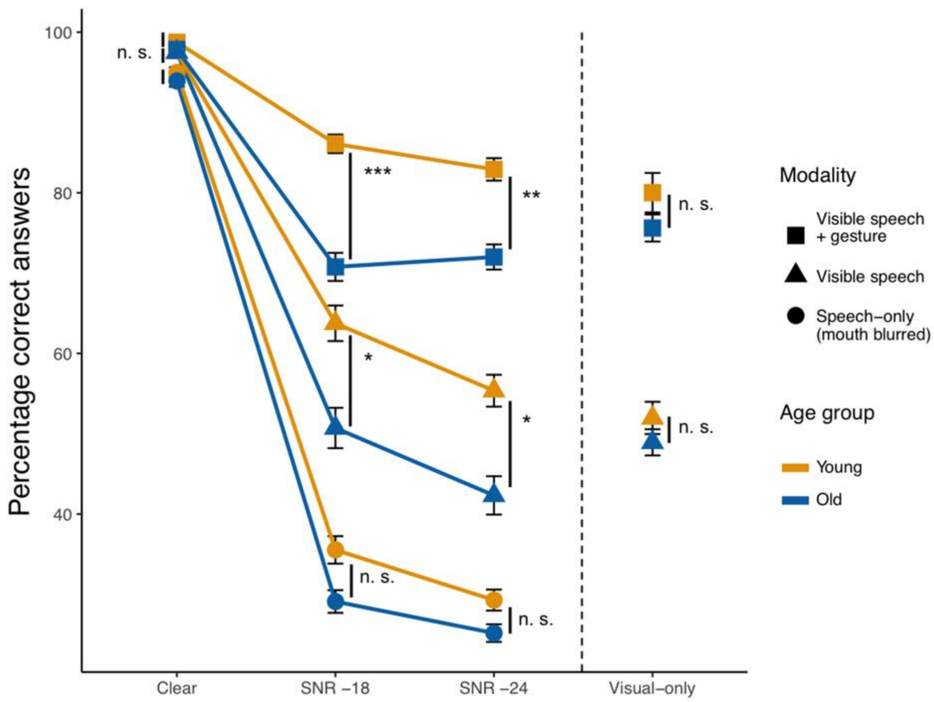


Figure 2. Response accuracy in percent per age group and condition. Error bars represent SE. The dotted line separates the audiovisual trials (left) from the visual-only trials (right).
*** p < .001, ** p < .01, * p < .05

Visual inspection of the data suggested that older adults did not perform better than chance in the speech-only, SNR-24 trials. A Wilcoxon signed rank test confirmed this (V = 97, p = 0.48). Since this concerns only one condition, we decided to conduct our analyses as planned. First, we compared response accuracies in the audiovisual trials based on age group and visual modality. In a second set of analyses, we followed up on the significant interaction of age by visual modality, analyzing audiovisual and visual-only trials separately per visual modality.

Audiovisual trials

Table 1. Model predicting response accuracy in multimodal trials, age group = young and visual modality = visible speech are on the intercept. N = 56.

	Response accuracy			
	β	SE	z	p
Intercept	.97	.07	13.49	< .001
Age group$_{old}$	-.40	.10	-4.07	< .001
Visual modality$_{Speech\text{-}only\ (mouth\ blurred)}$	-.83	.07	-11.32	< .001
Visual modality$_{Visible\ speech\ +\ gesture}$	1.17	.10	12.15	< .001
Age group$_{old}$: Visual modality$_{Speech\text{-}only\ (mouth\ blurred)}$	.25	.10	2.42	.02
Age group$_{old}$: Visual modality$_{Visible\ speech\ +\ gesture}$	-.32	.13	-2.55	.01

An initial model predicting response accuracies in the audiovisual trials based on age group, visual modality, and noise failed to converge. As our main research question and predictions related to the factors age group and visual modality, we decided to include only these two factors in this first part of the analyses, collapsing across noise levels. The younger adults' performance in the visible speech condition was used as a baseline level (intercept), to which we compared the older adults and other visual modality conditions. The best-fitting model (summarized in Table 1) shows significant effects for age and visual modality, such that younger adults outperformed older adults, while more visual articulators lead to higher accuracies. The significant interaction of the two factors indicates that the age-related performance difference was larger in the visible speech condition than in the speech-only condition, and again larger in the visible speech + gesture condition.[12]

[12] An alternative approach to addressing the convergence failure of the full model would have been to exclude the clear speech condition from the analysis, as both age groups performed near

Pairwise comparisons revealed that younger adults' response accuracy was not higher than older adults' in the speech-only (mouth blurred) condition (β = -.16, *SE* = .10, *z* = -1.65, *p* = .45), but it was significantly higher in the visible speech condition (β = -.40, *SE* = .10, *z* = -4.07, *p* < .001) and in the visible speech + gesture condition (β = -.73, *SE* = .12, *z* = -6.04, *p* < .001). Furthermore, both age groups scored significantly higher in the visible speech condition than in the speech-only (mouth blurred) condition (YAs: β = .83, *SE* = .07, *z* = 11.32, *p* < .001; OAs: β = .59, *SE* = .07, *z* = 8.28, *p* < .001). Likewise, both age groups scored higher in visible speech + gesture condition than in the visible speech condition (YAs: β = 1.17, *SE* = .10, *z* = 12.15, *p* < .001; OAs: β = .85, *SE* = .08, *z* = 10.63, *p* < .001).

In summary, although both age groups performed better the more visual articulators were present, the age-related performance difference also increased as more visual information was present. Note that hearing acuity did not improve the model fit.

Cognitive abilities in the audiovisual trials. Including the cognitive abilities yielded a significant effect of verbal WM, such that better WM was associated with higher accuracies (β = .11, *SE* = .04, *z* = 2.74, *p* = .006). The effect size of age group was reduced but remained significant (β = .32, *SE* = .10, *z* = -3.32, *p* < .001). Remaining effects or interactions were not affected.

Audiovisual and visual-only trials

To follow up on the significant interaction of age by visual modality and in order to be able to incorporate noise as a predictor in the analyses, we analyzed the audiovisual and, where applicable, visual-only trials separately per modality. Including the visual-only trials allowed us to investigate possible age differences in these conditions, and to draw direct comparisons between performance in visual-only and audiovisual trials.

---

ceiling in this condition and variation was low. Analyzing this subset of the data yielded significant main effects for age group and visual modality and a significant interaction between age group and visual modality, nearly identical to those reported in the main body of the paper. Additionally, there was a main effect for noise, but no interactions between noise and the other predictors (either 2-way or 3-way).

We nevertheless decided to report the analysis of the full dataset in the body of the paper, because including the clear speech condition is theoretically relevant and necessary in order to exclude the possibility that older adults perform worse than younger adults under optimal listening conditions, particularly in subsequent analyses.

Speech-only (mouth blurred) trials. Within the speech-only (mouth blurred) trials, performance was best predicted by hearing acuity and noise, such that participants with better hearing acuity performed significantly better, while louder noise levels lead to worse performance (see Table 2). There was no significant effect for age group on response accuracy and no interaction with noise, indicating that younger and older adults' performance did not differ significantly at any noise level (note though that the comparison between the two age groups at SNR-24 should be treated cautiously as the older adults' chance level performance in this condition may be masking lower actual performance).

Cognitive abilities in the speech-only (mouth blurred) trials. Verbal WM contributed significantly to the model fit (β = .13, *SE* = .05, z = 2.67, p = .008), reducing the size of the effect of hearing acuity (β = -.12, *SE* = .05, z = -2.42, p = .02).

Visible speech trials. Within the visible speech trials, older adults generally performed worse than younger adults, and both age groups performed worse at louder noise levels. The significant interaction of age group by noise indicates that the age-related performance difference was not equally large at all noise levels (Table 2). Pairwise comparisons revealed that younger and older adults differed from each other in their performance at SNRs -18 (β = -.57, *SE* = .18, z = -3.13, p = .02) and -24 (β = -.56, *SE* = .18, z = -3.10, p = .02), but not in clear speech or in visual-only trials (both p's > .5). Comparing the performance at the individual noise levels for the two age groups separately, we found that younger adults performed significantly better in SNR -18 than in SNR -24 and in visual-only trials (β = -.37, *SE* = .13, z = -2.93, p = .03, and β = -.51, *SE* = .12, z = -4.08, p < .001 respectively). There was no difference between SNR -24 and visual-only trials (p > .1). The older adults performed significantly better in SNR -18 than in SNR -24 (β = -.36, *SE* = .12, z = -2.9, p = .03), but there were no differences between SNR -18 and visual-only trials, or between SNR -24 and visual-only trials (both p's > .5). In summary, both age groups performed equally well in clear speech and visual-only trials, however, when background noise was added to the speech signal, younger adults significantly outperformed older adults. This was not related to differences in hearing acuity. Additionally, only for the younger adults, performance at the less severe noise level was better than in visual-only trials.

Cognitive abilities in the visible speech trials. Including verbal WM and inhibitory control improved the model fit (β = .14, *SE* = .07, z = 1.89, p = .059 and β = .18, *SE* = 08, z = 2.22, p = .03, respectively). This reduced the effect of age (β = -.29, *SE* = .19, z = -1.49, p > .1), but did not affect other effects or interactions.

Table 2. Models predicting response accuracy in speech-only (mouth blurred), visible speech, and visible speech + gesture trials, age group = young and noise = SNR -18 are on the intercept. N = 56[1].

	Speech-only (mouth blurred)				Visible speech				Visible speech + gesture			
	β	SE	z	p	β	SE	z	p	β	SE	z	p
Intercept	-.75	.07	-11.07	< .001	.59	.13	4.60	< .001	1.91	.15	12.40	< .001
Hearing acuity (PTA)	-.15	.05	-3.12	.002	-	-	-	-	-	-	-	-
Age group$_{old}$	-[2]	-	-	-	-.57	.18	-3.13	.002	-.99	.20	-4.93	< .001
Noise$_{clear}$	3.64	.15	24.29	< .001	3.17	.29	11.08	< .001	2.57	.40	6.44	< .001
Noise$_{SNR -24}$	-.24	.09	-2.57	.01	-.37	.13	-2.93	.003	-.25	.17	-1.49	.14
Noise$_{visual-only}$	n.a.[3]	n.a.	n.a.	n.a.	-.51	.13	-4.08	< .001	-.46	.16	-2.78	.006
Age group$_{old}$: Noise$_{clear}$	-	-	-	-	.60	.41	1.48	.14	.41	.50	.82	.41
Age group$_{old}$: Noise$_{SNR -24}$	-	-	-	-	.01	.18	.04	.97	.32	.22	1.47	.14
Age group$_{old}$: Noise$_{visual-only}$	n.a.	n.a.	n.a.	n.a.	.43	.18	2.47	.01	.72	.21	3.33	< .001

[1] In the model predicting response accuracy in the speech-only (mouth blurred) condition, N = 55.

[2] A hyphen indicates a non-significant predictor that was eliminated in the model-comparison process.

[3] Note that there were no visual-only trials in the speech-only (mouth blurred) condition.

Visible speech + gesture trials. Within visible speech + gesture trials, again, younger adults outperformed older adults, and louder noises lead to worse performance overall. As for visible speech, there was a significant interaction age group by noise (see Table 2). Pairwise comparisons revealed that younger and older adults differed from each other in their performance at SNRs -18 ($\beta = -.99$, $SE = .20$, $z = -4.93$, $p < .001$) and -24 ($\beta = -.68$, $SE = .20$, $z = -3.45$, $p = .005$), but not in clear speech or in visual-only trials (both p's > .5). Comparing the performance at the individual noise levels for the two age groups separately, we found that younger adults performed significantly better at SNR -18 than in visual-only trials ($\beta = -.46$, $SE = .16$, $z = -2.78$, $p = .047$), but there was no difference between SNRs -18 and -24 and between SNR -24 and visual-only (both p's > .5). For older adults, there were no significant differences between SNRs -18 and -24, between SNR -18 and visual-only, or between SNR -24 and visual-only (all p's > .5). Thus, as for visible speech, both age groups performed equally well in clear speech and in visual-only trials, but older adults performed significantly worse once background noise was added to the speech signal. Again, this was not related to hearing acuity. Additionally, only the younger adults performed better at the less severe noise level as compared to the visual-only trials.

Cognitive abilities in the visible speech + gesture trials. Including verbal WM significantly improved the model fit ($\beta = .29$, $SE = .07$, $z = 4.12$, $p < .001$). This reduced the effect size of age group without compromising its significant contribution as an explanatory variable ($\beta = -.79$, $SE = .19$, $z = -4.14$, $p < .001$). Other effects or interactions were not affected.

4.3.2. Relative benefit

The relative benefit indicates how much participants' performance improves due to the presence of visible speech compared to speech-only (visible speech benefit), visible speech + gesture compared to visible speech (gestural benefit), or visible speech + gesture compared to speech-only (double benefit). The best-fitting model predicting the influence of age, noise, and benefit type on the size of the relative benefit is summarized in Table 3. The main effect of age shows that overall, older adults received a smaller benefit from visual information than younger adults. There was a significant interaction of benefit type by noise, but no interactions between age group and noise, or between age group and benefit type, suggesting that the pattern of enhancement was comparable for the two age groups (see also Figure 3; note that we might be underestimating the size of the true benefits older adults received at SNR -24 due to their chance performance in the speech-only condition).

We followed the significant interaction between benefit type and noise up by paired comparisons, in order to test whether the size of the individual benefit types changes from one noise level to the next. The visible speech benefit did not change from one noise level to the other ($p > .10$). The gestural benefit increased from SNR -18 to SNR -24; this approached significance ($\beta = .11$, $SE = .04$, $z = 2.67$, $p = .057$). The double benefit (i.e. the benefit of visible speech + gesture compared to speech-only [mouth blurred]) did not significantly change from one noise level to the other (both p's > .1).

Table 3. Model predicting the size of the relative visual benefit, age group = young, benefit type = gestural benefit, and noise = SNR -18 are on the intercept. N = 56.

	Benefit size			
	β	SE	t	p
Intercept	.51	.04	14.28	< .001
Age group$_{old}$	-.14	.03	-4.50	< .001
Benefit type$_{Visible\ speech}$	-.07	.04	-1.53	.13
Benefit type$_{Double}$	.24	.04	5.68	< .001
Noise$_{SNR\ -24}$	.11	.04	2.67	.008
Benefit type$_{Visible\ speech}$: Noise$_{SNR\ -24}$	-.20	.06	-3.23	.001
Benefit type$_{Double}$: Noise$_{SNR\ -24}$	-.11	.06	-1.80	.07

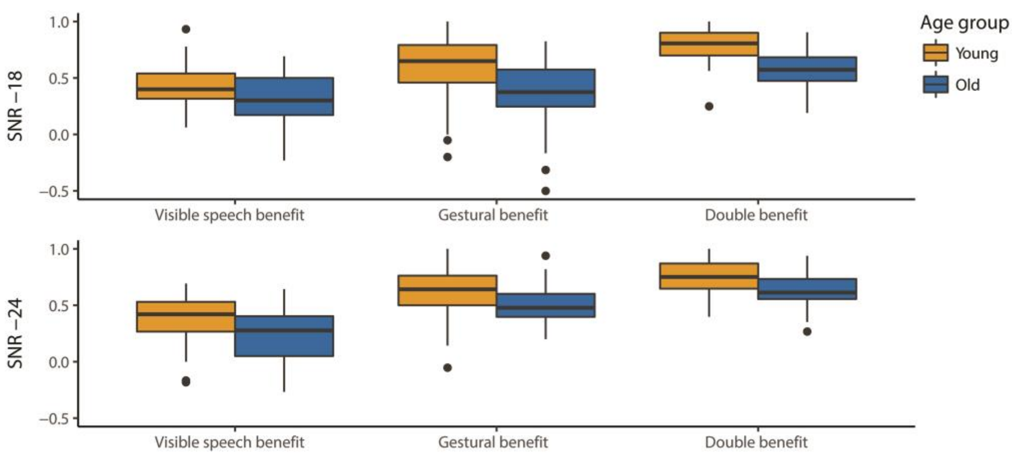


Figure 3. Relative benefit per age group, noise level, and benefit type. The black line represents the median; the two hinges represent the 1st and 3rd quartile; the whiskers capture the largest and smallest observation but extend no further than 1.5 * IQR (data points outside 1.5 * IQR are represented by dots).

Subsequently, we compared the size of the individual benefits per noise level, in order to test whether the benefit of visible speech and gesture combined exceeds that of either articulator individually. At SNR -18, the size of the gestural benefit did not differ significantly from that of the visible speech benefit ($p > .1$). The double benefit was larger than both the gestural benefit ($\beta = .24$, $SE = .04$, $z = 5.68$, $p < .001$) and the visible speech benefit ($\beta = .31$, $SE = .04$, $z = 7.21$, $p < .001$). At SNR -24, the gestural benefit was larger than the benefit of visible speech ($\beta = .26$, $SE = .04$, $z = 6.10$, $p < .001$), and the double benefit was again larger than the gestural benefit ($\beta = .13$, $SE = .04$, $z = 3.13$, $p = .01$) and the visible speech benefit ($\beta = .39$, $SE = .04$, $z = 9.29$, $p < .001$).

Overall then, younger adults benefitted more from visual information than older adults. At the same time, both age groups received a larger benefit from both visual articulators combined than from each articulator individually at both noise levels. Note that neither hearing acuity nor cognitive abilities significantly contributed to the model fit.

4.3.3. Proportion of semantic and phonological errors in visible speech + gesture trials

The best models predicting the proportion of semantic errors and of phonological errors in the visible speech + gesture trials both contained age group as the only significant predictor. Across all noise levels in this visual condition, older adults made a significantly higher proportion of semantic errors than younger adults ($\beta = 10.45$, $SE = 5.03$, $t = 2.08$, $p = .043$) and a significantly lower proportion of phonological errors ($\beta = -9.29$, $SE = 3.95$, $t = -2.35$, $p = .02$). For an overview of all answer types per age group and condition see Appendix L.

4.4. Discussion

The present study provides novel evidence that younger and older adults benefit from visible speech and iconic co-speech gestures to varying degrees when comprehending speech-in-noise (SiN). This variation is partly accounted for by individual differences in verbal WM and inhibitory control, but could not be attributed to age-related differences in hearing acuity. Furthermore, the difference could also not be attributed to differences in the ability to interpret visual information (i.e., how well listeners understood gestures in the absence of speech). The individual results are discussed in more detail below.

Both younger and older adults benefitted from the presence of iconic co-speech gestures in addition to visible speech. For both age groups, response accuracies in the visible speech + gesture condition were higher than in the visible speech condition, and the relative benefit of both visual articulators combined was larger than the relative benefit of either only visible speech or only gestural information. Hence, younger and older adults were able to perceive and interpret the semantic information contained in co-speech gestures and to integrate it with the phonological information contained in visible speech.

Our results are in line with and extend Drijvers and Özyürek's (2017) and Drijvers et al.'s (2018) findings on younger adults' comprehension of a degraded speech signal to multitalker babble noise. At the same time, the present study is the first to show that older adults' speech comprehension under adverse listening conditions, too, can benefit from the presence of iconic gestures. Earlier work on older adults' SiN comprehension had mainly focused on the benefit of visible speech without taking gestures into account (e.g. Sommers et al., 2005; Stevenson et al., 2015; Tye-Murray et al., 2010; 2016). While these studies consistently report a benefit from visual speech, they do not allow for any conclusions with respect to the role of co-speech gestures, which are ubiquitous in everyday talk. We extend this body of work by showing that iconic co-speech gestures can provide an additional benefit on top of the benefit provided by visible speech.

In the light of our findings, it is important to note that work by Thompson (1995) and Thompson and Guzman (1999) suggested that older adults could not benefit from co-speech gestures in addition to visible speech under other highly challenging listening conditions, like speeded speech or dichotic shadowing. We suggest that the difference in findings between these previous studies and the present one is due to differences in task demands. The results of the present study show that in circumstances in which the effort of speech processing is comparatively low (single action verbs rather than sentences, no production component), older adults are able to make use of gestures in addition to visible speech in order to improve their comprehension of SiN. In the communication with older adults then, it might be useful to consider that the benefit from visual cues is potentially enhanced if the linguistic content is simplified or shortened.

Yet, the relative benefit that older adults received from visible speech, gestures, or both articulators combined was significantly smaller than the benefit that younger adults experienced. Although older adults' chance performance in the more severe noise condition might mean that we underestimate their true ability to benefit from visual articulators at this noise level, the effects for the less severe noise level were reliable. Generally, our findings are in line with previous studies reporting a smaller benefit of visible speech for older adults under less favorable listening conditions (Stevenson et al., 2015; Tye-Murray et al., 2010). However, unlike reported in many previous studies on SiN, we did not find significant age-related performance differences in either of the unimodal conditions, i.e. the speech-only (mouth blurred) word recognition, or the visible speech and visible speech + gesture interpretation abilities (visual-only trials). Additionally, differences in hearing acuity did not predict performance in multimodal conditions or the size of the relative visual benefit. Therefore, in the present study, it seems unlikely that the age-related differences in response accuracies and in the relative visual benefit originated in age-related changes in hearing acuity, visual acuity, visual motion detection, or visual speech recognition. Yet, we would like to emphasize that based on our results, we do not make any claims as to whether visual-only speech recognition does or does not decrease in aging. It is possible that our design (using single action verbs, a cued recall task, and a small number of competitors) made the task relatively easier for older adults and therefore overestimates their true lip-reading ability. However, we feel confident to say that the age-related differences in the audiovisual conditions cannot be attributed to differences in visual-only speech recognition as it was assessed here.

Rather, age-related differences in the comprehension of SiN in the visible speech and visible speech + gesture conditions could at least in part be attributed to individual differences in verbal WM. In addition to that, individual differences in inhibitory control also predicted comprehension in the visible speech condition. This is in line with previous research on cognitive factors in SiN comprehension and visible speech (e.g. Baum & Stevenson, 2017; Rudner et al., 2016; Jesse and Janse, 2012; Tun et al., 2002). Our findings thus support the notion that due to the increased processing demands of the speech signal embedded in background talk, added WM and inhibitory resources are required for successful comprehension. Older adults were more strongly affected by the background noise than younger adults, presumably due to their relative decline in WM capacity and inhibitory control.

We therefore suggest that our findings reflect age-related changes in the processing of the auditory and visual streams of information during SiN comprehension. Younger adults used the visual information to enhance auditory comprehension where possible, resulting in higher response accuracies at the less severe noise level as compared to the visual-only trials. When the auditory signal was no longer at least minimally reliable at the more severe noise level, performance did not differ from the visual-only trials. This indicates that in more severe noise, visual information was the only valuable source of information (see also Drijvers & Özyürek, 2017).

For the older adults, on the other hand, performance in the audiovisual trials was not better than in the visual-only conditions. Potentially due to older adults' limited verbal WM resources, which were additionally challenged by the increased processing demands of SiN, it was not possible to simultaneously attend to, comprehend, or integrate all sources of information (see also Cocks et al., 2011). Unlike in previous studies where older adults focused on the auditory signal (Cocks et al., 2011; Thompson, 1995; Thompson & Guzman, 1999), in the present study, they appeared to focus on the visual signal, presumably due to the greater reliability of the visual as opposed to the auditory signal.

Our interpretation is further supported by the trend for older adults to perform worse in audiovisual trials with background noise than in visual-only trials, that we did not observe for the younger adults. Myerson et al. (2016) similarly report cross-modal interference, such that unrelated background babble hinders younger and older adults' ability to lip read (note however that Myerson et al. found no age difference in babble interference, but only in lip reading ability). They suggest that either the monitoring of the speech stream left fewer resources for the processing of visual stimuli, or that the (attempted) integration of visual and auditory speech streams led to interference in the interpretation of the visible speech signal. This suggests that older adults may have spent more WM and inhibitory resources trying to comprehend, integrate, or suppress the background babble, subsequently lacking those resources for visual processing.

Although in principle, it is also conceivable that due to age-related hearing deficits, older adults received insufficient information from the auditory signal at both noise levels, making visual enhancement of the auditory signal impossible, we deem this an unlikely explanation. As we found no significant age-related performance difference in speech-only (mouth blurred) trials, and hearing acuity did not affect response accuracies in multimodal trials, we feel confident to assume that age-related hearing deficits cannot explain why younger adults were able to benefit from visible speech and gesture beyond the simple effect of visual information, but older adults were not.

In addition to age-related differences in hearing acuity, visible speech and gesture interpretation, and cognitive functioning, we also tested the possibility that older adults might pay more attention to visible speech than younger adults (Thompson & Malloy, 2004), to the potential detriment of gesture perception. However, we found that when co-speech gestures were available, older adults made more semantic (i.e. gesture-based) and fewer phonological (i.e. visible speech-based) errors than younger adults. This suggests that older adults actually focused more on gestural semantic information than on articulatory phonological information. In the present task, gestures presented a very reliable signal, and they may have been visually more accessible to older adults than visible speech due to the larger size of the manual as compared to the mouth movements.

Yet, it is important to note that older adults did not focus exclusively on the information contained in gestures, as the benefit of visible speech and gestures combined was larger than the individual benefit of either articulator, also for the older adults. Thus, multimodality enhances communication, despite age-related changes in cognitive abilities.

We are aware that the two noise levels employed in the present study may be considered relatively severe and potentially do not reflect the level of noise accompanying speech in most every-day contexts. The chance performance of older adults at the more severe noise level additionally limited our ability to draw strong conclusions about the true size of their visual benefit in this condition. Yet, the finding that older adults can benefit from visual information even under these conditions is novel and noteworthy in itself. Future research using less severe noise levels may show whether under these conditions, older adults' ability to benefit from visible speech and gestures becomes more comparable to that of younger adults. Furthermore, we could only establish a gestural benefit for single words presented in isolation. Future research employing more complex linguistic material may show whether the beneficial effects of co-speech gestures also extend to longer stretches of speech.

4.5. Conclusion

The present study provides novel insights into how aging affects the benefit from visible speech and from additional co-speech gestures during the comprehension of speech in multitalker babble noise. We demonstrated that when processing single words in SiN, older adults could benefit from seeing iconic gestures in addition to visible speech, albeit to a lesser extent than younger adults. Age-related performance differences were absent in unimodal conditions (speech-only or visual-only) and only emerged in multimodal conditions. Potentially, age-related working memory limitations prevented older adults from perceiving, processing, or integrating the multiple sources of information in the same way as younger adults did, thus leading to a smaller visual benefit. Yet, our findings highlight the importance of exploiting the full multimodal repertoire of language in the communication with older adults, who are often faced with speech comprehension difficulties, be it due to age-related hearing loss, cognitive changes, or background noise.

Acknowledgments

We would like to thank Nick Wood for his help with video editing, Renske Schilte for her assistance with testing participants, and Dr. Susanne Brouwer for statistical advice. We would also like to thank two anonymous reviewers for their helpful comments and suggestions on an earlier version of this manuscript.

Chapter 5

General Discussion and Conclusion

In this thesis, I have studied the effects of aging and age-associated changes in cognitive functioning on multimodal utterances in language production and comprehension. My aim was to gain a better understanding of the multimodal communicative competences of a growing aging population in different communicative contexts. As laid out in the introductory chapter, previous research on older adults' use of the spoken and gestural modalities suggested systematic and significant age-related differences in language production and comprehension, such as differences in the ability to adapt verbal utterances to common ground shared with an addressee, differences in the ability to understand speech in noisy surroundings, differences in the frequency of representational co-speech gesture production, and differences in the ability to integrate information conveyed by co-speech gesture with information conveyed by speech in order to improve comprehension. However, a number of questions remained unanswered and were the focus of the present thesis: Are older adults' co-speech gesture production and comprehension affected by the context in which language is produced and perceived and if so, how? More specifically, how does the presence or absence of common ground with an addressee, or even the mere presence of a genuine addressee affect older adults' multimodal language production? How does background noise affect the ability to benefit from co-speech gestures during language comprehension? Finally, how do age-related changes in cognitive functioning affect the production and the comprehension of multimodal utterances?

Apart from placing a focus on co-speech gestures and looking beyond spoken language production and comprehension, the novelty of the approach used in this thesis was that it brings together contextual and cognitive factors in trying to understand age-related changes in multimodal language use. In addition, it also shows that aging provides a valuable testing ground to understand factors underlying multimodal language use more generally.

In the remainder of this chapter, I will first summarize the core findings of this thesis. I will then discuss these findings with respect to my initial hypotheses and consider how they broaden our understanding of older adults' communicative competences, comprising speech and co-speech gesture production and comprehension, more generally. I will also relate the findings to existing accounts of co-speech gesture production and comprehension. This chapter (and this thesis) end with some concluding remarks and suggestions for future research.

5.1. Summary of core findings

In Chapters 2 and 3 of this thesis, I investigated how older adults use co-speech gestures in interaction with others. Specifically, I investigated whether aging and cognitive factors modulate the adaptation of speech and co-speech gestures based on knowledge shared with an addressee, or common ground (Clark, 1996). I found no overall age-related differences in co-speech gesture frequency and rate. However, how older adults adapted their speech and co-speech gestures according to an addressee's knowledge state – and whether they behaved like younger adults or not – depended on the type of communicative task they had to perform.

In the narrative (or story telling) task reported in Chapter 2, only the younger, but not the older participants, provided longer and more informative narrations and gestured at a higher rate when relating unknown as opposed to mutually shared story content. Older adults showed no evidence of common ground-based recipient design either in their speech or in their gestures, and even produced fewer rather than more gestures in relation to speech when relating novel story content in this task. At the same time, both younger and older adults' gesture production was sensitive to addressee feedback. Age-related differences in cognitive abilities (verbal and visual working memory [WM], inhibitory control, semantic fluency) did not predict the differences in common ground-based recipient design. This suggests that other factors, such as differences in communicative goals, may have driven the older adults' communicative behavior.

This interpretation is supported by the findings of the second, spatial task reported in Chapter 3. Contrary to my expectations and unlike in the narrative task, I found no age-related differences in verbal and gestural adaptations to common ground. Rather, the two modalities were clearly affected by personal common ground induced per trial and by incremental common ground accumulating across trials for participants of both age groups. Rather than by age as such, the extent of common ground-based recipient design in this second task was modulated by individual differences in cognitive abilities, in particular verbal and visual WM and semantic fluency. Differences in inhibitory control did not affect the behavioral measures. From these findings, I concluded that under certain conditions, older adults have the capacity to engage in multimodal common ground-based recipient design: Older (and younger) adults' multimodal language use is determined by an interplay of cognitive factors, such as task demands and individual cognitive abilities, and contextual factors, such as the communicative setting and the speaker's communicative intention.

Whereas the two studies presented in the first part of this thesis investigated the role of aging and cognitive abilities on communicative co-speech gesture production, the study reported in Chapter 4 of this thesis addressed the impact of these two factors on speech and co-speech gesture comprehension. In particular, I tested whether older adults' speech comprehension could benefit from seeing iconic gestures in addition to visible speech under adverse listening conditions. I found that both age groups benefitted from gestures in addition to visible speech when perceiving speech-in-noise (SiN), i.e. they were better at understanding SiN when gestures were available as compared to when they were not. However, older adults benefitted significantly less than younger adults. This was, at least in part, predicted by individual differences in verbal WM. The ability to comprehend visible speech in noise (without gestures) was furthermore affected by inhibitory control. Differences in visual WM, however, did not affect the comprehension of multimodal utterances. From this, I concluded that it may be cognitively costly to perceive and process speech and gestures at the same time, at least under adverse listening conditions: Even though gestures may help comprehension, this does not come for free. Sufficient sensory and cognitive abilities are fundamental for speech and gesture perception, processing, and integration.

5.2. General discussion of findings and theoretical implications

As the summary of findings presented above shows, the major contribution of the present work is that it provides novel insights into the multimodal communicative behavior of a growing aging population and the factors that guide it. Unlike what previous research had suggested, co-speech gestures continue to play an important communicative role in older adults' language use and do not seem to decrease with age. By looking beyond the spoken modality, and by employing varying communicative contexts that entailed different communicative demands and challenges, I could show that older adults' behavior is guided by an interplay of cognitive and contextual factors. This has implications for general accounts of language use in older adults: language production and comprehension remain multimodal, as well as strongly context-dependent, and this needs to be considered in future investigations of older adults' communicative behavior, their communicative competences and practices.

In addition, the present thesis also contributes to existing knowledge about gesture production and comprehension more generally. By explicitly testing the influence of aging, cognitive abilities, and communicative context, I could show that all of these factors are at work to shape gesture production and comprehension. The novel contribution to the field is thus to further bridge the gap of communicative gesture research on the one hand, and cognitive gesture research on the other.

5.2.1. Age-related changes in multimodal language production

The main research questions that guided the studies on multimodal language production (Chapters 2 and 3) were: Do previous findings on age-related differences in gesture production generalize to a more communicative context? How is older adults' co-speech gesture production affected by the communicative needs of an addressee? How do age-related cognitive changes affect communicative gesture production?

I hypothesized that there would be age-related differences in common ground-based recipient design, such that older adults would show less evidence of speech and gestural adaptations than younger adults, potentially modulated by cognitive factors. Additionally, I considered two alternative outcomes that cognitive aging might have on co-speech gesture production itself: Older adults might rely relatively more on gestures in order to compensate for potential age-related cognitive limitations, as gestures are assumed to provide the speaker with a cognitive benefit (see introductory chapter, section 1.2.2.). Alternatively, I hypothesized that older adults may rely relatively less on gestures and more on speech in interaction with others, due to cognitive limitations which impede the production of communicative gestures. As summarized in section 5.1. above, I found that older adults' multimodal language use was determined by an interplay of contextual factors, such as the communicative context and the speaker's communicative intention, and cognitive factors, such as task demands and individual cognitive abilities. I will discuss the relevance of each of the two factors for our findings in turn.

The role of contextual factors in multimodal language production

Effects of common ground. As expected, the communicative context in which speech and gestures are produced has a major impact on younger and older adults' communicative behavior. First, let us consider the effects of the specific requirements of the communicative situation, in particular mutually shared knowledge between speaker and addressee or common ground. As summarized in section 5.1., in the narrative task, older adults showed no evidence of common ground-based recipient design in either speech or gesture. However, in the spatial task, they adapted both modalities to the same extent as younger adults. Only the findings for the narrative task were thus in line with earlier findings that older adults are less efficient in common ground-based adaptations than younger adults in the spoken modality (Horton & Spieler, 2007; Hupet, Chartraine, & Nef, 1993; Lysander & Horton, 2012). Importantly, I could show that this lack of recipient design extends to the gestural modality, too. Since there were no significant effects of cognitive factors on common ground-based adaptations of speech and gesture in this task, I speculated that age-related differences in task goals may have contributed to older adults' apparent lack of recipient design. While the younger adults presumably focused mainly on information transfer, i.e. providing the addressee with information she did not yet have, older adults may have interpreted it as a task where it mattered to be a 'good story teller', in the sense of providing an easy-to-follow narrative and being clear and exhaustive in terms of story events.

This interpretation seems to be supported by the unexpected findings from the spatial task. Since the same group of participants took part in both experiments, it shows that this group of older adults clearly had the capacity and the motivation to engage in common ground-based recipient design in speech and also in gestures. Consequently, there must be factors other than, or in addition to, age-related cognitive differences or cognitive factors more generally (discussed below), that determine whether older adults adapt their multimodal language use according to an addressee's knowledge state or not. Particularly interesting in this context is that the task goal of the spatial task may have been more straight-forward than that of the narrative task, i.e., to provide the addressee with sufficient information in order to enable her/him to assemble the wooden structure accurately. Distinguishing between known and novel information may have been more relevant to the older adults in the spatial than in the narrative task (for the role of the relevance that speakers attribute to the known/novel distinction for recipient design see also Galati & Brennan, 2014). Previous research similarly suggests that communicative intent and perceived relevance of a given task are important determinants of older adults' addressee-based adaptations (e.g., Adams, Smith, Pasupathi, & Vitolo, 2002). These findings highlight the importance of adopting different settings when investigating older adults' interactive language use, as the findings from one task may not generalize to another task (or to real life interactions, for that matter).

Effects of communicative setting. Next, let us turn to the effects that the communicative context in general, and the presence of a genuine, naïve addressee in particular, may have had on older adults' multimodal language production. There were no overall age-related differences in terms of spoken utterance length or information content in both the narrative and the spatial task. Older adults produced the same amount of words and semantic detail as younger adults. Furthermore, in both tasks, older adults also gestured as frequently as younger adults and expressed the same amount of information in their gestures. This clearly shows that healthy older adults use co-speech gestures alongside speech in their communication with others, for example to illustrate certain aspects of the story events they were talking about, or to convey relevant spatial information. Thus, older adults can exploit gestures as a communicative strategy and appear to be aware of their communicative value. This is not in line with earlier findings of an overall age-related difference in co-speech gesture production in visuo-spatial and other tasks (Arslan & Göksun, 2020; Cohen & Borsoi, 1996; Feyereisen & Havard, 1999; Theocharopoulou, Cocks, Pring, & Dipper, 2015). Where this previous research had suggested that older adults use fewer co-speech gestures and focus on the spoken message instead, the present results show that at least in more socially situated settings, older adults use co-speech gestures like younger adults do.

I would like to propose that this absence of age-related differences is primarily due to the communicative context employed here, in which older adults communicated with a genuine addressee rather than a video camera or an experimenter (see also Arslan & Göksun, 2020). Previous research with younger adults indicates that the presence of a visible, attentive addressee increases the relative frequency of representational gestures (e.g. Bavelas, Gerwing, Sutton, & Provost, 2008; Jacobs & Garnham, 2007; Kuhlen, Galati, & Brennan, 2012). Hence, older adults, too, might be more motivated to use gestures under such conditions, or they might put more effort into producing them (and potentially invest more cognitive resources, but see also the next sections for a more detailed discussion of cognitive effects). Thus, when trying to assess older adults' communicative abilities and their multimodal communication in particular, the communicative context in which language is produced is an important aspect to take into consideration.

The role of cognitive factors in multimodal language production

General effects of cognitive factors on gesture production. In addition to the contextual factors described above, cognitive factors also affected the use of speech and co speech gestures. First, however, let us consider the absence of a general effect of cognitive factors or (cognitive) aging on older adults' multimodal language production. For one thing, we found no evidence that aging or cognitive factors influenced gesture production as such (cf. Hostetter & Alibali, 2018; Kita, Alibali, & Chu, 2017). That is, in spite of measurable cognitive differences, older adults gestured neither more (so as to compensate for limited abilities) nor less (due to limited abilities, e.g., Arslan & Göksun,

2020) than younger adults (with the exception of the common ground condition in the narrative task). Similarly, there was no evidence for an age-related difference in the use of mental imagery during language production (Cohen & Borsoi, 1996), or an age-related shift in the speech-gesture system that leads to a stronger reliance on the spoken modality (c.f., Feyereisen & Havard, 1999; Theocharopoulou et al., 2015). However, it should also be pointed out that this absence of an overall effect of (cognitive) aging on gesture production went hand-in-hand with an overall absence of age-related differences in speech as measured by word count and semantic content. Hence, the present results might also indicate that in the tasks used here, older adults' cognitive resources were not so seriously taxed as to warrant either an increase or a decrease in gesticulation. It is possible that in tasks which pose more difficulty in terms of conceptualization, we would see a stronger influence of cognitive factors. Additionally, it is also possible that due to the addressee-oriented nature of the tasks used here, most gestures were produced with the intention to convey relevant meaning to the addressee, and that this masked the more cognitive, speaker-oriented potential of gesture production (see also Galati & Brennan, 2014). Finally, also due to the relatively small sample size employed in the studies, the absence of cognitive effects on gesture production need to be interpreted with caution. Although at this point, the present findings do not support the idea of a direct relationship between individual differences in cognitive abilities and gesture production, I do not want to suggest that this relationship does not exist.

Effects of individual differences in cognitive abilities on recipient design. In spite of the absence of a general effect of cognitive factors on gesture production, I did find direct effects of individual differences in cognitive abilities on the communicative adaptation of speech and co-speech gestures. As stated above, in the narrative task, there were no effects of cognitive abilities on speech and gesture production, in spite of significant age-related differences in multimodal recipient design. In the spatial task, however, individual differences in verbal and visual WM as well as semantic fluency influenced the extent of common ground-based adaptions in speech and also in gesture. Higher cognitive abilities allowed for more pronounced adaptations in both modalities. This suggests that the communicative adaptation of speech and of co-speech gestures may be cognitively costly (see also Horton & Gerrig, 2005; Mol, Krahmer, Maes, & Swerts, 2009) and presents a valuable contribution to previous research on interactive language use. Interestingly, while higher WM abilities led to more pronounced adaptations in speech and co-speech gesture across both age groups, the lower WM capacity of older relative to younger people did not predict their behavior on an age group level. Thus, there appears to be no simple, one-to-one relationship between communicative speech and gesture production and aging or age-related cognitive changes. Note that in neither of the two tasks we found an effect of individual differences in inhibitory control on verbal or gestural recipient design. However, as also stated in the discussion in chapter 2, it is possible that the particular task that we used to assess inhibitory control did not tap into the actual processes involved in multimodal recipient design as investigated here, or that inhibitory control plays a more prominent role in recipient design based on visual common ground (as e.g. in Long, Horton, Rohde, & Sorace, 2018, or Wardlow, 2013) as opposed to recipient design based on conversational common ground, as used in the present studies.

Effects of task demands on recipient design. Finally, let us consider the way in which the different tasks may have affected cognitive demands and thereby participants' communicative behavior. Importantly, the second, spatial task differed from the first, narrative task in two main ways that may have affected the associated cognitive load. The first and major difference was the spatial nature of the second task. Presumably, this task relied more strongly on visuo-spatial WM, including motor memory, than the narrative task. The sensory-motor experience in particular may have decreased the memory demands that older adults were faced with (e.g., Engelkamp, 1998). Also, the vocabulary to be used during the spatial descriptions was more restricted than in the narrative task, consisting mainly of geometric shape and size attributes and spatial prepositions, which may have decreased the demands of language/utterance planning. Finally, co-speech gestures played a much more prominent role during the spatial descriptions than during the narrations. This potentially gave speakers the opportunity to "off-load" some of the information onto visual space, thereby freeing up resources which then became available for other cognitive operations (see Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Wagner Cook, Yip, & Goldin-Meadow, 2012).

The second difference, relating more to the specific task design, was that the distinction between known and novel information was simpler and perhaps more salient in the spatial task than in the narrative task and may therefore also have been easier to encode and remember. Recall that in the spatial task, the addressee either saw the full model or not, whereas in the narrative task, the addressee always knew part of the story and the speaker had to keep track of which information was shared and which was not. Taken together with the presumed differences in task goals (see above), these task-related differences may have contributed to making the spatial task less cognitively demanding overall than the narrative task. This, in turn, may have affected older adults' ability to engage in verbal and gestural recipient design, assuming that such communicative behavior is dependent on cognitive resources (see previous section). Yet, I want to emphasize again that ultimately it is the interplay of the various contextual and cognitive factors discussed above, which determines older and younger adults' multimodal language use in interaction with others.

Implications for theoretical accounts of co-speech gesture production

My results have some implications for current accounts of co-speech gesture production. As discussed above, I demonstrated that the production of speech and crucially also of co-speech gestures is modulated by the specific requirements of the communicative context: in terms of the addressee's knowledge status, the speaker's communicative goal, and the type of information that needs to be communicated. Additionally, speakers' multimodal language production was sensitive to the verbal and non-verbal feedback that addressees gave. Our findings are thus in line with and support accounts of gestures as being communicatively intended and tightly coordinated with speech, both at the level of message conceptualization and at the level of utterance planning (e.g., Hostetter & Alibali, 2018; Kita & Özyürek, 2003; McNeill, 1992; 2005). Crucially, this was also true for older adults, such that like younger adults, older adults used gestures communicatively, and like younger adults, they coordinate speech and gesture use very carefully so as to fulfill communicative requirements. However, where previous accounts of co-speech gesture production have mainly focused either on the communicative aspects of gesture production (How are gesture and speech organized? How do we proceed from an initial communicative intention to a multimodal utterance? – de Ruiter, 2000; Kita & Özyürek, 2003), or on the cognitive aspects of gesture production (How do gestures facilitate speech production? What are the mechanisms by which this facilitation is achieved? Which cognitive mechanisms give rise to gestures in the first place? – Hostetter & Alibali, 2008; 2018; Kita et al., 2017; Krauss, Chen, & Gottesman, 2000), the unique contribution of my thesis for the literature is that in addition to being affected by the speaker's communicative intention and the addressee's communicative needs, how gestures are used for communication is modulated by the speaker's cognitive abilities. The present findings illustrate that the practice of investigating either the communicative or the cognitive functions of gestures can necessarily only yield an incomplete picture of gesture use in social interaction. Although previous research (e.g., Galati & Brennan, 2014; Hoetjes, Koolen, Goudbeek, Krahmer, & Swerts, 2015; Masson-Carro, Goudbeek, & Krahmer, 2016) had similarly addressed the simultaneous influence of communicative and cognitive demands on co-speech gesture production by varying the external cognitive load that was placed on participants in communicative settings, the present studies go beyond this work by explicitly testing how individual differences in cognitive abilities, i.e. internal cognitive load, affect communicative behavior, as well as by connecting this to cognitive change as part of the aging process.

For models and theories that emphasize the communicative, addressee-oriented aspect of gesture production (such as e.g. the Interface Hypothesis, Kita & Özyürek, 2003), this means that the role of cognitive abilities may need to be acknowledged at least in some detail, such that e.g. verbal WM resources are required for overall utterance planning, the coordination and execution of the multimodal message, as well as constant updating of local and global discourse aspects related to personal or incremental common ground. Visual WM may be needed for gesture planning and execution as well as the coordination of the gestural with the spoken message content. Clearly, more research is needed in order to establish which abilities support which processes in communicative speech and co-speech gesture production. Yet, even at this early stage, the present results suggest that multimodal language use requires cognitive resources and that taking individuals' cognitive abilities into account can improve our understanding of their communicative behavior.

For models and theories that emphasize the cognitive, speaker-oriented aspect of gesture production (e.g., Kita et al., 2017; Krauss et al., 2000), I recognize that their focus may be on different aspects or instances of language use. Still, the present findings suggest that even though gestures may have functions that facilitate the speaking process and may thus potentially compensate for cognitive limitations, in other contexts, sufficient cognitive resources may be a prerequisite in order to produce communicative gestures appropriately. That is, while theories like the Gesture as simulated action (GSA) framework (Hostetter & Alibali, 2008; 2018) or the Gesture-for-Conceptualization Hypothesis (Kita et al., 2017) assume that higher extrinsic or intrinsic cognitive load lead to an increase in gesticulation, they should also to be able to account for situations in which this is not the case and higher cognitive load has either no effects on gesture production, or even reverse effects. For example, while gesturing in principle may help structuring complex spatio-motoric information for utterance production, leading to an increase in gesture frequency in some settings, in other settings, the presence of an addressee and the associated communicative pressures or motivation may lead to a speaker's choice to distribute information differently across the two modalities, e.g. by putting more information into the verbal modality, in order to achieve her/his communicative goals. Additionally, sufficient cognitive resources may be a necessary prerequisite in order to be able to use gestures to manage high cognitive load in the first place (see also Özer & Göksun, 2020). Ideally, future models of speech – co-speech gesture production should acknowledge the role of both the speaker's communicative intent and speaker-internal (and -external) cognitive factors for gesture generation and execution, resulting in an integrative account of multimodal language production.

5.2.2. Age-related changes in multimodal language comprehension

Let us now turn to the age-related changes in multimodal language comprehension. The main research questions that guided the study reported in Chapter 4 were: How is older adults' ability to comprehend and benefit from co-speech gestures affected when the speech signal is embedded in noise? Are older adults able to integrate the semantic information conveyed by gestures with the phonological information conveyed by visible speech to maximally enhance their speech comprehension? How do age-related changes in cognitive functioning affect the comprehension of communicative co-speech gestures?

I hypothesized that (cognitive) aging could affect gesture comprehension and the ability to benefit from iconic co-speech gestures in either of two distinct ways. Older adults might rely relatively more on gestures than younger adults in order to compensate for age-related sensory and/or cognitive decline, hence older adults might receive a larger benefit from the additional visual information relative to younger adults. Alternatively, due to age-related cognitive limitations, which may affect the ability to perceive, process and/or integrate co-speech gestures, older adults may rely relatively less on co-speech gestures during language comprehension, and therefore receive a smaller benefit relative to younger adults. As summarized in section 5.1. above, I found that older adults indeed receive a smaller benefit from co-speech gestures than younger adults, which is partly attributable to cognitive differences. Yet, I also found that older adults strongly relied on co-speech gestures as a valuable source of information. In the following, I will discuss the influence of contextual factors, such as the presence of background noise and of visible speech, and of cognitive factors, such as individual differences in cognitive abilities, on our findings.

The role of contextual factors in multimodal language comprehension

The context in which speech and gestures are perceived greatly influences younger and older adults' ability to receive a communicative benefit from co-speech gestures. As expected, older adults' speech comprehension was more strongly affected than younger adults' when speech was embedded in background noise, even when visible speech was available (e.g., Stevenson, Nelms, Baum, Zurkovsky, Barense, Newhouse, & Wallace, 2015; Tye-Murray, Sommers, Spehar, Myerson, & Hale, 2010). Yet, under these adverse listening conditions, older adults were aware of and exploited the communicative potential of co-speech gestures. They focused on the information expressed in co-speech gestures and used this information to arrive at an interpretation of the speaker's utterance. Importantly, they were also able to integrate the gestural, semantic information with the phonological information they derived from the articulatory lip movements, suggesting that they were able to make use of these two distinct visual communicative signals in order to obtain a larger benefit.

Earlier research had suggested that older adults do not integrate gestures with speech, even under ideal listening conditions (Cocks, Morgan, & Kita, 2011). Furthermore, older adults could not benefit from co-speech gestures in addition to visible speech under highly challenging listening conditions, like speeded speech or dichotic shadowing, focusing on the auditory signal instead (Thompson, 1995; Thompson & Guzman, 1999). The differences in findings between those previous and the present study may in part be attributable to the context in which speech and gestures were perceived. For example, in Cocks et al. (2011), the speech signal was always clear. Since it appeared that they could glean all relevant information from the auditory signal, older adults may have been less motivated to rely on the gestural signal. Relatedly, as suggested in the introduction, adverse listening conditions may boost the reliance on gestures, also in younger adults (Obermeier, Dolk, & Gunter, 2012). In view of the above, it is also not surprising that older adults relied more strongly on speech in Cocks et al. (2011), and more strongly on gestures in the present study: while speech was always reliable in the previous study, gestures may have been the more reliable source of information in the present study. Finally, in Cocks et al. (2011), visible speech was not available as the speaker's face was covered. Therefore, it is unclear whether older adults really perceived speech and co-speech gestures as one integrated message, and this may have also affected their gesture processing and integration (see e.g. Kelly, Ward, Creigh, & Bartolotti, 2007).

In the case of Thompson's (1995) and Thompson and Guzman's (1999) studies, it is possible that the context in which co-speech gestures and visible speech were presented was too challenging for older adults. When the effort of speech processing is comparatively low, as in the present study (single action verbs rather than sentences, no production component), older adults are able to make use of gestures in addition to visible speech in order to improve their comprehension of SiN. However, when the processing effort is high, as in speeded speech (Thompson, 1995) or dichotic shadowing (Thompson & Guzman, 1999), this may become a context in which older adults indeed experience “overload” (see also next section). In summary, my results suggest that whether older adults can and do integrate and comprehend co-speech gestures along with the accompanying speech is highly context-dependent, in fact more so than for younger adults, potentially due to age-related changes in cognitive capacities which force older adults to distribute their (more limited) capacities differently.

The role of cognitive factors in multimodal language comprehension

As already hinted at above, not only contextual, but also cognitive factors clearly affected older adults’ ability to benefit from co-speech gestures in addition to visible speech.[13] As shortly summarized above, older adults’ greater difficulties at understanding SiN and their smaller benefit from visual information relative to younger adults were partly attributable to differences in verbal WM. Differences in inhibitory control additionally had an effect on the comprehension of visible speech (without gestures). In this sense, the present findings are in line with earlier proposals that due to limited cognitive capacities, older adults have more difficulties with processing and comprehending SiN (Sommers & Phelps, 2016) and crucially also with the processing and/or integration of information conveyed in the gestural modality (Cocks et al., 2011; Thompson; 1995; Thompson and Guzman, 1999). It appears that even though gestures can provide very valuable visual cues (in addition to visible speech) that serve to improve speech comprehension, they also present an additional signal that needs to be perceived, processed, and integrated with speech. Sufficient verbal WM capacity and potentially other cognitive resources are necessary in order to simultaneously attend to and perceive the individual signals, to process them, and to integrate them into one comprehensive message representation (see Özer & Göksun, 2020, for a similar argument). Hence, perceivers with limited verbal WM capacity, due to aging or otherwise, may be less efficient at co-speech gesture interpretation and integration (see also Wu & Coulson’s 2014 *verbal resources hypothesis*).

Although a number of questions remain unanswered with respect to the exact mechanisms underlying the age-related differences we observed in co-speech gesture comprehension (do older adults have difficulties perceiving and processing the auditory and visual signals simultaneously? Or do the difficulties arise at the level of speech-

[13] Note that the contextual factors mentioned in the previous section obviously also have implications for cognitive processes.

gesture integration?), our results still suggest that focusing on co-speech gestures can present a useful strategy for older adults when speech comprehension is difficult. Our findings furthermore highlight the importance of exploiting the full multimodal repertoire of language in the communication with older adults.

Implications for theoretical accounts of co-speech gesture comprehension

My results have some implications for current accounts of co-speech gesture comprehension. Previously, it has been claimed that perceiving gestures aids speech comprehension, e.g. by providing additional visual information that can disambiguate a degraded speech signal (e.g., Drijvers & Özyürek, 2017; Obermeier et al., 2012). Our findings are very much in line with this view: for younger and for older adults, gestures add to the comprehension of the speech signal in non-trivial ways when this signal is degraded. Importantly, this benefit of co-speech gestures was additive to the benefit derived from visible speech, showing that during the comprehension of SiN, multiple sources of visual information are exploited to obtain an accurate understanding of the speaker's message. The novel contribution of the present research was to show that in addition to other modulating factors identified previously, such as the perceived intentionality underlying the coupling of speech and gesture (Kelly et al., 2007), the temporal synchrony of speech – gesture onset (Habets, Kita, Shao, Özyürek, & Hagoort, 2011), the presence of background noise (Obermeier et al., 2012), or addressee status (Holler, Kokal, Toni, Hagoort, Kelly, & Özyürek, 2015; Holler, Schubotz, Kelly, Hagoort, Schuetze, & Özyürek, 2014), the comprehension of communicative co-speech gestures is additionally constrained by aging and verbal WM: Even though older adults did benefit from gestural information, they did not benefit as much as younger adults, which could partly be attributed to individual differences in verbal WM. This supports the idea that speech – co-speech gesture perception, processing, integration, and comprehension is indeed dependent on cognitive resources. Verbal WM capacity presumably is relevant for processing, storing and updating verbal information and integrating it with visual information, held in visual WM (see also Coulson & Wu, 2019; Wu & Coulson, 2014). At the same time, contextual factors may modulate the involvement of the individual abilities, such that, for example, the presence of visible speech may make the processing of speech easier due to the additional articulatory information that is provided, thereby "freeing up" more resources for gesture processing and integration. Certainly, more research is needed to establish which cognitive abilities support which processes during speech and co-speech gesture comprehension, under which circumstances, and how the different visual signals (such as articulatory lip movements and manual gestures) interact during these processes. Yet, even at this early stage, the present results suggest that current accounts of co-speech gestures comprehension need to be extended to cover the impact of speaker-internal constraints in interaction with contextual factors in order to describe the comprehension of co-speech gestures more accurately.

5.2.3. Age-related changes in spoken language production and comprehension

Before moving on to the overall conclusion, let us shortly turn to the issue of age-related changes in spoken language production and comprehension, and the following questions: What did we find out about older adults' spoken language use? What can we conclude based on these findings? And importantly, was it useful to also consider the gestural modality, and if so, why? I will argue that apart from the fact that gestures constitute an integral part of face-to-face language and should therefore always be considered in the investigation of such language use, including gestures in the present studies also provided us with new insights into older adults' communicative behavior that would have been missed had I only considered the spoken modality.

Age-related changes in spoken language production: Contextual and cognitive effects

As far as spoken language production in aging is concerned, the findings presented here were only partly in line with earlier investigations of verbal common ground-based recipient design in older adults (e.g., Horton & Spieler, 2007; Hupet et al., 1993; Lysander & Horton, 2012; Saryazdi, Bannon, & Chambers, 2019). These previous studies suggested that older adults would be less efficient than younger adults in adapting their verbal utterances either to incremental common ground (Horton & Spieler, 2007; Hupet et al., 1993; Lysander & Horton, 2012) or personal common ground based on visual scenes (Saryazdi et al., 2019). The narrative task apparently corroborated these findings, as older adults did not adapt their spoken language use to the personal common ground established at the outset of the interactions, either in terms of narration length or in terms of information conveyed. In the spatial task, however, older adults adapted both instruction length and information content to personal and incremental common ground to the same extent as younger adults. The reduction of instruction length across trials was additionally predicted by individual differences in verbal WM, such that participants with higher WM produced longer instructions on earlier as compared to later trials, while participants with lower WM produced relatively shorter instructions on all trials. I interpreted these findings such that older adults generally are able to adapt their verbal utterances according to personal and incremental common ground shared with an addressee; however, whether they do so or not depends on a number of factors: the nature of the task (e.g. narrative vs. spatial), the manner in which common ground is manipulated (i.e., whether the distinction between mutually shared and privileged knowledge is relatively easy or not), the speaker's task goals, and individual differences in verbal WM. It is possible that the process of incrementally building up common ground by establishing mutual reference to a limited set of objects during a referential communication task (as used by Horton & Spieler, 2007; Hupet et al., 1993; Lysander & Horton, 2012) is more difficult for older adults than the common ground manipulation employed in the present thesis. Hence, as already proposed in the introduction, an interplay of contextual and cognitive factors determines older adults' interactive spoken language use.

I would like to argue that it was useful to consider the gestural modality in addition to speech for several reasons. For one thing, we saw that co-speech gestures continue to be an important communicative strategy during older adults' language production. Like younger adults, older adults used co-speech gestures to communicate relevant information to their addressee, and like younger adults, older adults were able to adapt the use of co-speech gestures according to the communicative situation, at least in the spatial task. The spatial task also showed that older adults were as skillful as younger adults in distributing information across the two modalities, such that on earlier trials, they provided more information encoded in speech and in gesture, while on later trials, as common ground incrementally accrued, they tended to provide information either in speech or in gesture. This suggests that like younger adults, older adults seek to design their messages in the most efficient way, even if this means omitting certain pieces of information from speech altogether and expressing it only in gesture. Additionally, in spite of older adults' overall absence of common ground-based recipient design in the narrative task, older adults' gesture rate was sensitive to addressee feedback, such that more verbal and non-verbal feedback from the addressee was associated with higher gesture rates. Although addressee feedback similarly predicted a reduction in narrative events, i.e., a speech-based measure, this effect was mainly driven by the younger adults. The gesture-based findings therefore suggest an actual engagement of older speakers with their addressees, also in the narrative task, which would have gone unnoticed, had we only considered the verbal part of the utterances. Finally, if we assume that co-speech gestures originate from visual and motor imagery, the present findings suggest that older adults rely on these types of mental imagery to the same extent as younger adults during interactive language production – this, too, might have gone unnoticed if only the spoken modality had been considered.

Age-related changes in spoken language comprehension: Contextual and cognitive effects

For speech comprehension, I basically replicated the pattern reported in earlier investigations of SiN comprehension and the benefit of visible speech in aging (e.g., Stevenson et al., 2015; Tye-Murray et al., 2010): while both age groups' speech comprehension benefits from the presence of visible speech, older adults benefit less. I could additionally show that performance in the speech-only and in the visible speech condition were predicted by verbal WM and inhibitory control, both abilities have been proposed to be involved in the comprehension of SiN (e.g., Baum & Stevenson, 2017; Rudner, Mishra, Stenfelt, Lunner, & Rönnberg, 2016; Jesse and Janse, 2012; Tun, O'Kane, & Wingfield, 2002). One interesting difference between the present and previous findings is that older and younger adults' performance did not differ significantly in the speech-only condition (cf. Stevenson et al., 2015; Tye-Murray et al., 2010). However, as argued in Chapter 4, this finding may be specific to the present task and is additionally limited in its interpretability by the fact that older adults performed at chance at the worse noise level. Yet, as for speech production, we saw that older adults' speech comprehension is affected by an interplay of contextual and cognitive factors.

Considering the gestural modality in addition to speech provided me with valuable additional insights into older adults' ability to comprehend language. As stated in the Introduction, the communicative value of co-speech gestures does not only depend on the speaker's communicative intention, but crucially also on the listener's ability to perceive, process, interpret, and integrate the meaning conveyed in these gestures with the meaning conveyed in speech. The present results show that while older adults may have difficulties with one or more of these component processes, gestures still present a valuable source of information in addition to visible speech and may improve older adults' speech comprehension significantly. In fact, older adults may rely considerably on the gestural modality for language comprehension, depending on input quality. If background noise is severe and speech-only comprehension is at chance level, the presence of iconic co-speech gestures in addition to visible speech could boost comprehension accuracy to over 70 % in the present setting. If we keep in mind that every-day, face-to-face language use more often than not is accompanied by co-speech gestures, this also means that older adults' comprehension of such language might be much better than could be expected based solely on laboratory investigations of unimodal speech comprehension. In other words, the way in which language comprehension in aging has been investigated in the past may greatly underestimate older adults' true potential to understand interactive language.

5.3. Conclusion and outlook

5.3.1. Conclusion

In this thesis, I presented novel evidence for the contextual and cognitive factors that modulate interactive language use and showed that these factors continue to play an important role in normal human aging. By looking beyond the spoken modality, new insights could be gained into how older adults produce and comprehend language in interaction. We saw that the gestural modality remains an important communicative strategy, in spite of measurable age-related cognitive differences. Not to consider the gestural modality would mean to obtain an incomplete understanding of older adults' interactive language use, and hence co-speech gestures should be incorporated into general accounts of language use in aging. At the same time, we also saw that interactive, multimodal language use is determined by an interaction of contextual and cognitive factors, and future research may greatly benefit from taking both types of factors into account more systematically.

It is worth stating that the research presented in this thesis was initially inspired by the opposing effects that normal human aging may hypothetically have on co-speech gesture production and comprehension. Based on previous research on younger and older adults' co-speech gesture use and on age-related cognitive changes, relative to younger adults, older adults may either rely *more* on co-speech gestures in order to compensate for cognitive and/or perceptual deficits during language production and comprehension, alternatively, they may rely *less* on co-speech gestures due to cognitive and/or perceptual deficits which negatively affect the gesture production and comprehension processes. I propose that this dichotomy is misguided and that any hypothesis on co-speech gesture production or comprehension in older (but also in younger) adults must consider the influence and interaction of contextual and of cognitive factors at all times.

Previous research on the cognitive effects of gesticulation in younger adults and on gesture production in aging may have underestimated or neglected the role of communicative context. Focusing on the cognitively beneficial effects of co-speech gestures in contexts which do not involve interaction with a genuine addressee may overestimate the relevance of the speaker-oriented, cognitive function that co-speech gestures have in daily interactions. Moreover, we saw that in the case of older adults, an interactive setting appears to be essential for older adults to use co-speech gestures with the same communicative efficiency as younger adults. Earlier, less interactive research had suggested systematic, age-related differences in older adults' gesture production. Obviously, it is still conceivable that age-related cognitive changes affect gesture production in one direction or another, however, in the present studies, the communicative context really proved to be the key factor modulating older adults' multimodal utterance production.

For co-speech gesture comprehension, previous research with younger adults on the beneficial effects of gestures on language comprehension may have underestimated the role of cognitive abilities involved in this process. The present results suggest that in order to benefit from gestures, sufficient cognitive resources are needed, seemingly supporting the hypothesis that older adults rely less on gestures. However, my findings also suggest that whether older adults focus more on the spoken or more on the gestural modality may depend on how reliable the individual signals are: If gestures provide the more dependable signal, older adults might actually rely more on gestures than on speech. These considerations show once more that considering either cognitive or contextual factors in isolation will not yield the full picture of older adults' interactive language use – just like looking at either speech or gesture in isolation would not.

I would like to conclude this research endeavor by stating that normal human aging does not influence communicative behavior in any one predictable way – rather, I found an intricate interplay of the summary variable "age", individual differences in cognitive abilities, the specifics of the communicative situation, the affordances of the spoken and the gestural modality, and certainly many other factors yet to be identified. In the final paragraphs of this chapter, I will propose some directions for future research and implications for practice.

5.3.2. Suggestions for future research and implications for practice

In the studies investigating co-speech gesture production, I found no evidence for the cognitively beneficial functions ascribed to gesticulation. However, I did not systematically assess or manipulate the cognitive load that we presented our participants with. Future research might investigate whether in situations that are known to be cognitively demanding, communicative or other, gesticulation could be beneficial for older adults. For example, when faced with the task of having to memorize and to give more or less complex directions, it could be tested whether producing appropriate gestures improves older adults' memory, and whether more complex directions are accompanied by an increase in gesture frequency. If this were the case, it would provide evidence for speaker-directed, cognitive functions of gestures in aging and might additionally offer strategies for older adults to cope with certain types of memory deficits. Obviously, individual cognitive abilities should still be assessed independent of the gesture elicitation task.

Also, I found some evidence that cognitive factors influence the communicative adaptation of gestures. However, as these effects were only present in one of two tasks, and my explanation is still tentative, it is necessary to replicate these findings with a larger number of participants, and potentially also a variety of tasks.

Finally, I did not systematically compare qualitative features of the gestures produced by younger and older adults, such as gesture size or precision. Future research might investigate whether there are certain age-related differences with respect to these features, which may in turn affect the communicative value of the gestures.

In the study investigating gesture comprehension, one question that remained was whether older adults benefit less from iconic co-speech gestures than younger adults because of difficulties in simultaneous auditory and visual processing, or because of deficits in speech-gesture integration. This might be addressed in a more classic mismatch paradigm (e.g., Kelly, Özyürek, & Maris, 2010) in which an observed gesture either matches or mismatches the accompanying speech semantically, and participants' EEGs are recorded in order to detect potential age-related differences in the time course of speech and gesture processing and integration.

Furthermore, it would be interesting to test whether younger adults under high cognitive load (e.g., induced by a secondary task) behave like older adults, that is whether in younger adults, the processing and/or integration of speech and gesture is modulated by the availability of cognitive resources. If this were the case, it would support the interpretation of older adults' performance differences being due to cognitive limitations.

Finally, I would like to emphasize that co-speech gestures are only one of several visual communicative signals that accompany our speech in face-to-face interactions. Future research on older adults' face-to-face language use should ideally consider not only speech and co-speech gestures, but also other (and potentially subtler) signals, such as eye-gaze, body posture, or facial expressions, and may also reveal further insights by delving deeper into interactive processes during communication.

In addition to these suggestions for future research aimed at understanding healthy older adults' multimodal language production and comprehension in different contexts, the research presented in this thesis also has important implications for future investigations of gesture production and comprehension in age-related pathologies, such as aphasia, Parkinson's disease, or dementia (e.g., Akhavan, Göksun, & Nozari, 2018; Cleary, Poliakoff, Galpin, Dick, & Holler, 2011; Rousseaux, Rénier, Anicet, Pasquier, & Mackowiak-Cordoliani, 2012). As I hope to have convincingly shown, the context in which language is produced and perceived in interaction with cognitive functions is of great importance in the investigation of older adults' communicative abilities. The typical setting in clinical research involves older adults being given isolated gesture production and/or comprehension tasks, or interacting with a speech therapist or experimenter. The setting for the control group is similarly restricted. However, as the present research shows, the presence of a genuine, naïve addressee may be key in encouraging older adults to communicate with the same flexibility as younger adults. Testing older adults in less communicative situations may therefore put them at a systematic disadvantage relative to younger adults, and thus underestimate their true abilities to produce and comprehend gestures. It is therefore essential to consider developing new, more communicative assessment paradigms, involving for example the participation of a genuine addressee or other more contextualized, communicative tasks, in order to gain a more accurate understanding of older adults' abilities, also in age-related pathologies.

References

Abrams, L., & Farrell, M.T. (2011). Language processing in normal aging. In: J. Guendouzi, F. Loncke, & M.J. Williams (Eds.), *The Handbook of Psycholinguistic and Cognitive Processes: Perspectives in Communication Disorders* (pp. 49-73). New York, NY: Psychology Press.

Adams, C., Smith, M.C., Pasupathi, M., & Vitolo, L. (2002). Social context effects on story recall in older and younger women: Does the listener make a difference? *Journal of Gerontology: Psychological Sciences*, *57B*(1), 28-40.

Akeroyd, M. A. (2008). Are individual differences in speech recognition related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired individuals. *International Journal of Audiology, 47*(Suppl. 2), S53-S71.

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

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., 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*, 169-188.

Anderson, S., White-Schwoch, T., Parbery-Clark, A., Kraus, N. (2013). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hearing Research, 300*, 18–32.

Arslan, B. & Göksun, T. (2020). Aging, working memory, and mental imagery: understanding gestural communication in younger and older adults. *Quarterly Journal of Experimental Psychology*. doi: 10.1177/1747021820944696.

Avivi-Reich, M., Puka, K. & Schneider, B. A. (2018). Do age and linguistic background alter the audiovisual advantage when listening to speech in the presence of energetic and informational masking? *Attention, Perception, & Psychophysics, 80*(1), 242-262.

Baddeley, A. D. & Hitch, G. (1974). Working memory. *Psychology of learning and motivation, 8*, 47-89.

Bates, D., Maechler, M., & Bolker, B. (2017). lme4: Linear mixed-effects models using 'Eigen' and S4. R package version 1.1–13. Retrieved from https://cran.r-project.org/web/packages/lme4/

Baum, S. H. & Stevenson, R. A. (2017). Shifts in Audiovisual Processing in Healthy Aging. *Current Behavioral Neuroscience Reports, 4* (3), 198-208.

Bavelas, J.B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology, 79*(6), 941-952.

Bavelas, J.B., & Chovil, N. (2000). Visible acts of meaning. An integrated model of language in face-to-face dialogue. *Journal of Language and Social Psychology, 19*(2), 163-194.

Bavelas, J.B., Chovil, N., Coates, L., & Roe, L. (1995). Gestures specialized for dialogue. *Personality and Social Psychology Bulletin, 21*(4), 394-405.

Bavelas, J.B., Chovil, N., Lawrie, D.A., & Wade, A. (1992). Interactive gestures. *Discourse Processes, 15*, 469-489.

Bavelas, J.B., Gerwing, J., Sutton, C., & Provost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language, 58*, 495-520.

Bavelas, J.B., Kenwood, C., Johnson, T., & Phillips, B. (2002). An experimental study of when and how speakers use gestures to communicate. *Gesture, 2*(1), 1-17.

Boersma, P. & Weenink, D. (2015). Praat: Doing phonetics by computer. [Computer software]. Version 5.4.15. http://www.praat.org.

Bopp, K. L. & Verhaeghen, P. (2005) Aging and verbal memory span: a meta-analysis. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 60*(5), 223-233.

Bortfeld, H., Leon, S.D., Bloom, J.E., Schober, M.F., & Brennan, S.E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*, 123-147.

Brennan, S., Galati, A., & Kuhlen, A.K. (2010). Two minds, one dialogue: Coordinating speaking and understanding. *Psychology of Learning and Motivation, 53*, 301-344.

Brown, A.S. & Nix, L.A. (1996). Age-related changes in the tip-of-the-tongue experience. *The American Journal of Psychology, 109*(1), 79-91.

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review, 16*(5), 893–900.

Burke, D.M., MacKay, D.G., Worthley, J.S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in younger and older adults. *Journal of Memory and Language, 30*, 542-579.

Burke, D.M., & Shafto, M.A. (2008). Language and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (p. 373–443). Psychology Press.

Butterworth, B. (1974). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research, 4*(1), 75-87.

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics, 47*, 14-27.

CHABA (Working Group on Speech Understanding and Aging. Committee on Hearing, Bioacoustics, and Biomechanics, Commission on Behavioral and Social Sciences and Education, National Research Council). (1988). Speech understanding and aging. *Journal of the Acoustical Society of America, 83*, 859-895.

Chu, M., Meyer, A., Foulkes, L, & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: The role of cognitive ability and empathy. *Journal of Experimental Psychology: General, 143*(2), 694-709.

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

Clark, H.H., & Brennan, S. (1991). Grounding in communication. In L.B. Resnick, J.M. Levine, & J.S.D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). American Psychological Association.

Clark, H.H., & Murphy, G.L. (1983). Audience design in meaning and reference. In J.F. Leny & W. Kintsch (Eds.), *Language and comprehension* (pp. 287-299). Amsterdam: North-Holland Publishing Co.

Clark, H.H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*(1), 1-39.

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*, 753-756.

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

Corsi, P.M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International, 34*, 819B.

Coulson, S., & Wu, Y.C. (2019). Working memory and co-speech iconic gestures. In A.K. Goel, C.M. Seifert, & C.C. Freksa (Eds.), *Proceedings of the 41st Annual Meeting of the Cognitive Science Society (CogSci 2019)* (pp. 1553-1558). Montreal, QB: Cognitive Science Society.

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.

Davis, C. & Kim, J. (2006). Audio-visual speech perception off the top of the head. *Cognition, 100*(3), B21-B31.

Della Sala, S., Gray, C., Baddeley, A., & Wilson, L. (1997). *The visual patterns test: A new test of short-term visual recall.* Feltham, Suffolk: Thames Valley Test Company.

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*, 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, 177-178*, 7-17.

Drijvers, L., Özyürek, A. & Jensen, O. (2018). Hearing and seeing meaning in noise: Alpha, beta, and gamma oscillations predict gestural enhancement of degraded speech comprehension. *Human Brain Mapping, 39*(5), 2075-2087.

Dubno, J. R., Dirks, D. D, & Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *Journal of the Acoustical Society of America, 76*(1), 87-96.

ELAN (Version 4.9.4) [Computer software]. (2016-05-19). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from https://tla.mpi.nl/tools/tla-tools/elan/

Engelkamp, J. (1998). *Memory for actions*. Psychology Press/Taylor & Francis (UK).

Feyereisen, P. (1997). Use of a meta-analytic procedure shows an age-related decline in picture naming: Comments on Goulet, Ska, and Kahn (1994). *Journal of Speech, Language, and Hearing Research, 40*, 1328–1333.

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.

Fung, H.H., Carstensen, L.L., & Lutz, A.M. (1999). Influence of time on social preferences: Implications for life-span development. *Psychology and Aging, 14*, 595-604.

Fussel, S.R., & Krauss, R.M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology, 62*, 378-391.

Galati, A., & Brennan, S. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62*, 35-51.

Galati, A., & Brennan, S. (2014). Speakers adapt gestures to addressees' knowledge: implications for models of co-speech gesture. *Language, Cognition and Neuroscience, 29*(4), 435-451.

Gerwing, J., & Bavelas, J. (2004). Linguistic influences on gesture's form. *Gesture, 4*(2), 157-195.

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*, 174-180.

Glosser, G. & Deser, T. (1992). A comparison of changes in macrolinguistic and microlinguistic aspects of discourse production in normal aging. *Journal of Gerontology, 47*(4), 266-272.

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

Gould, O., & Shaleen, L. (1999). Collaboration with diverse partners: How older women adapt their speech. *Journal of Language and Social Psychology, 18*(4), 395-418.

Habets, B., Kita, S., Shao, Z., Özyürek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech–gesture integration during comprehension. *Journal of Cognitive Neuroscience, 23*, 1845-1854.

Hasher, L., Lustig, C., & Zacks, R. (2007). Inhibitory mechanisms and the control of attention. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in working memory* (pp. 227-249). New York, NY: Oxford University Press.

Hasher, L., & Zacks, R. (1988). Working memory, comprehension, and aging: A review and a new view. In G.H. Bower (Ed.), *The psychology of learning and motivation Vol. 2* (pp. 193-225). San Diego, CA: Academic Press.

Healey, H., & Grossmann, M. (2016). Social coordination in older adulthood: A dual-process model. *Experimental Aging Research, 42*(1), 112-127.

Hilliard, C., & Cook, S.W. (2015). Bridging gaps in common ground: Speakers design their gestures for their listeners. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 42*, 91-103.

Hoetjes, M., Koolen, R., Goudbeek, M., Krahmer, E., & Swerts, M. (2015). Reduction in gesture during the production of repeated references. *Journal of Memory and Language, 79-80*, 1-17.

Holle, H. & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience, 19*, 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*, 875–884.

Holler, J., & Bavelas, J. (2017). Multi-modal communication of common ground. In: R Breckinridge Church, M.W. Alibali, & S.D. Kelly (Eds.), *Why gesture? How the hands function in speaking, thinking and communicating* (pp. 213-240). Amsterdam: John Benjamins.

Holler, J., & Beattie, G. (2003a). How iconic gestures and speech interact in the representation of meaning: are both aspects really integral to the process? *Semiotica, 146*, 81-116.

Holler, J., & Beattie, G. (2003b). Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener? *Gesture, 3*, 127-154.

Holler, J., Kokal, I., Toni, I., Hagoort, P., Kelly, S.D., & Özyürek, A. (2015). Eye'm talking to you: Speakers' gaze direction modulates co-speech gesture processing in the right MTG. *Social Cognitive & Affective Neuroscience, 10*, 255-261.

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*, 692-697.

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., Turner, K., & Varcianna, T. (2013). It's on the tip of my fingers: Co-speech gestures during lexical retrieval in different social contexts. *Language and Cognitive Processes, 28*(10), 1509-1518.

Holler, J., & Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes, 24*(2), 267-289.

Holler, J., & Wilkin, K. (2011a). An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses. *Journal of Pragmatics, 43*, 3522-3536.

Holler, J., & Wilkin, K. (2011b). Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *Journal of Nonverbal Behavior, 35*, 133-153.

Horton, W.S., & Gerrig, R. (2005). The impact of memory demands on audience design during language production. *Cognition, 96*, 127-142.

Horton, W.S., & Gerrig, R. (2016). Revisiting the memory-based approach to common ground. *Topics in Cognitive Science, 8*, 780-795.

Horton, W.S., & Spieler, D.H. (2007). Age-related differences in communication and audience design. *Psychology and Aging, 22*(2), 281-290.

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. (2018). Gesture as simulated action: revisiting the framework. *Psychonomic Bulletin & Review, 26*, 721-752.

Hostetter, A.B., Alibali, M.W., & Kita, S. (2007). I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language and Cognitive Processes, 22*(3), 313-336.

Hothorn, R., Bretz, F., & Westfall, P. (2017). multcomp: Simultaneous inference in general parametric models. R package version 1.4-8. Retrieved from https://cran.r-project.org/web/packages/multcomp

Humes, L. E. (2002). Factors underlying the speech-recognition performance of elderly hearing-aid wearers. *Journal of the Acoustical Society of America, 112*, 1112-1132.

Humes, L. E. (2007). The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *Journal of the American Academy of Audiology, 18*, 590-603.

Humes, L. E., Watson, B. U., Christensen, L. A., Cokely, C. G., Halling, D. C., & Lee, L. (1994). Factors associated with individual differences in clinical measures of speech recognition among the elderly. *Journal of Speech and Hearing Research, 37*, 465-474.

Hupet, M., Chantraine, Y., & Nef, F. (1993). References in conversation between young and old normal adults. *Psychology and Aging, 8*(3), 339-346.

Isaacs, E.A., & Clark, H.H. (1987). References in conversations between experts and novices. *Journal of Experimental Psychology: General, 116*, 26-37.

Isaacs, B., & Kennie, A.T. The set test as an aid to the detection of dementia in old people. *British Journal of Psychiatry, 123* (1973), 467–471.

Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language, 56*, 291-303.

James, L.E., Burke, D.M., Austin, A., & Hulme, E. (1998). Production and perception of “verbosity” in younger and older adults. *Psychology and Aging, 13*, 355-367.

Janse, E. (2012). A non-auditory measure of interference predicts distraction by competing speech in older adults. *Aging, Neuropsychology and Cognition, 19*, 741-758.

Jesse, A., & Janse, E. (2012). Audiovisual benefit for recognition of speech presented with single-talker noise in older listeners. *Language and Cognitive Processes, 27*(7/8), 1167-1191.

Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition, 144*, 1-13.

Kavé, G., & Goral, M. (2017). Do age-related word retrieval difficulties appear (or disappear) in connected speech? *Aging, Neuropsychology, and Cognition, 24*(5), 508-527.

Keller-Cohen, D. (2014). Audience design and social relations in aging. *Research on Aging, 37*(7), 741-762.

Kelly, S.D. (2017). Exploring the boundaries of gesture-speech integration during language comprehension. In R.B. Church, M.W. Alibali & S.D. Kelly (Eds.), *Why gesture? How the hands function in speaking, thinking and communicating* (pp. 243-265). John Benjamins Publishing: Amsterdam.

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*, 577-592.

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

Kelly, S.D., Healey, M., Özyürek, A., & Holler, J. (2015). The processing of speech, gesture and action during language comprehension. *Psychonomic Bulletin & Review, 22*, 517-523.

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

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.

Kemper, S., Herman, R.E., & Liu, C.-J. (2004). Sentence production by young and older adults in controlled contexts. *The Journals of Gerontology Series B Psychological Sciences and Social Sciences, 59*(5), 220-224.

Kemper, S., Kynette, D., Rash, S., O'Brien, K., & Sprott, R. (1989). Life-span changes to adults' language: Effects of memory and genre. *Applied Psycholinguistics, 10*(1), 49–66.

Kemper, S. & Sumner, A. (2001). The structure of verbal abilities in young and older adults. *Psychology and Aging, 16*(2), 312-322.

Kendon, A. (1985). Some uses of gesture. In D. Tannen and M. Saville-Troike (Eds.): *Perspectives on silence* (pp. 215-234). Norwood, MA: Ablex.

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

Kessels, R., van Zandvoort, M., Postma, A., Kappelle, L.J., & de Haan, E. (2000). The Corsi Block-Tapping Task: Standardization and normative data. *Applied Neuropsychology, 7*(4), 252-258.

Keysar, B., Barr, D.J., & Horton, W.S. (1998). The egocentric bias of language use: Insights from a processing approach. *Current Directions in Psychological Science, 7*(2), 46-50.

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

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

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

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., van Gijn, I., & van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In Wachsmuth, I. & Fröhlich, M. (Eds) *Gesture and Sign Language in Human-Computer Interaction*. GW 1997. Lecture Notes in Computer Science, Vol 1371 (pp. 23–35). Springer: Berlin, Heidelberg.

Klencklen, G., Després, O., & Dufour, A. (2012). What do we know about aging and spatial cognition? Reviews and perspectives. *Aging Research Review, 11*(1), 123-135.

Koch, X., & Janse, E. (2016). Speech rate effects on the processing of conversational speech across the adult life span. *Journal of the Acoustical Society of America, 139*(4), 1618-1636.

Koeritzer, M. A., Rogers, C. S., Van Engen, K. J., & Peelle, J. E. (2018). The impact of age, background noise, semantic ambiguity, and hearing loss on recognition memory for spoken sentences. *Journal of Speech, Language, and Hearing Research, 61*(3),740-751.

Krauss, R.M., Chen, Y., & Gottesman, R.F. (2000). Lexical gestures and lexical access: a process model. In D. McNeill (Ed.), *Language and gesture* (pp. 93-116). Cambridge: Cambridge University Press.

Kuhlen, A.K., & Brennan, S.E (2013). Language in dialogue: when confederates might be hazardous to your data. *Psychonomic Bulletin & Review, 20*, 54-72.

Kuhlen, A.K., Galati, A., & Brennan, S.E. (2012). Gesturing integrates top-down and bottom-up information: Effects of speakers' expectations and addressees' feedback. *Language and Cognition, 4*, 17-41.

Kuznetsova, A., Brockhoff, P.B., & Bojesen Christensen, R.H. (2016). lmerTest: Tests in linear mixed effects models. R package version 2.0-33. Retrieved from http://cran.r-project.org/web/packages/lmerTest

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

Lenth, R. (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.1.3. Retrieved from https://cran.r-project.org/web/packages/emmeans

Levelt, W.J.M. (1989). *ACL-MIT Press series in natural-language processing. Speaking: From intention to articulation*. The MIT Press.

Lineweaver, T.T., Hertzog C. (1998). Adults' efficacy and control beliefs regarding memory and aging: Separating general from personal beliefs. *Aging, Neuropsychology, and Cognition, 5*, 264–296.

Long, M.R., Horton, W.S., Rohde, H., & Sorace, A. (2018). Individual differences in switching and inhibition predict perspective-taking across the lifespan. *Cognition, 170*, 25-30.

Long, M.R., Rohde, H., & Rubio-Fernandez, P. (2020). The pressure to communicate efficiently continues to shape language use later in life. *Scientific Reports, 10*, 8214.

Lysander, K., & Horton, W.S. (2012). Conversational grounding in younger and older adults: The effect of partner visibility and referent abstractness in task-oriented dialogue. *Discourse Processes, 49*(1), 29-60.

Mair, P., & Wilcox, R. (2017). WRS2: A collection of robust statistical methods. R package version 0.9-2. Retrieved from https://cran.r-project.org/web/packages/WRS2

Marini, A. & Andreetta, S. (2016). Age-related effects on language production: A combined psycholinguistic and neurolinguistic perspective. In H. H. Wright (Ed.), *Cognition, language and aging* (p. 55–79). John Benjamins Publishing Company.

Marini, A., Boewe, A., Caltagirone, C., & Carlomagno, S. (2005). Age-related differences in the production of textual descriptions. *Journal of Psycholinguistic Research, 34*(5), 439-63.

Markostamou, I., Coventry, K., Fox, C., & McInnes, L. (2015). Both symbolic and embodied representations contribute to spatial language processing: Evidence from younger and older adults. In: Noelle, D.C., Dale, R., Warlaumont, A.S., Yoshimi, J., Matlock, T., Jennings, C.D., & Maglio, P.P. (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society (CogSci 2015)*. Austin, TX: Cognitive Science Society (pp. 1518-1523).

Martin, A., Wiggs, C.L., Lalonde, F., & Mack, C. (1994). Word retrieval to letter and semantic cues: a double dissociation in normal subjects using interference tasks. *Neuropsychologia, 32*(12), 1487-1494.

Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2016). Can you handel this? The impact of object affordances on how co-speech gestures are produced. *Language, Cognition, and Neuroscience, 31*(3), 430-440.

Maylor, E.A. (1990). Age, blocking and the tip of the tongue state. *British Journal of Psychology, 81*, 123–134.

McNeill, D. (1992). *Hand and Mind*. Chicago, London: The Chicago University Press.

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

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

Melinger, A., & Kita, S. (2007). Conceptualization load triggers gesture production. *Language and Cognitive Processes, 22*(4), 473-500.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009). Communicative gestures and memory load. In: *Proceedings of the 31st Annual Conference of the Cognitive Science Society (CogSci 2009)*. Amsterdam, Netherlands: Cognitive Science Society (pp. 1569-1574).

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2011). Seeing and Being Seen: The effects on gesture production. *Journal of Computer-Mediated Communication, 17*(1), 77-100.

Morsella. E. & Krauss, R.M. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology, 117*(3), 411-424.

Mortensen, L., Meyer, A.S., & Humphreys, G.W. (2006). Age-related effects on speech production: A review. *Language and Cognitive Processes, 21*, 238-290.

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.

Nilsson, L.-G., Sternäng, O., Rönnlund, M., Nyberg, L. (2009). Challenging the notion of an early-onset of cognitive decline. *Neurobiology of Aging, 30*, 521-524.

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*, 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*, 857–870.

Özyürek, A. (2002). Do speakers design their co-speech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language, 46*(4), 688-704.

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

Özyürek, A. (2017). Function and processing of gesture in the context of language. In R.B. Church, M.W. Alibali, & S.D. Kelly (Eds.), *Why gesture? How the hands function in speaking, thinking and communicating* (pp. 39-58). Amsterdam: John Benjamins Publishing.

Özyürek, A. (2018). Role of gesture in language processing: Toward a unified account for production and comprehension. In S.-A. Rueschemeyer, & M. G. Gaskell (Eds.), *Oxford Handbook of Psycholinguistics* (2nd ed., pp. 592-607). Oxford: Oxford University Press.

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.

Parkington, J.E., & Leiter, R.G. (1949). Partington's Pathway Test. *The Psychological Service Center Bulletin, 1*, 9–20.

Parrill, F. (2010). The hands are part of the package: Gesture, Common ground and information packaging. In S. Rice & J. Newman (Eds.), *Empirical and Experimental Methods in Cognitive/Functional Research* (pp. 285-302). Stanford: CSLI Publications.

Peelle, J.E. (2019). Language and aging. In de Zubicaray and Schiller (Eds.), *The Oxford Handbook of Neurolinguistics* (pp. 295–316). Oxford: Oxford University Press.

Peelle, J.E. & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex, 68*, 169-181.

Pichora-Fuller, M.K., Alain, C., and Schneider, B.A. (2017). Older adults at the cocktail party. In J.C. Middlebrooks, J.Z. Simon, A.N. Popper, & R.R. Fay (Eds.): *The auditory system at the cocktail party* (pp. 227-259). New York: Springer International Publishing.

Pichora-Fuller, M.K., Schneider, B.A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America, 97*(1), 593-608.

Pichora-Fuller, M.K. & Singh, G. (2006). Effects of age on auditory and cognitive processing: implications for hearing aid fitting and audiologic rehabilitation. *Trends in Amplification, 10*(1), 29-59.

Pratt, M.W. & Robins, S.L. (1991). That’s the way it was: Age differences in the structure and quality of adults’ narratives. *Discourse Processes, 14*, 73–85.

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science, 6*(1), 5-42.

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*, 226-231.

R Development Core Team (2015). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org

Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L.-G. (2005). Stability, growth, and decline in adult life span development of declarative memory: Cross-sectional and longitudinal data from a population-based study. *Psychology and Aging, 20*(1), 3-18.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do You See What I Am Saying? Exploring Visual Enhancement of Speech Comprehension in Noisy Environments. *Cerebral Cortex, 17*(5), 1147–1153.

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.

Rowbotham, S. J., Holler, J., Wearden, A., & Lloyd, D. M. (2016). I see how you feel: Recipients obtain additional information from speakers' gestures about pain. *Patient Education and Counseling, 99*(8), 1333-1342.

Rudner, M., Mishra, S., Stenfelt, S., Lunner, T., & Rönnberg, J. (2016). Seeing the Talker's Face Improves Free Recall of Speech for Young Adults With Normal Hearing but Not Older Adults With Hearing Loss. *Journal of Speech, Language, and Hearing Research, 59*, 590-599.

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

de Ruiter, J.P. (2007). Postcards from the mind. The relationship between speech, imagistic gesture, and thought. *Gesture, 7*, 21-38.

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.

Sacks, H., Schegloff, E.A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696-735.

Salthouse, T.A. (1991). *Theoretical perspectives on cognitive aging*. Hillsdale, NJ: Lawrence Erlbaum.

Salthouse, T.A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging, 30*, 3-18.

Salthouse, T.A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society, 16*(5), 754-760.

Sanchez-Cubillo, I., Perianez, J.A., Adrover-Roig, D., Rodriguez-Sanchez, J.M., Rios-Lago, M., Tirapu, J., & Barcelo, F. (2009). Construct validity of the Trail Making Test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society, 15*, 438-450.

Saryazdi, R., Bannon, J., & Chambers, C.G. (2019). Age-related differences in referential production: A multiple-measures study. *Psychology and Aging, 34*(6), 791-804.

Schneider, B.A., Daneman, M., Murphy, D.R., & See, S.K. (2000). Listening to discourse in distracting settings: The effects of aging. *Psychology and Aging, 15*(1), 110–125.

Schneider, B.A., Daneman, M., & Pichora-Fuller, M.K. (2002). Listening in aging adults: From discourse comprehension to psychoacoustics. *Canadian Journal of Experimental Psychology, 56*(3), 139–152.

Schneider, B.A. & Pichora-Fuller, M.K. (2000). Implications of perceptual deterioration for cognitive aging research. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (p. 155–219). Lawrence Erlbaum Associates Publishers.

Schubotz, L., Holler, J., Drijvers, L. and 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*. doi:10.1007/s00426-020-01363-8.

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.

Smayda, K. E., Van Engen, K. J., Maddox, W. T., & Chandrasekaran, B. (2016). Audio-Visual and Meaningful Semantic Context Enhancements in Older and Younger Adults. *PLoS ONE 11*(3): e0152773.

Shao, Z., Roelofs, A., & Meyer, A. (2012). Sources of individual differences in the speed of naming objects and actions: The contribution of executive control. *Quarterly Journal of Experimental Psychology, 65*, 1927-1944.

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*, 115–125.

Sommers, M. D., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear and Hearing, 26*(3), 263-75.

Sommers, M. S. & Phelps, D. (2016). Listening effort in younger and older adults: A comparison of auditory-only and auditory-visual presentations. *Ear and Hearing, 37*, 62S-68S.

Stevenson, R. A., Nelms, C. E., Baum, S. H., Zurkovsky, L., Barense, M. D., Newhouse, P. A., & Wallace, M. T. (2015). Deficits in audiovisual speech perception in normal aging emerge at the level of whole-word recognition. *Neurobiology of Aging, 36*(1), 283–91.

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.

Sumby, W. H. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America, 26*, 212.

Theocharopoulou, F., Cocks, N., Pring, T., & Dipper, L.T. (2015). TOT phenomena: Gesture production in younger and older adults. *Psychology and Aging, 30*(2), 245-252.

Thompson, L. A. (1995). Encoding and memory for visible speech and gestures: A comparison between young and older adults. *Psychology and Aging, 10*(2), 215-228.

Thompson, L. & 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*, 347-349.

Thompson, L. A. & Malloy, D. (2004). Attention resources and visible speech encoding in older and younger adults. *Experimental Aging Research, 30*, 1-12.

Thornton, R., & Light, L.L. (2006). Language comprehension and production in normal aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the Psychology of Aging, 6th Ed*. (pp. 261-287). San Diego, CA: Elsevier Academic Press.

Tun, P. A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in younger and older listeners. *Psychology and Aging, 17*(3), 453-467.

Tun, P. A., & Wingfield, A. (1999). One voice too many: Adult age differences in language processing with different types of distracting sounds. *Journal of Gerontology: Psychological Sciences, 54B*(5), 317-327.

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

Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J., & Hale, S. (2010). Aging, audiovisual integration, and the principle of inverse effectiveness. *Ear and Hearing, 31*(5), 636-44.

Tye-Murray, N., Spehar, B., Myerson, J., Hale, S., & Sommers, M. (2016). Lipreading and audiovisual speech recognition across the adult lifespan: Implications for audiovisual integration. *Psychology and Aging, 31*(4), 380-389.

Underwood, K. (2010). Interactive remembering: Insights into the communicative competence of older adults. *Journal of Aging Studies, 24*, 145-166.

Verhaegen, P. (2011). Aging and executive control: Reports of a demise greatly exaggerated. *Current Directions in Psychological Science, 20*(3), 174-180.

Wagner Cook, S., 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.

Wardlow, L. (2013). Individual differences in speaker's perspective taking: The roles of executive control and working memory. *Psychonomic Bulletin & Review, 20*(4), 766-772.

Wechsler, D. (1981). *WAIS-R Manual: Wechsler Adult Intelligence Scale-Revised*. New York: Psychological Corp., 1981.

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.

Winneke, A. H., & Phillips, N. A. (2011). Does audiovisual speech offer a fountain of youth for old ears? An event-related brain potential study of age differences in audiovisual speech perception. *Psychology and Aging, 26*(2), 427-438.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. In Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. http://tla.mpi.nl/tools/tla-tools/elan

Wu, Y.C., & Coulson, S. (2007). Iconic gestures prime related concepts: an ERP study. *Psychonomic Bulletin & Review, 14*, 57-63.

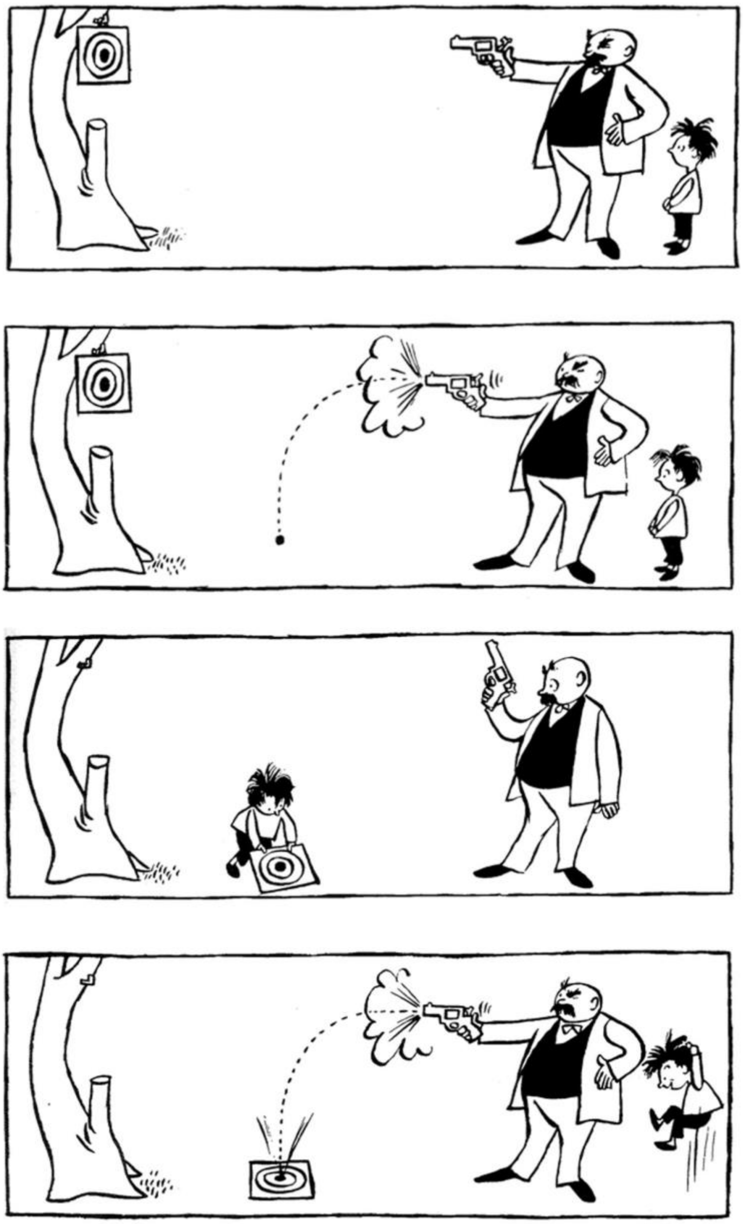
Wu, Y.C., & Coulson, S. (2010). Gestures modulate speech processing early in utterances. *Neuroreport, 21*, 522-526.

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

Appendices

Appendix A

Example comic strip



Artwork by cartoon artist e.o. plauen (Erich Ohser). Works by e.o. plauen are in the public domain.

Appendix B

Example narrative script

1.1 There are a man and a child/a father and a son.
1.2 A bull's eye is hanging in a tree.
1.3 The father has a gun.
1.4 The father and the son stand at a short distance from the tree.
1.5 The father is aiming at the bull's eye.
1.6 The boy is watching.

2.1 The father shoots.
2.2 The bullet doesn't go straight.
2.3 The bullet hits the ground.
2.4 It lands in between the tree and the father.
2.5 The boy is watching.

3.1 The boy has an idea.
3.2 He takes the bull's eye out of the tree.
3.3 He puts it on the ground, where the bullet had landed earlier.
3.4 The father looks confused.

4.1 The father shoots another time.
4.2 The bullet is not going straight again.
4.3 This time it hits the bull's eye exactly in the middle.
4.4 The boy jumps in the air.

Appendix C

Gesture frequency and gesture rate per narrative event

In addition to the analyses of gesture rate per 100 words and of the proportion of multimodal events reported in the main paper, here we also report analyses of simple gesture frequency (i.e., the number of gestures produced per narration) and of gesture rate per narrative event (i.e., dividing the number of gestures a given participant produced during each trial by the number of narrative events mentioned for this trial, see Galati & Brennan, 2014), for each condition within each trial separately.

Means and standard deviations per age group and condition for these additional gesture-based measures are reported in Table C1.

Table C1. Means (and SD) for gesture frequency and gesture rate per narrative event for each age group and condition. CG = common ground condition, no-CG = no common ground condition.

	Younger		Older	
	CG	No-CG	CG	No-CG
Gesture frequency	2.02 (1.39)	4.78 (2.39)	3.01 (2.09)	2.74 (2.18)
Gestures/narr. event	.56 (.42)	.81 (.34)	.60 (.43)	.51 (.43)

To investigate the influence of age and the common ground manipulation on gesture frequency and gesture rate per narrative event, we fitted linear mixed-effect models in R as described in the methods section of the main paper.

Table C2 summarizes the results for the models predicting the two dependent measures based on age and common ground manipulation.

Table C2. Linear mixed-effects models for the effects of age and common ground manipulation on gesture frequency and gesture rate per narrative event. Age group = young and Condition = CG[a] are on the intercept. N = 32.[b]

	Gesture frequency				Gestures/narrative event			
	β	SE	t	p	β	SE	t	p
Intercept	2.02	.50	4.07	< .001	.48	.12	3.94	< .001
Age group$_{old}$	.99	.50	1.62	.12	.13	.14	.94	.35
Condition$_{no\text{-}CG}$[a]	2.76	.36	7.65	< .001	.33	.06	5.10	< .001
Age group$_{old}$: Condition$_{no\text{-}CG}$	-3.03	.51	-5.94	< .001	-.42	.09	-4.61	< .001

[a] CG = common ground; no-CG = no common ground.

[b] Both models contain random intercepts for participants and items. The model predicting gestures/narrative event includes by-participant random slopes for common ground manipulation.

The absence of a main effect for age group indicates that there was no age-related difference for the two measures in the CG condition. The significant main effect for common ground manipulation indicates that for the younger adults, gesture frequency and gesture rate were higher in the no-CG as opposed to the CG condition. The significant interactions between age group and common ground manipulation indicate that the increase in gesture frequency and rate was only significantly present in the younger but not the older adults.

Individual contrasts confirm this analysis, with younger adults producing significantly more gestures and gesturing at a significantly higher rate in the no-CG as opposed to the CG condition ($\beta = 2.76$, $SE = .37$, $t = 7.41$, $p < .001$ and $\beta = .33$, $SE = .06$, $t = -5.10$, $p < .001$, respectively), whereas this difference was not significant for older adults ($p > .05$). Comparisons further showed that younger and older adults did not differ in the rate at which they gestured in the CG condition for both measures (both *p*'s > .05). However, there was a significant age-related difference in the no-CG condition (for gesture frequency, $\beta = 2.04$, $SE = .81$, $t = 2.52$, $p = .02$; for gestures/narrative event, $\beta = .29$, $SE = .14$, $t = 2.09$, $p = .04$), such that younger adults produced significantly more gestures and gestured at a significantly higher rate than older adults.

Appendix D

Cognitive test battery

Here, we provide a more detailed description of the different tasks we used to assess the individual cognitive abilities, including details on task administration and scoring procedure. With the exception of the Operation span task, which was computer-based, all other tasks used to measure cognitive skills were pen-and-paper versions.

Verbal working memory (Verbal WM)

The Operation span task (O-span) is a standard measure of verbal working memory (Turner & Engle, 1989). The Dutch version of the task used here, as well as the scoring procedure, are based on Shao, Roelofs, and Meyer (2012). Participants were required to evaluate the accuracy of 60 simple mathematical operations while remembering unrelated words for later serial recall. The O-span score was calculated as the sum of words that were recalled in the proper order on trials with correct responses to the mathematical problem, the highest possible score being 60. Due to time-out, O-span data could not be collected from one older male participant.

Visual working memory (Visual WM)

To assess the visuo-spatial component of visual WM, participants performed the Visual Patterns Test (VPT, Della Sala, Gray, Baddeley, & Wilson, 1997). Participants were briefly presented with visual patterns of increasing complexity, and had to reproduce these patterns. The VPT score is the highest level of complexity at which at least one of three patterns is recalled correctly. Due to time-out, VPT data could not be collected from two older female participants and one older male participant.

To assess the visuo-sequential component of visual WM, participants performed the Corsi Block-Tapping Task (CBT, Corsi, 1972). The task was administered based on the protocol proposed by Kessels, van Zandvoort, Postma, Kappelle, & de Haan (2000). In this test, participants were asked to imitate the experimenter in tapping nine black cubes mounted on a black board in sequences of increasing length, going in steps of two sequences per level. The final score for each participant was calculated as the length of the last correctly repeated sequence multiplied by the number of correctly repeated sequences (i.e. the number of correct trials).

Inhibitory control

Participants performed the Trail Making Test part A and B (TMT A and TMT B) (Parkington & Leiter, 1949) in order to assess their inhibitory control. In part A, participants used a pencil to connect a series of 25 encircled numbers in numerical order. In part B, participants connected 25 encircled numbers and letters in numerical and alphabetical order, alternating between numbers and letters, requiring the continuous shifting of attention between numbers and letters. The difference between the time to complete part A and part B (TMT B-A) is seen as a measure of inhibition/interference control (isolating the switching component of part B by subtracting the visual search and speed component of part A, see Sanchez-Cubillo, Perianez, Adrover-Roig, Rodriguez-Sanchez, Rios-Lago, Tirapu, & Barcelo, 2009).

Semantic fluency

The animal naming task is a standard measure of semantic fluency (Isaacs & Kennie, 1973). Participants were asked to generate as many unique animal names as possible within 60 seconds. Every unique response is given a point, with repetitions receiving no point.

Appendix E

Full model summaries additional analyses

Table E1. Linear mixed-effects models for the effects of age and common ground manipulation on explicit references to common ground and addressee feedback. Age group = young and Condition = CG[a] are on the intercept. N = 32.[b]

	Explicit reference to common ground				Addressee feedback			
	β	SE	t	p	β	SE	t	p
Intercept	.74	.07	9.96	< .001	.07	.01	10.93	< .001
Age group$_{old}$	-.41	.10	-3.87	< .001	-.02	.01	-1.99	.06
Condition$_{no\text{-}CG}$[a]	-.67	.07	-9.84	< .001	-.02	.01	-2.96	.004
Age group$_{old}$: Condition$_{no\text{-}CG}$	.51	.10	5.33	< .001	-	-	-	-

[a] CG = common ground; no-CG = no common ground.

[b] Both models contain random intercepts for participants and items, but no by-participant random slopes for common ground manipulation.

Table E2. Linear mixed-effects models for the effects of age, common ground manipulation, and addressee feedback on narrative event count and gesture rate per 100 words. Age group = young and Condition = CG[a] are on the intercept. N = 32.[b]

	Narrative events				Gesture rate per 100 words			
	β	SE	t	p	β	SE	t	p
Intercept	6.43	.85	7.55	.009	3.11	1.97	1.58	.17
Addressee feedback	-10.63	4.15	-2.56	.01	36.16	9.90	3.65	< .001
Age group$_{old}$	-	-	-	-	1.59	1.48	1.07	.29
Condition$_{no\text{-}CG}$[a]	-	-	-	-	3.01	.87	3.48	< .001
Age group$_{old}$: Condition$_{no\text{-}CG}$	-	-	-	-	-4.42	1.19	-3.73	< .001

[a] CG = common ground; no-CG = no common ground.

[b] Both models contain random intercepts for participants and items. The model predicting narrative event count includes by-participant random slopes for the common ground manipulation.

Appendix F

Correlations between cognitive predictors and dependent measures

Tables F1 and F2 list the correlations between dependent variables and cognitive predictors (z-scored) for younger and older adults respectively. Note that we multiplied the inhibitory control task's scores with -1, so that higher scores would represent better performance. In the younger adults, none of the cognitive measures were significantly correlated with the dependent variables. In the older adults, verbal WM was positively correlated with word and narrative event count, such that the higher the verbal WM, the larger the number of words and narrative events.

Table F1. Spearman's rank correlation rho for the dependent measures and cognitive predictors (z-scored). Younger adults.

	Words	Narrative events	Gestures/ 100 words	% Multimodal events
Verbal WM	-0.01	-0.04	0.12	0.1
Visuo-sequential WM	.01	.1	.26	.18
Visuo-spatial WM	-.27	.02	-.15	-.36
Executive control	0.06	0.16	0.15	0.15
Semantic fluency	0.26	0.38	0.23	0.28

*** = p < .001, ** = p < .01, * = p < .05

Table F2. Spearman's rank correlation rho for the dependent measures and cognitive predictors (z-scored). Older adults.

	Words	Narrative events	Gestures/ 100 words	% Multimodal events
Verbal WM	0.58*	0.59*	-0.31	-0.08
Visuo-sequential WM	.37	.22	.15	.26
Visuo-spatial WM	.47	.48	-.31	-.27
Executive control	0.25	0.3	-0.33	-0.39
Semantic fluency	0.01	-0.18	0.24	0.33

*** = p < 0.001, ** = p < 0.01, * = p < 0.05

Appendix G

Sums of semantic features normalized by word and gesture count

In the main manuscript, we report the effects of common ground on the number of words and gestures, as well as on the total sum of semantic features encoded in speech and in gesture. These measures are frequently used when the effects of common ground on verbal and gestural behavior are investigated (e.g., Clark & Wilkes-Gibbs, 1986; Fussell & Krauss, 1992; Galati & Brennan, 2010; 2014; Holler & Wilkin, 2009; Schubotz et al., 2019). However, even though a reduction in verbal description length and semantic features in the presence of common ground can be seen as an indicator of more efficient language use, these measures actually give little insight into how efficiently speakers truly use their words and gestures in order to communicate information. That is, these raw counts do not indicate whether the reduction in e.g. word count is proportional to the reduction in semantic features encoded verbally. If the presence of common ground indeed means that communication becomes more efficient, we might actually expect a disproportionate reduction, such that speakers encode relatively more information per word or gesture, as compared to the absence of common ground.

Therefore, here we additionally report the analyses of the semantic features encoded in speech and the semantic features encoded in gestures normalized by word and gesture count respectively. That is, we divided the total sum of features encoded in speech by the number of words, and the total sum of features encoded in gesture by the number of gestures. These normalized measures provide an index of the information density, i.e. how efficiently speakers used their words and their gestures in order to convey information.

Analyses and Results

The statistical analyses were conducted in the same fashion as described in the main manuscript.

Sum of semantic features in speech divided by word count. The sum of features contained in speech divided by the number of words (verbal information density) was predicted by personal common ground condition, such that there was a lower information density in no-CG trials (β = -.02, *SE* = .01, $t(154.80) = -3.15$, p =.002). Additionally, incremental common ground had a significant effect, such that information density was higher on later trials (β = .01, *SE* = .002, $t(154.80) = 5.35$, $p < .001$). There was no effect for age (β = -.003, *SE* = .02, *t*(31.78) = -.16, *p* = .87). There were no effects for cognitive predictors.

Sum of semantic features in gesture divided by gesture count. The sum of features contained in gesture divided by the number of gestures (gestural information density) was not predicted by any of the experimental predictors, age (β = -.08, *SE* = .11, *t*(32) = -.71, *p* = .49), personal common ground condition (β = .002, *SE* = .05, *t*(160) = .04, *p* = .97), or incremental common ground (β = .02, *SE* = .01, *t*(160) = 1.40, *p* = .16). There were no effects for cognitive predictors.

Discussion

The purpose of the additional analyses reported here was to investigate whether common ground not only affects the number of words or gestures used, or the amount of information expressed in the two modalities, but also the efficiency with which speech and gestures are used to express information. The number of semantic features expressed in speech and in gesture divided by the number of words and gestures respectively is a relational measure that gives an indication of how much information speakers actually express per word or gesture, i.e., how high the information density is.

We observed that the information density in speech was larger in the CG as compared to the no-CG condition, and increased across the experiment. This is interesting with respect to the findings reported in the main manuscript, where we observed fewer words in the CG compared to the no-CG condition, increasingly fewer words across trials, and increasingly less information across trials. The findings presented here thus suggest that participants did not only use fewer words in the CG condition, but they also used these words more efficiently, since they expressed more information per word relative to the no-CG condition. Similarly, even though both number of words and the total sum of information expressed verbally decreased across the experiment, the information density increased. Again, this indicates that participants used their fewer words more efficiently. Therefore, the present findings suggest that the presence of common ground does not only allow for a reduction in spoken utterance length and in the amount of information expressed, but also for a truly more efficient spoken language use. To the best of our knowledge, we are the first to have applied this relational measure and to have obtained this novel insight into the effects of common ground on spoken interaction. However, it must be noted that the reduction in utterance length was to some extent attributable to an omission of all but content words, which prescriptively speaking resulted in “ungrammatical” utterances. In the present task, this was possible, as an instruction like “long ones, short ones, arc, triangles” would still be comprehensible to the addressee. Therefore, it is unclear whether we would have obtained similar results in other settings that require the formulation of complete utterances.

The absence of any effect for common ground on information density in gestures suggests that regardless of personal common ground condition or accumulating incremental common ground, the amount information expressed per gesture remained constant. This is likely due to the holistic nature of gestures as compared to words. It is perfectly possible to speak of “the small triangle sitting on its base” on one trial (encoding size, shape, and orientation), and of “the triangle” on another trial (encoding only shape). However, it would be much harder if not impossible to alter one gesture, for example tracing a triangle shape (encoding size, shape, and orientation), to a gesture encoding only shape information without simultaneously also providing information with respect to the other two features.

Appendix H

Means and SDs for each dependent measure by age group and trial number

Younger adults

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6
Number of words	62.06 (27.77)	54.44 (25.42)	46.31 (18.28)	34.63 (10.52)	43.94 (14.65)	37.94 (14.08)
Number of gestures	10.31 (6.69)	8.56 (4.35)	7.13 (5.15)	5.19 (3.04)	6.81 (4.18)	5.25 (3.57)
Gestures/100 words	17.92 (11.18)	16.40 (7.21)	15.25 (7.53)	15.87 (9.63)	16.04 (9.43)	15.38 (12.06)
Speech info total	7.69 (1.70)	7.56 (1.59)	7.94 (1.73)	6.56 (1.93)	7.13 (1.78)	7.63 (1.50)
Gesture info total	5.50 (2.03)	5.59 (2.20)	4.50 (2.73)	4.19 (2.69)	4.88 (2.76)	4.25 (3.11)
% Speech unique info	37.18 (23.83)	38.26 (23.40)	48.85 (30.73)	46.65 (32.35)	44.96 (30.03)	50.21 (36.54)
% Gesture unique info	14.11 (12.24)	16.64 (14.57)	10.50 (13.03)	19.17 (14.21)	18.70 (14.28)	11.81 (12.97)
% Info speech&gesture	48.72 (17.68)	45.10 (19.95)	40.66 (25.18)	34.18 (25.06)	36.33 (22.84)	37.98 (30.47)

Older adults

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6
Number of words	53.88 (39.16)	52.25 (25.43)	47.06 (26.99)	31.94 (18.29)	45.25 (28.34)	35.81 (20.23)
Number of gestures	10.38 (11.91)	7.75 (5.85)	8.19 (6.91)	5.38 (4.92)	6.69 (5.02)	6.75 (5.39)
Gestures/100 words	19.62 (10.66)	18.41 (11.99)	18.09 (12.61)	18.55 (14.27)	18.21 (12.47)	20.61 (15.08)
Speech info total	6.94 (1.88)	6.94 (2.24)	7.00 (2.53)	5.81 (2.20)	6.19 (2.01)	6.25 (2.29)
Gesture info total	4.91 (2.94)	4.88 (3.38)	4.94 (3.23)	4.31 (3.43)	4.50 (3.19)	4.06 (2.73)
% Speech unique info	40.54 (30.73)	44.31 (35.85)	42.37 (37.62)	48.7 (40.52)	47.53 (34.51)	48.05 (33.25)
% Gesture unique info	12.35 (16.41)	19.10 (17.83)	14.61 (20.45)	21.76 (21.96)	24.34 (20.41)	20.61 (19.13)
% Info speech&gesture	47.11 (26.88)	36.60 (26.88)	43.03 (28.71)	29.54 (25.57)	28.13 (20.10)	31.34 (22.73)

Appendix I

Full model summaries of analyses reported in sections 3.3.2 and 3.3.3

All models contain random intercepts for participants and items.

Table I1. Linear mixed-effects model for the effects of age, personal common ground, and incremental common ground (trial number) on word count. Age group = young and personal common ground condition = CG[a] are on the intercept. N = 32.

	Word count			
	β	SE	t	p
Intercept	56.70	5.65	10.03	<.001
Age group$_{old}$	-2.90	6.49	-.34	.74
Condition$_{no\text{-}CG}$[a]	8.73	2.06	4.23	<.001
Trial Number	-4.15	.60	-6.87	<.001

[a] CG = common ground; no-CG = no common ground.

Table I2. Linear mixed-effects models for the effects of age, personal common ground, and incremental common ground (trial number) on gesture count and gesture rate. Age group = young and personal common ground condition = CG[a] are on the intercept. N = 32.

	Gesture count				Gesture rate			
	β	SE	t	p	β	SE	t	p
Intercept	9.34	1.35	6.92	<.001	17.00	2.56	6.64	<.001
Age group$_{old}$	.31	1.60	.20	.85	2.77	3.27	.85	.40
Condition$_{no\text{-}CG}$[a]	1.40	.58	2.42	.02	-.88	.97	-.91	.37
Trial Number	-.81	.17	-4.80	<.001	-.12	.28	-.42	.68

[a] CG = common ground; no-CG = no common ground.

Table I3. Linear mixed-effects models for the effects of age, personal common ground, and incremental common ground (trial number) on the total number of semantic features encoded in speech and the total number of semantic features encoded in gesture. Age group = young and personal common ground condition = CG[a] are on the intercept. N = 32

	Number of features in Speech				Number of features in Gesture			
	β	SE	t	p	β	SE	t	p
Intercept	7.81	.45	17.54	<.001	5.39	.68	7.95	<.001
Age group$_{old}$	-.90	.56	-1.60	.12	-.22	.90	-.24	.81
Condition$_{no\text{-}CG}$[a]	.21	.18	1.17	.25	.33	.20	1.75	.08
Trial number	-.14	.05	-2.70	.008	-.21	.06	-3.75	<.001

[a] CG = common ground; no-CG = no common ground.

Table I4. Linear mixed-effects models for the effects of age, personal common ground, and incremental common ground (trial number) on the percentage of semantic features encoded uniquely in speech and the percentage of semantic features encoded uniquely in gesture. Age group = young and personal common ground condition = CG[a] are on the intercept. N = 32.

	% features speech unique				% features gesture unique			
	β	SE	t	p	β	SE	t	p
Intercept	38.76	7.60	5.10	<.001	11.58	3.82	3.03	<.001
Age group$_{old}$	.90	10.13	.09	.93	3.64	4.69	.78	.44
Condition$_{no\text{-}CG}$[a]	-2.48	2.20	-1.13	.26	.41	1.56	.26	.79
Trial number	1.95	.64	3.04	.003	.96	.46	2.12	.04

[a] CG = common ground; no-CG = no common ground.

Table I5. Linear mixed-effects model for the effects of age, personal common ground, and incremental common ground (trial number) on the percentage of semantic features encoded twice, in speech and in gesture. Age group = young and personal common ground condition = CG[a] are on the intercept. N = 32.

	% features speech and gesture			
	β	SE	t	p
Intercept	49.67	5.67	8.75	<.001
Age group$_{old}$	-4.54	7.13	-.64	.53
Condition$_{no\text{-}CG}$[a]	2.07	2.04	1.01	.31
Trial number	-2.92	.60	-4.88	<.001

[a] CG = common ground; no-CG = no common ground.

Appendix J

Pretest

10 younger adults (4 female), aged 22 to 30, and 10 older adults (9 female), aged 62 to 78, participated in the pretest. None of the participants from the pretest participated in the main experiment.

The task participants performed was identical to the task in the main experiment. Videos were presented with audio in either clear speech, or embedded in 8-talker babble at SNRs -6, -12, -18, or -24 dB. There were two multimodal conditions (speech + visible speech, speech + visible speech + gesture). For a detailed description of the materials, see section 4.2.4 of the main paper.

We performed logistic regression analyses for both age groups separately using the function glmer from the package lme4 in the statistical software R, as described in the main paper, section 4.2.6. The contrasts between individual noise levels reported below were obtained using the package lsmeans (Lenth, 2017).

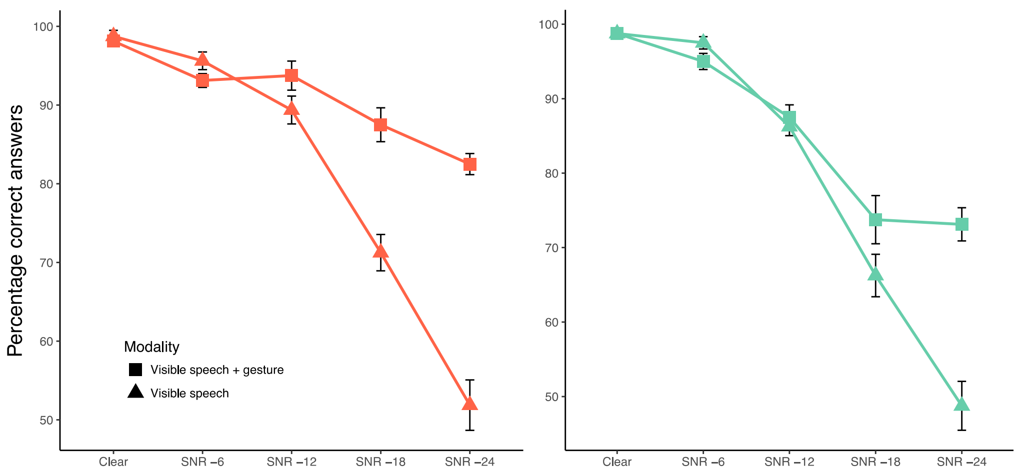


Figure J2. Response accuracies in percent per noise level and visual modality. Error bars represent SE.

Both younger and older adults performed at or above 85% accuracy in the visible speech condition in SNRs -6 and -12, with no significant performance difference between the visible speech and the visible speech + gesture conditions (younger adults: SNR -6, β = .48, *SE* = .5, *z* = .97, *p* = .33; SNR -12, β = -.58, *SE* = .42, *z* = -1.4, *p* = .16; older adults: SNR -6, β = 7.2e-01, *SE* = .62, *z* = 1.16, *p* = .25; SNR -12, β = -1.11e-01, *SE* = .33, *z* = -.33, *p* = .74; see also Figure J1). For the younger adults, as performance decreased in the visible speech condition, the added value of gestures became significant at SNRs

-18 (β = -1.05, *SE* = .3, *z* = -3.53, *p* < .001) and -24 (β = -1.5, *SE* = .26, *z* = -5.69, *p* < .0001). For older adults, performance similarly decreased with increasing noise in the visible speech condition, however, the difference between visible speech and visible speech + gesture remained non-significant at SNR -18 (β = -3.64e-01, *SE* = .25, *z* = -1.47, *p* = .14) and became significant only at SNR -24 (β = -1.07e, *SE* = .24, *z* = -4.45, *p* < .0001).

Appendix K

Response latencies

We evaluated the effects of age group, visual modality, and noise on the log transformed response latencies of the correct trials by fitting a linear mixed effects model, using the function lmer from the package lme4 in the statistical software R, as described in the main paper, section 4.2.6. The best-fitting model contained a significant three-way interaction of the predictors (the likelihood-ratio test was significant at p < .001 for comparing models with and without the three-way interaction term). In order to explore this three-way interaction further, we analyzed the response latencies of older and younger adults separately.

For the younger adults, more visual articulators led to shorter response latencies, and more severe noise led to longer response latencies (see Table K1, also Figure K1). The (non-)significant interactions of noise by modality indicate that the differences in response latencies between the three modalities were significantly larger at SNR -18 than at clear but did not differ between SNRs -18 and -24.

For the older adults, we found significant main effects of noise and of modality, but no interaction of the two predictors (see Table K1, also Figure K1). As for the younger adults, more severe noise led to longer response latencies. Response latencies were shorter in the visible speech + gesture trials than in the visible speech trials, but there was no difference between visible speech and speech-only trials.

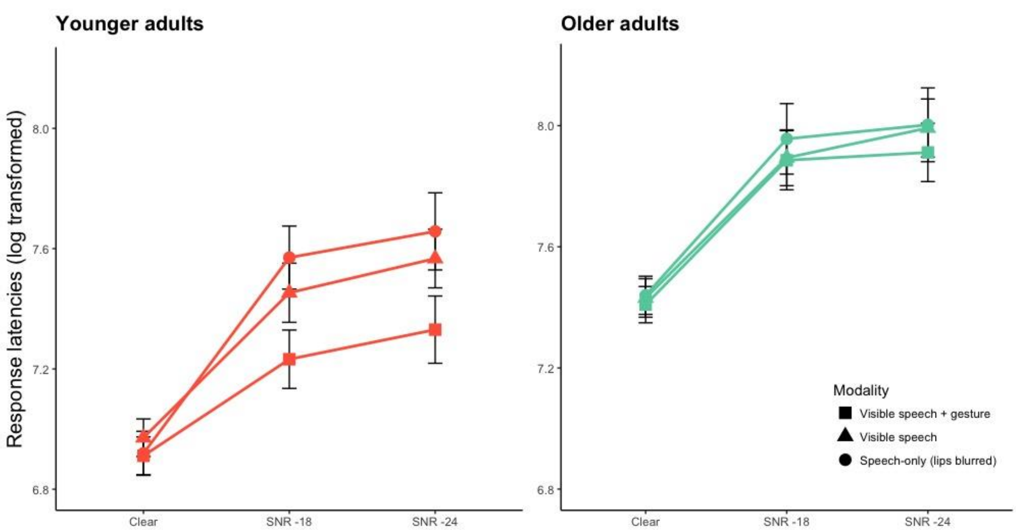


Figure K1. Log transformed response latencies per noise level and visual modality. Error bars represent SE.

Table K1. Models predicting log transformed response latencies for younger and older adults separately. Noise = SNR -18 and modality = visible speech are on the intercept.

	Younger adults				Older adults			
	β	SE	t	p	β	SE	t	p
Intercept	7.44	.05	152.64	<.001	7.92	.04	207.62	< .001
Noise$_{clear}$	-.48	.03	-16.95	< .001	-.49	.02	-28.05	< .001
Noise$_{SNR\ -24}$	.13	.03	3.86	< .001	.05	.02	2.39	.02
Visual modality$_{Speech\text{-}only\ (mouth\ blurred)}$	.11	.04	3.04	.002	.02	.02	.90	.37
Visual modality$_{Visible\ speech\ +\ gesture}$	-.21	.03	-7.20	< .001	-.04	.02	-2.29	.02
Noise$_{clear}$ *: Visual modality*$_{Speech\text{-}only\ (mouth\ blurred)}$	-.16	.04	-3.57	< .001	-	-	-	-
Noise$_{SNR\ -24}$ *: Visual modality*$_{Speech\text{-}only\ (mouth\ blurred)}$	.02	.05	-.44	.66	-	-	-	-
Noise$_{clear}$ *: Visual modality*$_{Visible\ speech\ +\ gesture}$	.15	.04	4.02	< .001	-	-	-	-
Noise$_{SNR\ -24}$ *: Visual modality*$_{Visible\ speech\ +\ gesture}$	-.04	.04	-.84	.40	-	-	-	-

[1] A hyphen indicates a non-significant predictor that was eliminated in the model-comparison process.

Appendix L

Average responses per answer type in percent (SD) per age group and condition.

		Young				Old			
Modality	**Noise**	*Target*	*Semantic competitor*	*Phonological competitor*	*Unrelated foil*	*Target*	*Semantic competitor*	*Phonological competitor*	*Unrelated foil*
Speech-only (mouth blurred)	*clear*	95.00 (4.71)	0.36 (1.31)	1.43 (3.56)	3.21 (2.44)	93.93 (5.33)	0.00 (0.00)	2.32 (4.19)	3.75 (2.59)
	SNR-18	35.54 (12.79)	20.36 (8.60)	24.46 (8.85)	19.64 (9.71)	29.11 (10.55)	24.64 (9.71)	25.54 (9.94)	20.71 (9.88)
	SNR-24	29.29 (9.79)	25.18 (10.93)	22.14 (7.38)	23.39 (10.37)	25.18 (8.22)	23.39 (10.98)	27.68 (10.14)	23.75 (10.15)
Visible speech	*clear*	97.50 (5.53)	0.18 (0.94)	2.14 (5.52)	0.18 (0.94)	97.50 (3.19)	0.36 (1.31)	2.14 (2.86)	0.00 (0.00)
	SNR-18	63.75 (16.59)	7.32 (8.33)	21.79 (12.56)	6.79 (6.27)	50.71 (18.79)	15.89 (8.50)	21.07 (9.16)	11.79 (10.65)
	SNR-24	55.36 (14.84)	12.86 (11.50)	23.39 (7.94)	8.39 (6.95)	42.32 (17.87)	18.04 (9.46)	23.39 (11.39)	15.89 (12.25)
	visual-only	51.96 (15.05)	10.18 (8.55)	24.11 (9.82)	13.75 (10.24)	48.93 (12.20)	14.46 (8.20)	23.21 (9.74)	13.39 (12.70)
Visible speech + gesture	*clear*	98.75 (2.93)	0.36 (1.31)	0.89 (2.38)	0.00 (0.00)	97.85 (3.46)	0.89 (1.95)	0.90 (1.97)	0.36 (1.31)
	SNR-18	86.09 (8.62)	8.33 (7.16)	3.41 (3.07)	2.17 (3.20)	70.77 (13.09)	21.67 (9.32)	4.14 (5.80)	3.42 (4.35)
	SNR-24	82.91 (10.46)	10.33 (6.51)	3.27 (4.42)	3.49 (4.28)	72.00 (11.71)	20.23 (9.20)	3.05 (4.16)	4.72 (4.96)
	visual-only	80.01 (18.40)	11.90 (8.08)	4.30 (6.77)	3.79 (6.90)	75.61 (12.61)	16.59 (8.47)	3.26 (4.20)	4.54 (5.06)

Nederlandse samenvatting

Menselijke communicatie is multimodaal: naast taal gebruiken we bijvoorbeeld ook gebaren, onze houding en onze blik. Met name betekenisvolle gebaren vormen een integraal onderdeel van ons taalsysteem, met vele belangrijke cognitieve en communicatieve functies: functies voor de spreker, en functies voor degene tot wie gesproken wordt, de geadresseerde. Zulke gebaren kunnen de spreker helpen bij het structureren van haar gedachten wanneer zij haar uitspraken formuleert. Daarnaast kunnen uitbeeldende gebaren, ook wel 'iconische gebaren' genoemd, gebruikt worden, om dat wat gezegd wordt beter te illustreren voor een geadresseerde. Ook kunnen gebaren helpen te begrijpen wat er gezegd wordt als dit bemoeilijkt wordt door bijvoorbeeld lawaaierige situaties.

Maar hoe veranderen de alledaagse communicatieve interacties naarmate we ouder worden? Zijn er leeftijdsgebonden verschillen in hoe we multimodaal taalgebruik hanteren? Hoe succesvol zijn oudere volwassenen in het communiceren met en begrijpen van anderen? Er zijn redenen om aan te nemen dat oudere volwassenen betekenisvolle gebaren anders gebruiken dan jongere volwassenen. Deels omdat hun gesproken taalproductie en -begrip verschilt van die van jongere volwassenen, en deels omdat ouder worden vaak gepaard gaat met veranderingen in de cognitie en het waarnemingsvermogen. Gezien de positieve functies die bij jongere volwassenen aan betekenisvolle gebaren worden toegeschreven, zou men kunnen denken dat oudere volwassenen veel baat zouden kunnen hebben bij het maken en zien van deze gebaren. Iconische gebaren zouden oudere volwassenen bijvoorbeeld kunnen helpen bij het overwinnen van moeilijkheden om de juiste woorden te vinden tijdens de taalproductie, of ze zouden belangrijke visuele informatie kunnen verschaffen wanneer taalbegrip moeilijk is door achtergrondlawaai en leeftijdsgebonden gehoorverlies. Toch kan het maken en begrijpen van iconische gebaren naast spraak ook cognitief veeleisend zijn, wat zou kunnen betekenen dat oudere volwassenen zich toch meer concentreren op spraak en geen baat hebben bij het maken of zien van extra gebaren.

Het doel van het in dit proefschrift gepresenteerde onderzoek was om te achterhalen, of en hoe het multimodale taalgebruik van oudere volwassenen verschilt van dat van jongere volwassenen en daarnaast, wat de rol is van de specifieke communicatieve context en van (leeftijdsgebonden) verschillen wat betreft cognitie en waarnemingsvermogen.

In hoofdstukken 2 en 3 van dit proefschrift heb ik onderzocht, hoe oudere volwassenen iconische gebaren gebruiken in interactie met anderen. Specifiek onderzocht ik de effecten van veroudering en van cognitieve factoren op het vermogen om spraak en iconische gebaren aan te passen op basis van wederzijds gedeelde kennis met een geadresseerde. In beide studies werkten deelnemers samen in duo's, waarbij de ene deelnemer als spreker optrad en de andere als geadresseerde. Deze duo's bestonden uit óf jongere óf oudere volwassenen. De studie die in hoofdstuk 2 wordt gepresenteerd, was een narratieve taak – hier zagen beide deelnemers samen de ene helft van een aantal korte stripverhalen, daarna kreeg alleen de spreker de volledige verhalen te zien. Hun taak was vervolgens om het volledige verhaal aan hun gesprekspartner te vertellen. Uit onderzoek met jongere volwassenen weten we, dat in dergelijke situaties, waarin de hoeveelheid wederzijds gedeelde kennis tussen spreker en geadresseerde wordt gemanipuleerd, sprekers geneigd zijn minder te spreken en minder iconische gebaren te maken, wanneer de inhoud van het verhaal al bekend is bij de geadresseerde, en meer te spreken en te gebaren wanneer de inhoud van het verhaal nieuw is voor de geadresseerde. Sprekers passen dus hun taalgebruik aan de behoefte van hun gesprekspartner aan. Dit is precies wat we vonden bij de jongere volwassenen: ze spraken minder en ze maakten minder gebaren wanneer ze vertelden over het onderdeel van het verhaal dat ze samen met hun gesprekspartner hadden gezien, en ze spraken meer en maakten meer gebaren wanneer ze vertelden over het onderdeel dat onbekend was voor hun gesprekspartner. Oudere volwassenen, daarentegen, vertoonden geen dergelijke aanpassingen. Hoewel zij gemiddeld even veel gebaren maakten als jongere volwassenen, spraken ze in beide situaties even veel en maakten daarbij tevens evenveel gebaren, ongeacht of hun gesprekspartner al bekend was met de inhoud van het verhaal of niet. Hoewel dit gedrag niet geassocieerd werd met (leeftijdsgebonden) cognitieve verschillen, moest ik op basis van deze studie concluderen dat oudere volwassenen er minder goed in zijn, hun multimodaal taalgebruik aan te passen dan jongere volwassenen.

De studie die in hoofdstuk 3 van dit proefschrift wordt gepresenteerd, betreft een ruimtelijke taak. Opnieuw manipuleerde ik de hoeveelheid gedeelde kennis tussen spreker en geadresseerde, alleen gebruikte ik deze keer eenvoudige lijntekeningen van kleine kastelen en houten bouwstenen waaruit deze kastelen konden worden samengesteld. Bij de helft van de proeven zagen beide deelnemers aan het begin kort de lijntekening, bij de andere helft van de proeven zag alleen de spreker deze. De spreker zou vervolgens achter een scherm het kasteel bouwen en daarna de geadresseerde instructies geven vanuit haar of zijn herinnering. Deze keer vond ik dat zowel jongere als oudere volwassenen hun spraak en hun iconische gebaren aanpasten, afhankelijk van of hun geadresseerde het kasteel al had gezien of niet. Beschrijvingen van beide leeftijdsgroepen waren korter en bevatten minder gebaren voor bekende kastelen, en langer en bevatten meer gebaren voor nieuwe kastelen. Bovendien werden de beschrijvingen na verloop van tijd, naarmate zowel de spreker als de geadresseerde meer vertrouwd raakten met het materiaal, over het algemeen korter en minder informatief. Verschillen in de aanpassingen konden niet toegeschreven worden aan de leeftijdsgroep, maar gedeeltelijk wel aan verschillen in de cognitieve vaardigheden van de deelnemers. Op basis van de resultaten van deze twee studies concludeerde ik dat of oudere volwassenen hun multimodale taalgebruik in dezelfde mate aanpassen als jongere volwassenen of niet, waarschijnlijk afhangt van een samenspel tussen a) de specifieke communicatieve context, d.w.z. een narratieve of ruimtelijke taak, b) de communicatieve of taakdoelen, en c) de bijbehorende cognitieve eisen.

In hoofdstuk 4 van dit proefschrift onderzocht ik de effecten van veroudering, cognitieve vaardigheden en waarnemingsvermogen op het begrijpen van spraak in rumoer en de rol die lipbewegingen en iconische gebaren hierbij spelen. Meer specifiek wilde ik weten, of oudere volwassenen naast lipbewegingen baat kunnen hebben bij het zien van iconische gebaren, wanneer ze spraak proberen te verstaan dat belast is door achtergrondlawaai. Onderzoek bij jongere volwassenen toont aan dat de visuele semantische informatie van gebaren het taalbegrip verbetert, net zoals de visuele fonologische informatie van lipbewegingen dat doet bij oudere volwassenen. Het naast spraak zien en begrijpen van gebaren kan echter ook cognitieve inspanning vergen. Hierdoor hebben oudere volwassenen wellicht minder baat bij deze extra visuele informatie dan jongere volwassenen. In het onderzoek waarover in dit hoofdstuk wordt bericht, bekeken jongere en oudere volwassenen korte filmpjes waarin een vrouw te zien was die een werkwoord uitsprak. Soms waren haar lippen te zien en soms niet; soms maakte ze een iconische handbeweging en soms niet. Het geluid in de filmpjes was óf volledig duidelijk, óf er was sprake van achtergrondlawaai, met name meerstemmig gebrabbel. De resultaten toonden aan dat beide leeftijdsgroepen significant voordeel hadden van iconische gebaren wanneer er sprake was van achtergrondlawaai – oudere volwassenen hadden echter een kleiner voordeel dan jongere volwassenen. Dit werd in ieder geval gedeeltelijk beïnvloed door een verschil in werkgeheugen. Dit suggereert dat, hoewel oudere volwassenen kunnen profiteren van multimodale communicatie, er voldoende cognitieve capaciteiten nodig zijn om dit te doen.

Het laatste hoofdstuk van dit proefschrift presenteert een samenvatting en discussie van de hiervoor genoemde studies, waarbij de nadruk ligt op de belangrijkste bijdrage van dit proefschrift: de combinatie van contextuele en cognitieve factoren in het onderzoek naar het multimodale taalgebruik van oudere volwassenen. Samenvattend kan gesteld worden dat normale veroudering bij mensen het communicatief gedrag niet op één voorspelbare manier beïnvloedt – ik vond eerder een ingewikkeld samenspel van de samenvattende variabele "leeftijd", verschillen in cognitieve capaciteiten, de specifieke kenmerken van de communicatieve situatie, de mogelijkheden van de gesproken en de visuele modaliteit (in dit geval gebaren), en met zekerheid nog vele andere nog nader te identificeren factoren.

Acknowledgements

One of the good things about growing older is that you forget about some of the hard times that you went through. The years go by and memories begin to fade. One of the bad things is that you also forget about some of the very good people who accompanied you along the way. So, if you miss your name here – please don't take it personally and blame it on my advancing age.

Two people who I shall not forget anytime soon are my two supervisors, **Asli** and **Judith**. Thank you both for your support and your guidance throughout these years – the good times and the not so good. You took me on as a PhD before I had even settled on a topic for my MA thesis. At the time, I didn't think much of it, but now I've come to see it as a sign of your trust and faith in my abilities, which you have expressed time and time again – even at times when I could not see it myself. I am truly grateful for that. I have learned so much from you both and I certainly became a better researcher and a better writer because of you. Thank you.

Thank you to my reading committee – **Antje**, **Esther**, and **Sotaro**, for taking the time to read my thesis and for your constructive criticism. Thank you also to **Alexia** and **Spencer**, who will make my examination board complete – I am very happy with this group of people!

A big Thank you to **MaxNetAging** for providing funding for the majority of the research that went into this thesis – thank you **Mirko** and **Jutta** for giving me this opportunity. Moreover, I have to thank **Peter** and the NBL for granting me a fourth year to finish this PhD – though it didn't quite work out as planned, I certainly would never have finished without this support.

Thank you to the technical group at the MPI, to **Alex** and **Peter** for their excellent PC and Mac support. A special Thank you also to **Nick**: for your help with video editing, for teaching me how to name my files, for explaining how to cut my own hair (I finally tried it – it works!), and for all the other small interactions that we had. I will not forget you.

Thank you also to all my participants, younger and older. It goes without saying: without your participation and your informed consent, this thesis would not have been possible.

The rest of these acknowledgements will be in semi-chronological order, starting back in January 2012 with the MaxNetAging cohort: **Conni**, **Tuuli**, **Elli**, **Paul** – thank you for your brief but intense friendship. Those six months in Rostock were some of the best of my life. I hope you all are happy and healthy and aging well!

Next, I really want to thank the MLC lab members past and present (where by "present" I mean "as of 2018") for their support, their useful feedback and the fun moments during lunch and at social events: **Reyhan**, **Emanuela**, **Chu**, **David**, **Beyza**,

Gerardo, **Zeynep**, **Renske**, **Kazuki**, **Linda**, **Tom**, **Hükümran**, **Dilay**, **Marlou**, **Francie**, **Susanne**, **Vicky**, **Ezgi**, **James**, and **Marlijn**. A special thanks to Marlijn for extending my MPI account again and again – indefinitely it seems.

Thank you to other MPI colleagues who became more than that: **Will** and **Elliott** for all the fun we had together. **Matthias**, for the good conversations and some dubious advice. **Peter**, thank you for your visits to my office that started it all, for countless walks out in the woods, for inspiring conversations, for agreeing to be my paranymph, and for your friendship. I still hope you'll move to Berlin one day.

A huge Thank you goes out to all the strong women in my life. Though very different individuals, what you all have in common is that you choose your path and pursue it, no matter how rocky it may be. You keep inspiring me every day and I feel honored and privileged to call you my friends. **Franziska**, I really admire you for knowing exactly what you want and then going for it. I have no doubt that you will reach your goals. **Linda**, you are exceptional in so many ways – still, what amazes me most is how a single person can be so dedicated to their work and still be so genuinely nice. **Meldy** – you're the most independent person I've ever met. I think of you often and I hope we'll see each other again soon. **Sophie**, you are one of my oldest friends. I am so grateful for our friendship throughout these years, no matter where on this planet you or I were or what we went through. Thank you for all those moments, big and small, and that I can always rely on you. **Isa**, going to the Stabi to work on this thesis was something to look forward to – because you were there! Although our personalities clash every now and then, sometimes I secretly wish I were a little more like you. **Martha**, thank you for all your support in recent years. Our friendship is very special to me, all the more, because we built it so carefully. Always stay true to yourself. **Oma Gertrud**, you also belong in this paragraph. I can't imagine what your life would have been like, had you been born later. Thank you for encouraging me to pursue all the opportunities that you never had.

The final thanks go out to the rest of my family, Dutch and German.

Barny and **Lianne**: Thank you for welcoming me into your family and for always making me feel at home.

Thank you to the people without whom none of this would have ever happened: my parents **Anne** und **Winni**. Lieber Papa – ich hätte dich so oft um Rat fragen wollen in den letzten neun Jahren. Und wusste eigentlich doch immer, was du mir geraten hättest. Danke dafür. Und trotzdem: du fehlst mir jeden einzelnen Tag. Liebe Mama – danke, dass du immer für mich da warst, dass du dich mit mir gefreut und mit mir gelitten hast und dass du bedingungslos hinter mir stehst.

And finally: Thank you **Bart**. For everything you have done for me during these past nine years. For supporting me and believing in me when I needed it most. For building a home with me in Nijmegen and then packing it all up and (almost literally) building a new home here in Berlin. For all the things we have done and seen together, all the places we have been to, all the people we have met, and all the sad and the beautiful and the very silly moments. And for designing the cover of my thesis. Now let's grow old together!

Curriculum Vitae

Louise Schubotz (*24-07-1985) was born and raised in Berlin. She obtained a bachelor's degree in Scandinavian Studies and English & American Studies from the University of Freiburg, Germany, and a research master's degree in Language & Communication from Radboud University Nijmegen. After a 6-month stay at the MPI for Demographic Research in Rostock, Germany, Louise returned to Nijmegen to do her PhD in the *Multimodal Language and Cognition Group* (back in 2012 still known as the *Gesture and Sign Language Lab*) at the MPI for Psycholinguistics.

In summer 2018, Louise moved back to Berlin, where she is currently working in the cultural sector on the development of an AI-based, interactive museum guide.

Louise lives in Schöneberg with her partner Bart and their cat Snorri.

Author publications

Schubotz, L., Özyürek, A., & Holler, J. Working memory and semantic fluency predict younger and older adults' multimodal recipient design in a spatial task (under review).

Schubotz, L., Holler, J., Drijvers, L**.,** & Özyürek, A (2020). Aging and cognitive decline modulate the integration of visible speech and iconic gestures with speech-in-noise. *Psychological Research*. doi:10.1007/s00426-020-01363-8.

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., & Özyürek, A. (2015). Age-related differences in multi-modal audience design: Young, but not old speakers, adapt speech and gestures to their addressee's knowledge. In G. Ferré, & M. Tutton (Eds.), *Proceedings of the 4th GESPIN - Gesture & Speech in Interaction Conference* (pp. 211-216). Nantes: Université of Nantes.

Schubotz, L., Oostdijk, N., & Ernestus, M. (2015). Y'know vs. you know: What phonetic reduction can tell us about pragmatic function. In S. Lestrade, P. De Swart, & L. Hogeweg (Eds.), *Addenda: Artikelen voor Ad Foolen* (pp. 361-380). Njimegen: Radboud University.

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, 692-697.

Holler, J., **Schubotz, L.**, Kelly, S., Schuetze, M., Hagoort, P., & Özyürek, A. (2013). Here's not looking at you, kid! Unaddressed recipients benefit from co-speech gestures when speech processing suffers. In M. Knauff, M. Pauen, I. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (CogSci 2013) (pp. 2560-2565). Austin, TX: Cognitive Science Society.

MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. Miranda van Turennout
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. Niels O. Schiller
3. Lexical access in the production of ellipsis and pronouns. Bernadette M. Schmitt
4. The open-/closed-class distinction in spoken-word recognition. Alette Haveman
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. Kay Behnke
6. Gesture and speech production. Jan-Peter de Ruiter
7. Comparative intonational phonology: English and German. Esther Grabe
8. Finiteness in adult and child German. Ingeborg Lasser
9. Language input for word discovery. Joost van de Weijer
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. James Essegbey
11. Producing past and plural inflections. Dirk Janssen
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. Anna Margetts
13. From speech to words. Arie van der Lugt
14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. Eva Schultze-Berndt
15. Interpreting indefinites: An experimental study of children's language comprehension. Irene Krämer
16. Language-specific listening: The case of phonetic sequences. Andrea Weber
17. Moving eyes and naming objects. Femke van der Meulen
18. Analogy in morphology: The selection of linking elements in Dutch compounds. Andrea Krott
19. Morphology in speech comprehension. Kerstin Mauth
20. Morphological families in the mental lexicon. Nivja H. de Jong
21. Fixed expressions and the production of idioms. Simone A. Sprenger

22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). Birgit Hellwig
23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. Fermín Moscoso del Prado Martín
24. Contextual influences on spoken-word processing: An electrophysiological approach. Daniëlle van den Brink
25. Perceptual relevance of prevoicing in Dutch. Petra M. van Alphen
26. Syllables in speech production: Effects of syllable preparation and syllable frequency. Joana Cholin
27. Producing complex spoken numerals for time and space. Marjolein Meeuwissen
28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. Rachèl J. J. K. Kemps
29. At the same time...: The expression of simultaneity in learner varieties. Barbara Schmiedtová
30. A grammar of Jalonke argument structure. Friederike Lüpke
31. Agrammatic comprehension: An electrophysiological approach. Marlies Wassenaar
32. The structure and use of shape-based noun classes in Miraña (North West Amazon). Frank Seifart
33. Prosodically-conditioned detail in the recognition of spoken words. Anne Pier Salverda
34. Phonetic and lexical processing in a second language. Mirjam Broersma
35. Retrieving semantic and syntactic word properties. Oliver Müller
36. Lexically-guided perceptual learning in speech processing. Frank Eisner
37. Sensitivity to detailed acoustic information in word recognition. Keren B. Shatzman
38. The relationship between spoken word production and comprehension. Rebecca Özdemir
39. Disfluency: Interrupting speech and gesture. Mandana Seyfeddinipur
40. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation. Christiane Dietrich
41. Cognitive cladistics and the relativity of spatial cognition. Daniel B.M. Haun
42. The acquisition of auditory categories. Martijn Goudbeek
43. Affix reduction in spoken Dutch. Mark Pluymaekers

44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. Valesca Kooijman
45. Space and iconicity in German Sign Language (DGS). Pamela Perniss
46. On the production of morphologically complex words with special attention to effects of frequency. Heidrun Bien
47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. Amanda Brown
48. The acquisition of verb compounding in Mandarin Chinese. Jidong Chen
49. Phoneme inventories and patterns of speech sound perception. Anita Wagner
50. Lexical processing of morphologically complex words: An information-theoretical perspective. Victor Kuperman
51. A grammar of Savosavo, a Papuan language of the Solomon Islands. Claudia Wegener
52. Prosodic structure in speech production and perception. Claudia Kuzla
53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. Sarah Schimke
54. Studies on intonation and information structure in child and adult German. Laura de Ruiter
55. Processing the fine temporal structure of spoken words. Eva Reinisch
56. Semantics and (ir)regular inflection in morphological processing. Wieke Tabak
57. Processing strongly reduced forms in casual speech. Susanne Brouwer
58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. Miriam Ellert
59. Lexical interactions in non-native speech comprehension: Evidence from electro-encephalography, eye-tracking, and functional magnetic resonance imaging. Ian FitzPatrick
60. Processing casual speech in native and non-native language. Annelie Tuinman
61. Split intransitivity in Rotokas, a Papuan language of Bougainville. Stuart Robinson
62. Evidentiality and intersubjectivity in Yurakaré: An interactional account. Sonja Gipper
63. The influence of information structure on language comprehension: A neurocognitive perspective. Lin Wang
64. The meaning and use of ideophones in Siwu. Mark Dingemanse
65. The role of acoustic detail and context in the comprehension of reduced pronunciation variants. Marco van de Ven

66. Speech reduction in spontaneous French and Spanish. Francisco Torreira
67. The relevance of early word recognition: Insights from the infant brain. Caroline Junge
68. Adjusting to different speakers: Extrinsic normalization in vowel perception. Matthias J. Sjerps
69. Structuring language. Contributions to the neurocognition of syntax. Katrien R. Segaert
70. Infants' appreciation of others' mental states in prelinguistic communication: A second person approach to mindreading. Birgit Knudsen
71. Gaze behavior in face-to-face interaction. Federico Rossano
72. Sign-spatiality in Kata Kolok: how a village sign language of Bali inscribes its signing space. Conny de Vos
73. Who is talking? Behavioural and neural evidence for norm-based coding in voice identity learning. Attila Andics
74. Lexical processing of foreign-accented speech: Rapid and flexible adaptation. Marijt Witteman
75. The use of deictic versus representational gestures in infancy. Daniel Puccini
76. Territories of knowledge in Japanese conversation. Kaoru Hayano
77. Family and neighbourhood relations in the mental lexicon: A cross-language perspective. Kimberley Mulder
78. Contributions of executive control to individual differences in word production. Zeshu Shao
79. Hearing speech and seeing speech: Perceptual adjustments in auditory-visual processing. Patrick van der Zande
80. High pitches and thick voices: The role of language in space-pitch associations. Sarah Dolscheid
81. Seeing what's next: Processing and anticipating language referring to objects. Joost Rommers
82. Mental representation and processing of reduced words in casual speech. Iris Hanique
83. The many ways listeners adapt to reductions in casual speech. Katja Poellmann
84. Contrasting opposite polarity in Germanic and Romance languages: Verum Focus and affirmative particles in native speakers and advanced L2 learners. Giuseppina Turco
85. Morphological processing in younger and older people: Evidence for flexible dual-route access. Jana Reifegerste

86. Semantic and syntactic constraints on the production of subject-verb agreement. Alma Veenstra
87. The acquisition of morphophonological alternations across languages. Helen Buckler
88. The evolutionary dynamics of motion event encoding. Annemarie Verkerk
89. Rediscovering a forgotten language. Jiyoun Choi
90. The road to native listening: Language-general perception, language-specific input. Sho Tsuji
91. Infants'understanding of communication as participants and observers. Gudmundur Bjarki Thorgrímsson
92. Information structure in Avatime. Saskia van Putten
93. Switch reference in Whitesands. Jeremy Hammond
94. Machine learning for gesture recognition from videos. Binyam Gebrekidan Gebre
95. Acquisition of spatial language by signing and speaking children: a comparison of Turkish sign language (TID) and Turkish. Beyza Sümer
96. An ear for pitch: on the effects of experience and aptitude in processing pitch in language and music. Salomi Savvatia Asaridou
97. Incrementality and Flexibility in Sentence Production. Maartje van de Velde
98. Social learning dynamics in chimpanzees: Reflections on (nonhuman) animal culture. Edwin van Leeuwen
99. The request system in Italian interaction. Giovanni Rossi
100. Timing turns in conversation: A temporal preparation account. Lilla Magyari
101. Assessing birth language memory in young adoptees. Wencui Zhou
102. A social and neurobiological approach to pointing in speech and gesture. David Peeters
103. Investigating the genetic basis of reading and language skills. Alessandro Gialluisi
104. Conversation Electrified: The Electrophysiology of Spoken Speech Act Recognition. Rósa Signý Gisladottir
105. Modelling Multimodal Language Processing. Alastair Smith
106. Predicting language in different contexts: The nature and limits of mechanisms in anticipatory language processing. Florian Hintz
107. Situational variation in non-native communication. Huib Kouwenhoven
108. Sustained attention in language production. Suzanne Jongman
109. Acoustic reduction in spoken-word processing: Distributional, syntactic, morphosyntatic, and orthographic effects. Malte Viebahn

110. Nativeness, dominance, and the flexibility of listening to spoken language. Laurence Bruggeman
111. Semantic specificity of perception verbs in Maniq. Ewelina Wnuk
112. On the identification of FOXP2 gene enhancers and their role in brain development. Martin Becker
113. Events in language and thought: The case of serial verb constructions in Avatime. Rebecca Defina
114. Deciphering common and rare genetic effects on reading ability. Amaia Carrión Castillo
115. Music and language comprehension in the brain. Richard Kunert
116. Comprehending Comprehension: Insights from neuronal oscillations on the neuronal basis of language. Nietzsche H.L. Lam
117. The biology of variation in anatomical brain asymmetries. Tulio Guadalupe
118. Language processing in a conversation context. Lotte Schoot
119. Achieving mutual understanding in Argentine Sign Language. Elizabeth Manrique
120. Talking Sense: the behavioural and neural correlates of sound symbolism. Gwilym Lockwood
121. Getting under your skin: The role of perspective and simulation of experience in narrative comprehension. Franziska Hartung
122. Sensorimotor experience in speech perception. Will Schuerman
123. Explorations of beta-band neural oscillations during language comprehension: Sentence processing and beyond. Ashley Lewis
124. Influences on the magnitude of syntactic priming. Evelien Heyselaar
125. Lapse organization in interaction. Elliott Hoey
126. The processing of reduced word pronunciation variants by natives and foreign language learners: Evidence from French casual speech. Sophie Brand
127. The neighbors will tell you what to expect: Effects of aging and predictability on language processing. Cornelia Moers
128. The role of voice and word order in incremental sentence processing. Sebastian Sauppe
129. Learning from the (un)expected: Age and individual differences in statistical learning and perceptual learning in speech. Thordis Neger
130. Mental representations of Dutch regular morphologically complex neologisms. Laura de Vaan
131. Speech production, perception, and input of simultaneous bilingual preschoolers: Evidence from voice onset time. Antje Stoehr

132. A holistic approach to understanding pre-history. Vishnupriya Kolipakam
133. Characterization of transcription factors in monogenic disorders of speech and language. Sara Busquets Estruch
134. Indirect request comprehension in different contexts. Johanne Tromp
135. Envisioning Language - An Exploration of Perceptual Processes in Language Comprehension. Markus Ostarek
136. Listening for the WHAT and the HOW: Older adults' processing of semantic and affective information in speech. Juliane Kirsch
137. Let the agents do the talking: on the influence of vocal tract anatomy on speech during ontogeny and glossogeny. Rick Janssen
138. Age and hearing loss effects on speech processing. Xaver Koch
139. Vocabulary knowledge and learning: Individual differences in adult native speakers. Nina Mainz
140. The face in face-to-face communication: Signals of understanding and non-understanding. Paul Hömke
141. Person reference and interaction in Umpila/Kuuku Ya'u narrative. Clair Hill
142. Beyond the language given: The neurobiological infrastructure for pragmatic inferencing. Jana Bašnáková
143. From Kawapanan to Shawi: Topics in language variation and change. Luis Miguel Rojas-Berscia
144. On the oscillatory dynamics underlying speech-gesture integration in clear and adverse listening conditions. Linda Drijvers
145. Understanding temporal overlap between production and comprehension. Amie Fairs
146. The role of exemplars in speech comprehension. Annika Nijveld
147. A network of interacting proteins disrupted in language-related disorders. Elliot Sollis
148. Fast speech can sound slow: Effects of contextual speech rate on word recognition. Merel Maslowski
149. Reason-giving in everyday activities. Julija Baranova
150. Speech planning in dialogue - Psycholinguistic studies of the timing of turn taking. Mathias Barthel
151. The role of neural feedback in language unification: How awareness affects combinatorial processing. Valeria Mongelli
152. Exploring social biases in language processing. Sara Iacozza

153. Vocal learning in the pale spear-nosed bat, Phyllostomus discolor. Ella Lattenkamp

154. The effect of language contact on speech and gesture: The case of Turkish-Dutch bilinguals in the Netherlands. Elif Zeynep Azar

155. Language and society: How social pressures shape grammatical structure. Limor Raviv

156. The moment in between: Planning speech while listening. Svetlana-Lito Gerakaki

157. How speaking fast is like running: Modelling control of speaking rate. Joe Rodd

158. The power of context: How linguistic contextual information shapes brain dynamics during sentence processing. René Terporten

159. Neurobiological models of sentence processing. Marvin Uhlmann

160. Individual differences in syntactic knowledge and processing: The role of literacy experience. Saoradh Favier

161. Memory for speaking and listening. Eirini Zormpa

162. Masculine generic pronouns: Investigating the processing of an unintended gender cue. Theresa Redl

163. Properties, structures and operations: Studies on language processing in the brain using computational linguistics and naturalistic stimuli. Alessandro Lopopolo

164. Investigating spoken language comprehension as perceptual inference. Greta Kaufeld

165. What was that Spanish word again? Investigations into the cognitive mechanisms underlying foreign language attrition. Anne Mickan

166. A tale of two modalities: How modality shapes language production and visual attention. Francie Manhardt

167. Why do we change how we speak? Multivariate genetic analyses of language and related traits across development and disorder. Ellen Verhoef

168. Variation in form and meaning across the Japonic language family with a focus on the Ryukyuan languages. John Huisman

169. Bilingual sentence production and code-switching: Neural network simulations. Chara Tsoukala

170. Effects of aging and cognitive abilities on multimodal language production and comprehension in context. Louise Schubotz