Hearing the message and seeing the messenger: The role of talker information in spoken language comprehension

D i s s e r t a t i o n

zur Erlangung des akademischen Grades

Doktor der Philosophie

in der Philosophischen Fakultät

der Eberhard Karls Universität Tübingen

vorgelegt von

Thanh Lan Truong, M.A.

aus

Pforzheim

2023

Gedruckt mit Genehmigung der Philosophischen Fakultät
der Eberhard Karls Universität Tübingen

Dekan: Prof. Dr. Jürgen Leonhardt

Hauptberichterstatterin: Prof. Dr. Andrea Weber
Mitberichterstatterin: Prof. Dr. Claudia Friedrich

Tag der mündlichen Prüfung: 20. Juli 2023

Universitätsbibliothek Tübingen: TOBIAS-lib

Hearing the message and seeing the messenger:
The role of talker information in spoken language comprehension

Thanh Lan Truong

To my grandparents and parents

Một đời hy sinh
Mẹ cha là cả bầu trời
Cháu con là của một đời hy sinh
Làm sao mà để chứng minh
Bao la rộng lơn sắc thinh hải hà

Acknowledgements

Being able to write these words makes me emotional and proud of my own work and determination, but this achievement would not have been possible without the support of the people I want to thank here.

I would like to start by expressing my gratitude to Professor Andrea Weber, my primary supervisor. Thank you, Andrea, for your faith in me. Although my first year kicked off well with various exciting experiments and my first big ICPhS conference in Melbourne, Australia, the majority of my PhD journey was overshadowed by the pandemic. While everything was closed down due to the pandemic, your door always remained open (both online and offline), providing me with a boost of energy whenever I needed it. Your cheerful attitude always kept me going, and your prompt feedback on my work was particularly helpful, especially toward the end of my PhD journey. I would also like to extend my thanks to my co-supervisor, Professor Claudia Friedrich. Your insightful guidance and suggestions

were greatly appreciated, especially during the early stages of my PhD journey.

I extend my heartfelt gratitude to Sara D. Beck and Simón Ruiz, both current and former colleagues and friends. Sara, I thank you for your unwavering support and guidance from the start, as well as for the enjoyable lunch breaks and delightful baked treats. Simón, I appreciate your introduction to LaTex and for guiding me through my data analyses, even when I was feeling discouraged. I have gained a lot from your expertise. Maria, I am especially grateful for your cheerful energy and the chocolate bites you always brought on most German holidays.

I also wish to express my deepest gratitude to my other colleagues in the department who were nothing but kind to me. Among those: Karin Klett, Beate Starke, Shawn Raisig, and Ann-Kathrin Grohe. Thank you for the laughter and chats.

This thesis could not have been carried out without speakers willing to lend their voices for recordings. Thank you to all speakers and I especially would like to thank Johanna Gijswijt. The studies would not have been possible without your invaluable contribution. Thank you for being so engaged and enthusiastic!

I am deeply thankful to our research assistants (present and past), Johanna Geyer, Sabrina Bucher, Eliane Leberer, Marc Stoyanopoulos, Anna-Lena Lämmle, Stacie Boswell, and Christiana Chaidaridou who played a crucial role in supporting and conducting the experiments for this dissertation.

I am deeply thankful for the opportunity to have been a part of the SFB. I appreciate the financial support provided by the SFB which allowed me to gather data on multiple occasions. I have gained a wealth of knowledge through workshops (such as programming and statistics in R) and DoKo organized by fellow PhD

students.

I would also like to express my gratitude to my friends who made my life outside of the university rich and complete, and who were always there to listen and support me throughout my PhD journey. Thank you to Toni Marcel Truong Mai, Kristina Grin, Mastan Jaza Raoof, Ida Sanders, Hannah Kontos, Hieu-Dinh Nguyen, Sandra and Maxime Courcelle, Brock Schardin, and Thien Thanh Nguyen (also those that I forgot to mention here).

I am especially grateful to my friend and partner, Swen Schreiter, who had to endure all of my PhD ups and downs. Thank you for your love, support and for encouraging me to strive for new heights and learn new things.

And I definitely would not have been here where I am now, if it were not for my grandparents and parents! Thank you, *ông bà và cha mẹ*, for your unconditional love and unwavering support. You are the four pillars without whom I cannot stand. I am proud to be your grandchild and daughter.

Contents

CHAPTER 1

General Introduction

In everyday listening, we are exposed to speech produced by talkers varying in age, gender, and language background, introducing a rich array of information and variation to the speech signal. Generally, information that is available in the speech signal is typically categorized into linguistic and indexical properties (Abercrombie, 1967). While linguistic properties convey the content of an utterance, indexical properties contain information about the talker such as the talker's language background (e.g., native or non-native), physical attributes of the talker (e.g., age, gender, health, and physiological state), and the talker's emotional state (e.g., anger, happiness, and sadness) (Abercrombie, 1967; Creel & Bregman, 2011). For example, when an adult female talker uses the word "table", the linguistic information would be something like a "piece of furniture with a level surface for eating or working at", whereas the indexical information would convey information about the talker's gender and her age, but also about her mood. When a male adult talker uses the same word "table", the linguistic information would remain the same but the indexical information would differ at least in gender information, and maybe also in age and mood, for example. Hence, realizations of the same word by different talkers will never sound exactly the same. In fact, talkers will add all kinds of variations to the speech signal. They might talk more slowly or quickly and their pronunciation might vary from the standard norms of native speech, ultimately revealing information about their social background, such as which social class they belong to, their educational history, and ethnic affiliations. The most important and often overlooked factor in delivering a talker's utterance is the medium of the voice. As Crystal puts it:

> The sound of our voice is produced by the configuration of the organs in our vocal tract. The shape of our tongue, the height of our palate, the thickness of our vocal cords, the size of our nose, the width of

> our windpipe, the contour of our lips...all of this results in a personal anatomical architecture that is unique. (Crystal & Crystal, 2014, p. 17)

Thus, the reasons for variability in speech are manifold, and the signal is rife with it. However, despite this variability, native listeners can usually interpret the speech signal with ease. How do listeners interpret such a varied speech signal correctly and effortlessly? Current theories of speech comprehension postulate that listeners map the variable signal onto pre-existing mental representations (McQueen, 2005; Weber & Scharenborg, 2012). Classically, indexical information was believed to be redundant in language comprehension. It was suggested that variability (i.e., indexical information) in the speech input was "irrelevant" information for first-pass online comprehension. Thus, indexical information was stripped away and *normalized* in order to arrive at abstract lexical representations that were needed for further linguistic analysis (Nygaard, Sommers, & Pisoni, 1994).

Recent findings from research on foreign-accented speech, for example, suggest however that indexical information might play a role in the comprehension processes quite early on. Foreign accents have been studied from different angles. By definition, non-native talkers who did not acquire their second language in their early childhood are very likely to deviate to some extent in their pronunciation from the norms of the target language (Steinlen, 2005). That is, they keep a foreign accent in their pronunciation. Psycholinguistic studies have repeatedly investigated the intelligibility of foreign-accented speech (Bradlow & Bent, 2008; Munro & Derwing, 1995), and they found that often only a brief amount of time is necessary for listeners to adjust their listening and to improve their understanding of foreign-accented speech (C. Clarke & Garrett, 2004), which also depends on the strength of foreign accent and listeners' familiarity with the accent (Witteman, Weber, & McQueen, 2013). Thus, in the scope of foreign-accented talkers, the

non-nativeness of talkers has been found to affect the comprehension process, thereby illustrating the significance of talker information in spoken language comprehension. This raises questions regarding the role of other types of speech variation like the speech of children. Similar to foreign-accented talkers, children's pronunciations also often deviate from the standard norms of native adult speech. An additional factor that comes into play is the voice, that is, the indexicality, of child talkers. Child speech is generally characterized by greater acoustic-phonetic variation than adult speech (S. Lee, Potamianos, & Narayanan, 1999), mostly due to children's distinct physical characteristics. For example, children's vocal folds are shorter in length, they have higher fundamental frequencies, and thus unsurprisingly have higher voices than adults. Thus, from the voices alone, listeners can identify the approximate age of talkers quite easily (Ptacek & Sander, 1966), and listeners appear to integrate information about the talker with the content of the message, indicating that talker information can influence listeners' interpretation of conceptual messages (Van Den Brink et al., 2010).

The processing of talker information in the speech input has recently received a rise in attention in spoken language research. More knowledge about the role of talker information is greatly needed for the development of adequate theories of spoken-language comprehension. Furthermore, studying child speech is informative for the speech perception system in general, because it can show how this system processes variation in speech, and therefore provides insights into the relevance of talker information in spoken-language comprehension. The goal of the present dissertation was to study the role of talker information and its influence on speech comprehension. More specifically, we addressed this issue from three distinct angles: We focused on talker information that is delivered through the auditory channel alone, talker information that is delivered through the audio-visual channel,

and the social impact induced by talker information.

Chapter 4 presents experiments devoted to the phonetic-to-lexical mapping of native adult and child speech. Chapters 5.1 to 5.3 are devoted to talker information in memory encoding by means of a speech intelligibility paradigm and subsequent cued-recall task. Chapter 6 is concerned with talker information for the evaluation of credibility testing native child speech as well as native and non-native adult speech. The section that follows this introduction, which is Chapter 2, contains the theoretical framework. Since the investigation of the role of talker information in spoken-language comprehension is addressed from three different angles, we present individual theoretical frameworks for each area. Finally, the research questions that are central to this dissertation, the general methodology, and the outline of the present dissertation are described in Chapter 3.

CHAPTER 2

Theoretical Foundations

This chapter describes the theoretical foundations for each study presented in Chapters 4 through 6. These theoretical foundations present a broader theoretical background for the aforementioned chapters, because each individual study focuses much more narrowly on a certain topic, a certain empirical study, and its results. To this end, this chapter is organized into three sections. The first two sections cover theories that are in the linguistic context. More specifically, the first section gives a selective description of models of spoken language comprehension and relevant empirical findings, covering the phonetic, lexical, and sentence level. While the first section provides theories covering auditory comprehension of spoken language, the second section provides theories of audiovisual speech input. Lastly, the third section presents theories covering the role of talker effect in a socio-linguistic context.

Identifying message and messenger

Spoken language contains an enormous amount of information. Not only does spoken language carry information about the message, but it also carries information about the messenger. Both information sources are simultaneously present in the acoustic signal, which makes the decoding of spoken language inherently complex, because listeners need to keep track of both sources for the interpretation of the speech stream (Abercrombie, 1967; Levi & Pisoni, 2007). In this view, speech conveys the content of an utterance (i.e., linguistic information; the message) and at the same time information about the talker who produced an utterance (i.e., indexical information; the messenger). Linguistic information entails phonological, morphological, syntactic, semantic, and pragmatic information embodied within sounds, words, and sentence structures (Levi & Pisoni, 2007). Indexical information, also known as paralinguistic or extralinguistic information, complements linguistic information with cues about, for example, the talkers' language background (e.g.,

native and non-native), characteristics of the talker (e.g., age, gender, health, and physiological state), and emotions (e.g., anger, happiness, sadness) (Abercrombie, 1967; Creel & Bregman, 2011). Thus, the speech signal contains a multidimensional array of information that exceeds the literal translation from speech signal to sounds and words. Despite the complexity, listeners can usually decode the speech signal with apparent ease, even though listeners are regularly exposed to speech that is produced by talkers varying, for example, in age, gender, and language background. The fact that individual talkers contribute significantly to the variability of the phonetic realization of linguistic categories has been known since the early days of acoustic phonetics.

In 1952, Peterson and Barney (1952) measured formant frequencies of American English vowels which had been produced by different talkers (e.g., male adult talkers, female adult talkers, and child talkers). They found, albeit not surprisingly, that vowel productions differed between talkers based on their regional and dialectal background, gender, and their age. Furthermore, research in speech articulation observed that while talkers do show similarities in their articulatory movements, there are also observable differences. For example, Johnson, Ladefoged, and Lindau (1993) found differences in jaw movements between native talkers when producing low vowels in American English, and Bordon and Gay (1979) observed that while some talkers produced the /s/ sound with the tongue tip up, others produced it with the tongue tip down.

Taken together, variation in the speech signal that is caused by differences between talkers is ubiquitous in spoken language. The question that arises from this is: What is the role of indexical information in speech comprehension? Abercrombie (1967) highlighted the importance of indexical information and said that "such 'extra-linguistic' properties of the medium, however, may fulfil other functions which

may sometimes even be more important than linguistic communication, and which can never be completely ignored" (Abercrombie, 1967, p. 5). Especially on the level of lexical processing, the potential contribution of indexical information had often been neglected in the 1970s and 1980s. Subsequently, the view on the role of linguistic and indexical information has been rather divided. On one end of the scale were so-called abstractionist theories (McClelland & Elman, 1986; Norris, 1994), and on the other end episodic theories Goldinger (1996, 1998). On the lexical level, abstractionist theories treat indexical information as "noise" and deem it as "unimportant" information for further processing. To illustrate this, Halle (2003) states that

> when we learn a new word we practically never remember most of the salient acoustic properties that must have been present in the signal that struck our ears; for example, we do not remember the voice quality of the person who taught us the word or the rate at which the word was pronounced. (Halle, 2003, p. 122)

Abstractionist theories assume that the speech input is mapped onto an abstract prelexical level of processing and subsequently matching words in the mental lexicon are identified (e.g., models of spoken-word recognition like *TRACE* McClelland & Elman, 1986; *Cohort* Marslen-Wilson & Warren, 1994, and *Shortlist* Norris, 1994; see Weber & Scharenborg, 2012 for an overview). Thus, indexical information is not assumed to be part of the listeners' mental lexicon but is rather assumed to be stripped away or *normalized* during the encoding process in order to arrive at a canonical representation for further linguistic analysis (e.g., phonemes, words, utterances) (Ladefoged & Broadbent, 1957; Nygaard & Pisoni, 1998).

The influence of abstract knowledge on word recognition has been

extensively investigated in both first (L1) and second language learning (L2) (for review, McQueen, Dahan, & Cutler, 2003). Here, L2 provides insight into abstract and prelexical processing, because unlike L1 word learning, which largely depends on heard exemplars, L2 learners can be influenced by a variety of knowledge sources other than the nature of the input. In other words, L2 learners already have a set of phonemic categories that may induce or confuse the interpretation of L2 speech. For example, confusions include the /r/-/l/ contrast of Chinese or Japanese learners of English such that they cannot distinguish between *right* and *light*. Similarly, the English /æ/-/ɜ/ contrast is equally hard for Dutch and German learners of English to differentiate. This means that two distinct phonemes collapse into a single category for L2 listeners. Nonetheless, even though L2 listeners cannot distinguish between *right* and *light* (i.e., /r/-/l/) or *kettle* and *cattle* (i.e., /æ/-/ɜ/), L2 listeners do not store them as simple homophones in their lexicon (i.e., phonetically indistinguishable which should naturally compete with each other for recognition). Studies provided evidence that listeners differentiate between those words even though they could not successfully make this distinction in their L2 acoustic-phonetic processing.

Specifically, the eye-tracking paradigm has been shown to be sensitive enough to record the spoken-word recognition process as it unfolds over time, thus highlighting the influence of abstract knowledge on L2 lexical representations. This paradigm has been proven to be extremely efficient when investigating phonetic-to-lexical mapping in L2. In an eye-tracking experiment, the gaze direction of participants' eyes are being recorded. Typically, participants are presented with a number of objects on a computer screen while voice instructions, played over headphones, are asking participants to click on one of the picture objects. In Weber and Cutler (2004), for example, Dutch listeners were shown a four-picture display: a *panda*, a *pencil* and two distractor pictures like a *duck*, and a *strawberry*.

When instructed to click on the *panda*, listeners were more likely to first look at the *pencil*. However, when asking them to click on the *pencil*, they hardly looked at the *panda*. This effect was replicated in the Japanese context as well (Cutler, Weber, & Otake, 2006). For example, Japanese listeners were asked to click on the picture of a *rocket*, but they first looked at the *locker* instead and thus experienced interference. Identical to Weber and Cutler (2004), the reversed pattern did not occur. This effect also remained robust with novel words. In Escudero, Hayes-Harb, and Mitterer (2008), Dutch listeners learned novel English bisyllabic nonwords, for example, *tenzer* and *tandik*, with *tenzer* containing /ɜ/ and *tandik* containing /æ/. Again, Dutch listeners first looked at drawings of non-objects that were associated with the auditory target word *tenzer*, when they heard the fragment *-tan*. However, listeners rarely looked at *tandik* when presented the fragment *-ten*. This asymmetry in all of the above mentioned studies suggests that L2 listeners do not store English words containing /æ/ and words containing /ɜ/ as interchangeable homophones in their lexicon even though they cannot reliably distinguish them in their phonetic processing. If they had been storing them as homophones, then confusion would have occurred in both directions. These findings indicate that abstract knowledge about the new language (i.e., L2) affects the formation of L2 lexical representations.

While the models mentioned above are mainly concerned with the lexical level, that is, word recognition, there are also models that primarily focus on speech-sound perception. In line with the models of spoken-word recognition, speech-sound perception models treat indexical information as a by-product of the articulation process which is seen as an interference that listeners need to overcome in order to process linguistic content correctly (e.g., models of speech sound perception like *Motor theory* Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985 and *Fuzzy Logic Model of Perception - FLMP*

Massaro, 1987, 1989, 1998). In this view, speech-sound perception focuses on how acoustic-phonetic cues are mapped directly onto stored mental representations, with no account of lexical factors (Klatt, 1979 but see McQueen, 2005). Overall, research on speech comprehension predominantly concentrates on the processing of linguistic information, that is, the message, rather than considering the potential contributions of indexical information, that is, the messenger.

Furthermore, research in neurolinguistics demonstrates that indexical information is processed in different parts of the brain than linguistic information even though both information sources are simultaneously transmitted in the speech signal (Winters, Levi, & Pisoni, 2008). For example, it has been found that listeners who suffer from phonagnosia can understand spoken utterances in a familiar language, but they do not have the ability to identify the voices of familiar talkers (Van Lancker, Cummings, Kreiman, & Dobkin, 1988). Another study found hemispheric specialization for indexical information but not for linguistic information (Landis, Buttet, Assal, & Graves, 1982). In addition, Stevens (2004) observed that while voice discrimination tasks activate the right frontoparietal area, lexical discrimination stimulates the left frontal and bilateral parietal areas. These observations suggest that understanding speech and recognizing the talker can operate independently of each other. Although the presented studies seemingly support the assumption that linguistic and indexical information are processed separately, behavioral studies provide evidence that both information sources can be closely intertwined during spoken language comprehension. Numerous studies have documented that speech-sound characteristics and talker characteristics are processed in parallel and interact with each other during speech perception (Bricker & Pruzansky, 1966; Goldinger, 1996; Ladefoged & Broadbent, 1957; Mullennix & Pisoni, 1990; Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008). In a

speeded-classification task by Mullennix and Pisoni (1990), participants were asked to classify a pre-specified speech sound (e.g., /b/ and /p/) in a set of words that varied in the voice of the talker and the word-initial consonants. Results showed that talker-voice variation slowed down and led to incorrect identification of speech sounds, suggesting that talker and linguistic information are both integrated during speech perception and that neither one can be selectively ignored.

Specifically, episodic theories that take into account indexical influences have been put forward as a counterargument to abstractionism on the lexical level. In contrast to abstractionist theories, episodic theories postulate that indexical information (e.g, talker-specific information) is preserved as traces in memory and is integral to later perception processes (Goldinger, 1996, 1998) (*Minerva2* Hintzman, 1986). Prelexical representations are deemed unnecessary in episodic theories. Goldinger (1998), for example, investigated with Minerva2 an episodic theory of spoken-word recognition, propelled by the fact that the speech signal is highly variable. The presumed mechanism postulates, that whenever a word is heard, it is compared directly to all stored versions of the word in the mental lexicon. Other behavioral studies have also obtained results that support the assumption that indexical and linguistic information is encoded and stored together in representations of spoken words in memory, thereby facilitating word recognition when words are produced by familiar talkers (Goldinger, Pisoni, & Logan, 1991; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Palmeri, Goldinger, & Pisoni, 1993; Schacter & Church, 1992). From the studies presented so far, it becomes clear that both abstract knowledge and episodic experience, including indexical content, are relevant for speech comprehension. In other words, speech sound perception and lexical processing can be shaped by talker information. A similar discussion about the role of linguistic and indexical information can be found on the sentence level.

Since the Chomskyan era (Chomsky, 1957), many theories assumed that the syntactic structure of a specific sentence is sufficient to derive its basic meaning with context playing little or no role in an initial analysis. The implication of this notion has led to the *standard two-step model of language interpretation* (Van Berkum et al., 2008). According to this model, the process of sentence meaning comprehension is split into two steps. In the first step, listeners attempt to figure out the context-free meaning of a sentence. In the second step, world knowledge and information about the talker are then integrated with the sentence meaning. However, the concept of context-free sentence meaning is difficult to corroborate since words like "I" and "you" automatically set up the ground for context in a communicative situation (Van Berkum et al., 2008). Theorists thus have come to realize that there is more to the meaning of an utterance in communication than just its literal interpretation. This is maybe not surprising, given that the fundamental goal of human language is not just to encode, transfer, and decode a message (Van Den Brink et al., 2010), but also to support social and interpersonal interaction (Prieur, Barbu, Blois-Heulin, & Lemasson, 2020). People use language to convey emotions and thoughts, share experiences, coordinate actions, and deepen relationships. However, language is also used to manipulate, threaten, seduce, and fool others.

For this reason, identifying speech and identifying talkers are not separate processes, they are intertwined (Clark, 1996). The *one-step model* challenges the *standard two-step model* of interpretation, because it assumes that linguistic and talker information is integrated immediately when combining the meaning of individual words into a larger whole. Research using eye tracking has shown that listeners relate specific lexical forms with speakers' characteristics (e.g., Barr & Keysar, 2006; Hanna & Tanenhaus, 2004; Hanna, Tanenhaus, & Trueswell,

2003; Metzing & Brennan, 2003; Tanenhaus & Trueswell, 2005). That is, when participants are in a dialogue with an experimenter, they connect particular lexical descriptions of items in a scenario with the dialogue partner (e.g., "the shiny cylinder" for a cylindrical object). Delayed saccade times occurred for participants when the dialogue partner unexpectedly used a different lexical description to refer to the same object (e.g., "the silver pipe"), but saccade times were not delayed when an additional new dialogue partner used that description (Metzing & Brennan, 2003). Van Berkum et al. (2008) carried this line of research further by investigating exactly when during language comprehension talker information is integrated. Van Berkum et al. (2008) examined the integration of word identity and talker identity using event-related potentials (= ERPs). Participants listened to sentences spoken by talkers varying in age and gender. What was being said either matched or did not match stereotypical talker characteristics. For example, participants listened in a mismatch condition to a child saying, "Every evening I drink some wine before I go to sleep" Such a mismatch between content and talker elicited an N400 effect (in comparison to a match condition) which is typically found when sentences contain a semantic mismatch (e.g., "He spread socks on his bread" versus "He spread butter on his bread") (Kutas & Hillyard, 1980). Findings of Van Berkum et al. (2008) showed that talkers' identity is considered as early as 200-300 ms after word onset, showing that this information is relevant for language comprehension from the earliest moment on. Following up on this work, Tesink et al. (2008) found in a functional magnetic resonance imaging study (= FMRI) overlap in brain regions involved in processing talker information, semantic, and world knowledge information, providing further evidence for a unified system of linguistic and indexical information during language comprehension.

The relevance of indexical information for spoken-language comprehension

has recently also been shown for non-native talkers. That is, listeners have been found to include knowledge about the non-nativeness of talkers during spoken language comprehension. In comparison to native speech, non-native speech, that is, speech produced by second language talkers, is even more variable, both within talkers and across talkers (Nissen, Dromey, & Wheeler, 2007; Wade, Jongman, & Sereno, 2007). Non-native talkers are often recognizable by an accent in pronunciation or by grammatical errors, as it is difficult to achieve native-like proficiency in a second language. The typical deviations from the target norms of a language are mostly due to interference from the talkers' native language (Bent & Bradlow, 2003). Such inconsistency in the speech signal can negatively affect sentence comprehension (Goslin, Duffy, & Floccia, 2012); nevertheless, listeners make use of non-natives' idiosyncrasies to adapt to them in order to understand the sentence (Hanulíková, van Alphen, van Goch, & Weber, 2012). Research has shown that listeners' knowledge about foreign-accented speech and the likelihood of deviations have modified their processing (Hanulíková & Weber, 2012). Specifically, in an ERP study native listeners were found to be more forgiving of grammatical errors produced by non-native talkers than by native talkers (Hanulíková et al., 2012), and they have been found to relax their vowel categories to accept deviating forms more willingly for non-native talkers than for native talkers (Hay, Nolan, & Drager, 2006).

The influence of social information on speech perception might be most striking in situations of phonemic mergers (Campbell-Kibler, 2010). Phonemic mergers involve sounds that previously belonged to two different categories but then over time merged into a single sound category, with talkers producing the sounds similarly which in turn makes it harder for listeners to distinguish them. In order to still distinguish those sounds, listeners can make use of what they know

about the talkers. One of the prominent examples of phonemic mergers in New Zealand English are the diphthongs of "near" (i.e., /iə/) and "square" (i.e., /eə/). As the experiment of Maclagan and Gordon (2000) progressed, the researchers noticed that the production of those two diphthongs progressively moved toward the single diphthong "near". Most striking, while younger talkers pronounced the diphthongs in "near" and "square" similarly, older talkers produced them still quite differently. The merger thus resulted in distinct pronunciation patterns for younger and older talkers. It was found that listeners utilize information about a talker's age in order to determine which diphthong was being heard. For example, Hay, Warren, and Drager (2006) manipulated this initial experiment in two ways. The speech was accompanied by different photographs of talkers, with the goal of influencing the perceived age of a talker. Particularly in New Zealand, it is known that the /iə/ and /eə/ merger most prominently occurs in younger talkers with lower social class. Hay, Warren, and Drager (2006) found that listeners who distinguished word pairs like "near-square" in their own speech were more accurate at distinguishing these word-pairs when a photograph of an older talker was shown while the speech signal was heard than when the photograph showed a younger talker. Unsurprisingly, those who did not distinguish word-pairs like "near-square" in their own speech lacked the ability to use talker information and had more difficulties distinguishing word pairs. Analogous to this study, the researchers Koops, Gentry, and Pantos (2008) have reproduced this effect in Houston Texas, where the p/iə/n - p/e/n (i.e., pin-pen) contrast is disappearing. Participants had considerable difficulties distinguishing between those two vowels when presented with a photograph of an older talker.

Phonemic mergers, however, are not the only case where social information has been known to have an effect on speech perception. In the pioneering work of Niedzielski (1999), participants made judgments about several vowels, comprising

the /au/ diphthong, which has a raised nucleus in the speech of both Canadians and Detroiters. Note that the raised segments of /aʊ/ are a commonly known stereotype of Canadian English, and Detroiters are mostly unaware of the fact that they also produce raised variants. In order to show how social information can affect speech perception, half of the respondents were told that the talker was from Detroit, the other half were told that the talker was from Canada. It was predicted that if listeners were told that the talker was from Detroit, it was less likely that they noticed the raising and report hearing a non-raised, more "standard" variant, but if the Detroit listeners were told that the Detroit talker was in fact Canadian, the listeners were more likely to notice the Canadian-raised /aʊ/. As a result, when listeners were told that the talker was Canadian, respondents tended to choose the raised vowel, but when the talker was described as coming from Detroit, respondents selected the lower more standard vowel instead. This research was again replicated in the New Zealand context and introduced two intriguing twists. In New Zealand, some speech patterns seem to be gender-split. While men tend to shift away from the Australian norm, women tend to move toward it (Hay & Drager, 2010; Warren, Hay, & Thomas, 2007). This shift even held without having the listener believe that the talker was in fact Australian (Warren et al., 2007). Indeed, simply exposing listeners to stuffed animals that were representative to the relevant country (i.e., the koala for Australia and the kiwi bird for New Zealand) sufficed to demonstrate that speech perception can by influenced by social characteristics (Hay & Drager, 2010).

In addition, listeners used information about talkers for metalinguistic judgments. Strand (1999) showed that presenting participants with a female or male photograph can affect the categorization of ambiguous stimuli between sibilants /s/ and /ʃ/ and between the back vowels in "hood" and "hud". Intriguingly, stereotypical

men and women exhibited greater effects than less stereotypical men and women. Additionally, Casasanto (2008) also showed that not only phonetic identification was sensitive to the influence of sociolinguistic knowledge but that lexical access was susceptible to that as well. They found a significant impact on listeners' perception of a string of sounds like "mæs" when listeners were presented with faces of African American or White American people. While white talkers are typically associated with less consonant cluster reduction, African American employ a stronger use of reduction. For example, the string "mæs" can either be a complete word (i.e., mass) or it can be a reduced word form which has undergone consonant cluster reduction (i.e., mast). Participants exhibited longer processing time for sentence endings which violated or mismatched the expectations. The studies presented here have shown that phonetic realizations vary depending on various factors and that listeners are able to determine a talker's age, ethnicity, and nationality based on their phonetic realizations, highlighting the close relationship between social and phonetic factors. This is particularly remarkable given the fine-grained nature of these variations (Hay & Drager, 2010).

The debate about the role of linguistic and indexical information has been long-standing and thoroughly examined on the sound, lexical, and sentence level. There have been corroborative results on all levels for both views: On the one hand, indexical information can be discarded from the speech stream and linguistic content remains the primary source for the initial analysis of the speech input. On the other hand, indexical information can complement linguistic information, thus providing listeners with detailed information which has been found to influence comprehension immediately. More recently, the trend is moving toward the direction that indexical information matters right from the beginning, that is, as soon as the speech signal hits the ear drum.

Taken together, identifying the message and the messenger has been shown to be bound to one another and the two information sources can interact during language comprehension. Listeners can integrate talker information with linguistic information from the speech signal and employ their sociolinguistic knowledge to form connections between social information and language in order to better understand speech. The linguistic brain seemingly takes into account all information available to achieve an effortless and successful comprehension of spoken language.

Hearing and seeing speech

This section presents relevant information about audiovisual speech perception, which serves as a theoretical framework for **Chapter 5.1, 5.2, and 5.3**. Speech can be seen as a multimodal phenomenon, since speech cannot only be heard but also seen (when a talker's face is visible). In the previous section, we have introduced theories of speech perception, that focus on talker information as it is delivered through the auditory channel. Besides the auditory channel (i.e., hearing the talker), talker information can also be transmitted through the visual channel (i.e., seeing the talker). In this dissertation, visual information of the talker will not entail visual features like skin color and/or facial features that can reveal a talker's ethnic background. Instead, it will entail information about a talker's visible articulatory gestures (i.e., lip and jaw movements and the general mouth region), serving as additional linguistic content that is presented along with audible cues (i.e., the sound of the talker's voice).

For many years there had been a focus on speech perception as an auditory process. The auditory signal was deemed sufficient for the interpretation of a talker's

message (e.g., as it is when being on the telephone and listening to the radio). In face-to-face communication, however, listeners are presented with information that spans across two modalities. Research has shown indeed that listeners use both visual information (lipreading) and auditory information (voice) whenever they see a talker talk (Jesse & Massaro, 2010), thus suggesting that speech comprehension is inherently multimodal. When both visual and auditory speech input is available, listeners typically gain a so-called audiovisual benefit or boost in speech perception accuracy, which results in a general improvement of speech comprehension of spoken language. This benefit applies to various types of listeners, including those with hearing difficulties and listener groups varying in age (Grant, Walden, & Seitz, 1998; Jesse, Vrignaud, Cohen, & Massaro, 2001-2002; Kaiser, Kirk, Lachs, & Pisoni, 2003; MacDonald & McGurk, 1978; MacLeod & Summerfield, 1987).

A classic demonstration of audiovisual speech perception is the McGurk effect (McGurk & Macdonald, 1976). The McGurk effect shows that visual information can change what is being perceived when visual information is in conflict with auditory information. For the original study, McGurk and his research assistant, MacDonald had asked a technician to dub the audible syllable /ba/ onto a silent video presenting a person producing /ga/. When the compiled video was presented to participants, they did not experience the expected mismatching sensation of auditory and visual input. Instead, they perceived a syllable that was neither /ba/ nor /ga/; they perceived /da/. The McGurk effect thus shows that listeners integrate both auditory and visual information automatically and in case of mismatching signals this integration can even overrule the auditory signal (see Figure 2.1).

The McGurk effect is therefore an example that demonstrates nicely how seeing the lip and jaw movements of a talker can affect our interpretation of what

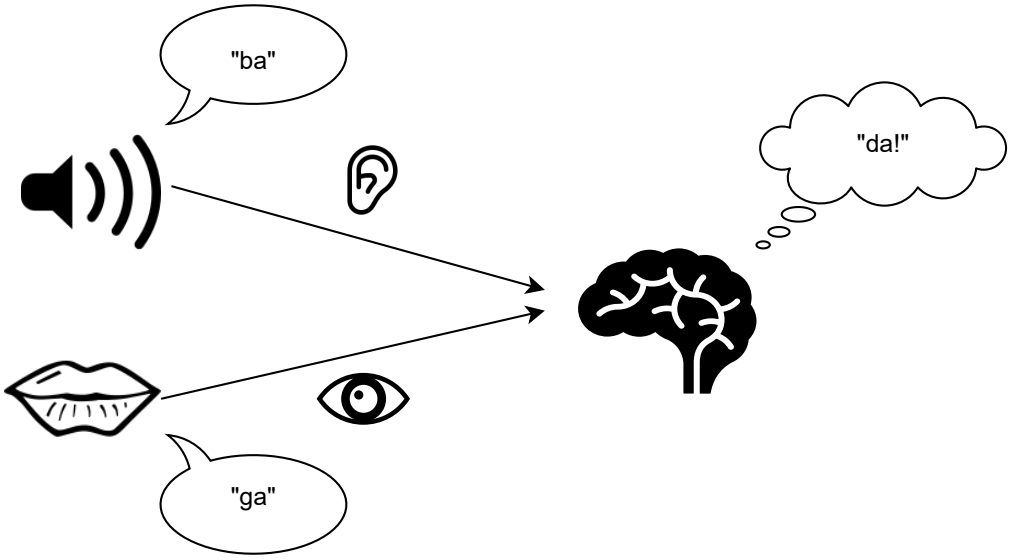

Figure 2.1: The McGurk illusion. Adapted from Lüttke (2018)

is being said. The integration of auditory and visual input furthermore happens automatically and robustly as it even occurs when participants are instructed to concentrate on only one of the two input sources (McGurk & Macdonald, 1976; Reisberg, McLean, & Goldfield, 1987; Soto-Faraco, Navarra, & Alsius, 2004).

Findings of audiovisual speech perception, like the McGurk effect, led to a shift in research. They also present a challenge for classical theoretical accounts of speech perception that were specifically developed with unimodal speech input in mind (i.e., only the auditory signal), leaving aside the important contribution of visual speech input to speech perception. For example, some of the most influential models of auditory speech perception such as TRACE (McClelland & Elman, 1986), Cohort (Marslen-Wilson & Warren, 1994), and Shortlist (Norris, 1994) do not address visible speech in their models. It is not immediately obvious how or if theories, which were developed specifically for auditory-only input, can account for the influence of visible articulatory information.

There are, however, two accounts that have been described in the previous section that can potentially explain audiovisual phenomena even though they are primarily concerned with speech-sound perception, those are: the motor theory by Liberman and Mattingly (1985) and the FLMP model by Massaro (1987). The motor theory maintains that speech perception is closely intertwined with articulatory production mechanisms. The basic assumption of the motor theory is that the speech input is based on the premise that *speech is special*.

Speech is operated by a special speech processing module and thus functions differently compared to general auditory signals (Liberman & Mattingly, 1985). For instance, listeners have a speech processing module, which is innate and uniquely human. It analyzes the speech input by creating motor representations based on how the sounds are produced. The special module, therefore, functions as a mediator for speech perception. Overall, the motor theory maintains the idea that the primary object of the speech perception function is not acoustical but gestural. According to Massaro (1987), the major challenge of the motor theory is the lack of empirical evidence that corroborates the mediation of speech perception by gestures. In addition, there is as of date no compelling evidence for articulatory gestures to be the primary object in speech perception (Massaro & Jesse, 2007).

Although the motor theory is among the most cited theories and one of the most recognized theories in fields such as cognitive psychology, it received mixed scientific responses, particularly in the domain of speech perception (Galantucci, Fowler, & Turvey, 2006). For example, the motor theory has been revised (Liberman & Mattingly, 1985), reviewed positively (Galantucci et al., 2006), and revisited negatively (Massaro & Chen, 2008). One of the most appealing theories and the most influential challenge to the motor theory, with a large body of evidence for humans' bimodal speech perception, is the FLMP model Massaro (1987). In contrast to the

motor theory, the FLMP model does not support the idea that *speech is special*. Rather, the FLMP model assumes that speech perception can also be described as a form of pattern recognition that is divided into three stages: (1) evaluation, (2) integration, and (3) decision. Incoming information from multiple sources will be evaluated individually and then integrated, unraveling what is being said (Massaro, 1989).

Although visual and auditory cues together provide more detailed information than when presenting either of the single modalities individually, this does not necessarily imply that visible speech is more informative than the auditory signal. In fact, the amount of phonemes that can be differentiated visually is lower than the amount of phonemes that can be differentiated auditorily (Van der Zande, 2013). For example, bilabial place of articulation as in stop consonants (e.g., /p/ and /b/) is visually recognizable, lip roundings as in (/i/ versus /u/), and the jaw openness that correlates with the height of vowels (e.g., more open jaw for the vowel /a/ and less open jaw for /i/). Thus, visual phonetic categories are overall harder to recognize than auditory phonetic categories (Owens & Blazek, 1985; Van Son, Huiskamp, Bosman, & Smoorenburg, 1994). Auditory and visual information presented together can be both redundant and complementary (Grant et al., 1998; Jesse & Massaro, 2010; Massaro, 1998; Sumby & Pollack, 1954). On the one hand, information from both input sources can be redundant, meaning that they convey the same information, thus contributing supplementary strength to the interpretation of the signal. On the other hand, information coming from the auditory and visual channels can also be complementary. This means that certain cues are more easily distinguishable in one modality than in the other. For example, manner of articulation (e.g., difference between /ba/ and /ma/) and voicing (e.g., difference between /ba/ and /pa/), are acoustically easier to differentiate than visually, but the

place of articulation is more informative in the visual signal (e.g., difference between /ma/ and /na/) (Massaro & Jesse, 2007). In addition, visual speech is often earlier available than auditory speech, because mouth movements often precede the sound output. For example, when producing a voiceless bilabial plosive like /p/, talkers close their lips and build up air pressure to release the plosive sound. This closure results into a silent auditory sound, which is not informative about the place of articulation. Thus, visual articulatory movements and closure of the talker's mouth contribute imperative support to the identification of place of articulation (Jesse & Massaro, 2010).

In general, the influence of visual articulatory information is most easily observed when the auditory speech signal is distorted by noise. Seeing the talker's face can improve (1) speech intelligibility (MacLeod & Summerfield, 1987; Sumby & Pollack, 1954), (2) speech detection in noise (Bernstein, Auer, & Takayanagi, 2004; Grant & Seitz, 2000), and (3) speech comprehension (Summerfield, 1992). This audiovisual speech benefit has been demonstrated for various items such as single syllables (e.g., Massaro & Cohen, 1993), words (De la Vaux, 2004; Sumby & Pollack, 1954), sentences (Jesse, Vrignaud, Cohen, & Massaro, 2000/2001; MacLeod & Summerfield, 1987), and whole sections (Reisberg et al., 1987). The benefits of audiovisual speech signals may also extend to foreign-accented speech in which audiovisual input enhances recognition of perceptual ambiguities of an unfamiliar accent compared to audio-only input (Arnold & Hill, 2001; Janse & Adank, 2012; Kawase, Hannah, & Wang, 2014; Yi, Phelps, Smiljanic, & Chandrasekaran, 2013). For example, foreign-accented speech can deviate from the standard norms of native speech, causing an ambiguous production of a phoneme or a word. Observing the lip movements of the talker may then help the listener to resolve the perceptual ambiguity and help narrow down the intended word of the talker. Hence, speech

recognition of foreign-accented speech can be improved when being face-to-face with a non-native talker (Arnold & Hill, 2001; Hazan et al., 2006; Reisberg et al., 1987). However, the visual speech signal is not only used in situations where the auditory speech signal is difficult to understand. The effect of visual speech remains robust even when listening conditions are excellent (Arnold & Hill, 2001; McGurk & Macdonald, 1976; Reisberg et al., 1987). These findings clearly show that seeing the talker enhances speech perception, but at the same time it may increase listening effort.

Pichora-Fuller et al. (2016) describe listening effort as "the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task, with listening effort applying more specifically when tasks involve listening" (Pichora-Fuller et al., 2016, p. 5S). That is, listening effort is associated with increased cognitive processing, meaning that resources or energies are utilized by listeners to fulfill cognitive demands (Peelle, 2018). However, this in turn can have downstream consequences for ongoing cognitive functions like memory encoding (Rabbitt, 1991). Memory encoding is the initial stage of the learning of information. Information coming from the sensory input is modified into a construct that can be stored and retrieved later from mental storage (Baddeley, Eysenck, & Anderson, 2020). In Rabbitt (1991), for instance, a group of listeners were presented with lists of digits in either a quiet or a noisy background. Overall, participants shadowed the digits with high accuracy. However, the group in the noisy listening condition recalled fewer digits compared to the group in the quiet listening condition. This result can be attributed to the fact that a noisy background might increase the effort that would otherwise be available for rehearsal and other processing functions that can enhance memory recall (Pichora-Fuller et al., 2016; Rabbitt, 1968). Similarly, when speech is degraded by noise, it leads to higher cognitive

demands, thereby making word identification less accurate. That is, listeners are forced to reallocate the energy from memory performance back to perception. By contrast, more cognitive resources remain available for storing information in memory when speech is easy to understand (Peelle, 2018; Pichora-Fuller et al., 2016). Multiple studies indeed found worse performance in identifying previously perceived words and recalling them in adverse listening conditions for conversational speech and for unfamiliar accents (Gilbert, Chandrasekaran, & Smiljanic, 2014; Grohe & Weber, 2018; Keerstock & Smiljanic, 2019). These findings are in line with the *effortfulness hypothesis* (McCoy et al., 2005; Rabbitt, 1968) and the Ease of Language Understanding (ELU) model (Rönnberg et al., 2013), saying that less cognitive resources remain available for storing information in memory when speech is hard to understand. This leaves open the question of how listening effort interacts with audiovisual speech input. The literature rather shows mixed results so far.

The relationship between listening effort and audiovisual speech processing has been drawing more attention in the last couple of years (Fraser, Gagné, Alepins, & Dubois, 2010; Gosselin & Gagné, 2011; Mishra, Lunner, Stenfelt, Rönnberg, & Rudner, 2013a, 2013b; Sommers & Phelps, 2016). While current theories like ELU and the effortfulness hypothesis do not tackle the question of whether audiovisual input reduces or increases listening effort, other research studies found arguments supporting both sides: First, audiovisual speech perception may provide more detailed information complementary to the auditory speech input which in turn reduces effort. Second, it can also increase listening effort, because listeners need to integrate sources from two modalities, which require additional cognitive resources (Pichora-Fuller et al., 2016). The literature shows support for both arguments. For example, Fraser et al. (2010) and Gosselin and Gagné (2011) found that audiovisual input increases listening effort, because monitoring two modalities simultaneously

and incorporating both information sources into a unified percept can demand greater cognitive load compared to when only monitoring input coming from a single source. Conversely, Sommers and Phelps (2016) obtained evidence that auditory-only speech makes listening more effortful compared to audiovisual speech. For example, seeing the talker's mouth movements in addition to hearing the talker's voice might be expected to decrease effort (e.g., Sommers & Phelps, 2016), because the visual signal provides complementary information which may limit the search for matching input onto phonetic representations in memory and in turn reduces the cognitive effort of lexical competition (Kuchinsky et al., 2013; Wagner, Toffanin, & Başkent, 2016). Furthermore, there is also evidence for comparable effort for both audiovisual and auditory-only speech input (Keidser, Best, Freeston, & Boyce, 2015). This pattern can be attributed to the fact that effort is not only affected by the type of input, but it also varies based on experimental conditions (e.g., level of difficulty by adding noise to the stimuli) (Mishra et al., 2013a). As mentioned above, audiovisual speech input may increase listening effort - even when the auditory input is highly intelligible. In adverse listening conditions, however, visual articulatory information reduces effort and improves spoken word recognition (Tye-Murray, Spehar, Myerson, Hale, & Sommers, 2016). This is in alignment with the ELU framework and the effortfulness hypothesis.

Taken together, it is now widely accepted that the addition of visual cues to auditory cues augments speech perception. Human communication most often involves face-to-face interaction in which listeners both hear and see speech. Hence, being able to see the talker's lips and jaw movements presents a distinct advantage for the listener, such as a substantial improvement in speech recognition (Jesse & Janse, 2012), even though it may involve increased listening effort (McCoy et al., 2005; Rönnberg et al., 2013). Nonetheless, audiovisual speech provides listeners with

the most absolute source of speech information, helping them to overcome obstacles in speech comprehension.

The influence of talker information on credibility judgment

This section introduces the construct of truth judgments, which serves as a theoretical background for the experiments presented in **Chapter 6**. Research presented so far showed that speech variability, ranging from stereotypes about gender, ethnicity, age, and so forth is not eliminated from the speech stream during speech processing but that information about the talker provides a multidimensional array of information that can induce implicit social evaluation like credibility judgment (Giles & Trudgill, 1983; Kinzler & DeJesus, 2013). Note that in this dissertation, the terms *credibility* and *truth* are treated as synonyms and thus used interchangeably. Higher credibility/ truth judgment (i.e., judging that something is true) is associated with the feeling of increased trustworthiness toward a person. According to Brashier and Marsh (2020), the literature differentiates between three types of inferences that can assist people in making truth judgments: (1) base rates, (2) feelings, and (3) memories (see Figure 2.2). In addition, we propose a fourth inference in this model: talker identity, which is conveyed through speech (see Figure 2.3). This inference is particularly relevant for the experiments in **Chapter 6**. Each inference will be introduced and discussed separately in more detail below.

Judging truth from base rates

When people are presented with trivia statements like "A camel's hump holds water", "Earthworms have five brains", or "Richard Feynman was a famous

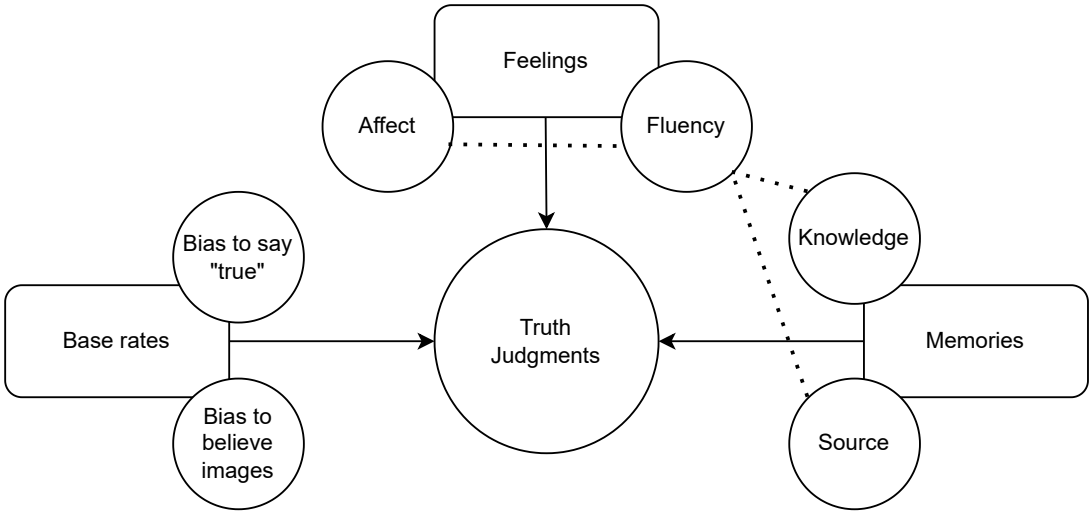

Figure 2.2: Model of truth judgments containing three inferences like base rates, feelings, and knowledge. Dotted lines indicate interactions. This figure is a reproduction from Brashier and Marsh (2020).

chemist", and they do not know for sure whether these statements are true or not, they nevertheless often believe these statements to be true, even though camel humps store fat, earthworms have five hearts, and Richard Feynman was a famous physicist (Brashier & Marsh, 2020). Why do they believe these incorrect statements? Generally speaking, judgments of truth depend on several factors like (1) how reasonable the information is, (2) how credible the source is, (3) how the talker formulates it, and (4) what the talker sounds like (e.g., G. R. Miller & Hewgill, 1964). A typical first approach to evaluating a statement's truthfulness is, however, to rely on one's initial intuition. This is often a fast-acting and effortless tactic. Since "unbelieving" requires considerably more energy than simply nodding along and agreeing to trivia statements. This is especially the case when trivia statements, which do not sound totally unreasonable, are intuitively more often judged to be true than not. This type of inference is called *base rates inference*. This fast and effortless processing stands in contrast to a second processing mechanism which is

slower and more effortful. Research in social sciences (e.g., Petty & Cacioppo, 1986) and cognitive psychology (Kahneman, 2011) have intensively discussed these two mechanisms. While the slow and effortful processing mechanism has been defined as being *analytic* and *systematic*, and is often called *System 2*, the fast and automatic processing mechanism has been referred to as being intuitive and heuristic, and is often called *System 1* (N. Schwarz & Jalbert, 2020). Considering the amount of energy that is involved utilizing in System 2, Kahneman (2011) assumes that there is a predominant preference for fast and easy processing (i.e., System 1) compared to effortful processing (i.e., System 2), because it exhausts less cognitive resources. Ideally, System 2 endorses the suggestions of System 1 with little or no alteration. This makes System 1 an important gatekeeper that can indicate whether or not it is necessary for System 2 to engage in further critical thinking (i.e., by comparing and analyzing the input with stored knowledge in memory) or to simply agree and nod along (N. Schwarz & Jalbert, 2020). Note that System 1 and 2 are always both active. While System 1 operates automatically, System 2 operates in parallel at a low-effort mode with only a small, engaged fraction of its potential. Only when System 1 detects a discrepancy, System 2 becomes more actively engaged, and measures and calculates information, as much as time and cognitive resources permit it (Kahneman, 2011). System 2 then, for example, considers the following five criteria for further assessment (N. Schwarz, 2015, pp. 211-212):

1. Is the claim compatible with what they know?
2. Is it internally consistent and coherent?
3. Does it come from a trustworthy source?
4. Do other people agree with it?
5. Is there much evidence to support it?

In summary, impressions, intuitions, intentions, and feelings are first monitored by System 1. This information is then forwarded to System 2 for approval. Impressions and intuitions become beliefs, and impulses become intended actions. Since it is less effortful to accept information in statements as true than to evaluate its wrongness, System 1 has a bias toward accepting statements as true. For example, participants exhibited a modest bias to accept new claims as true. Although participants only rarely accepted ambiguous claims as true, when they had seen them for the first time, their judgments still skewed toward "true" responses (Brashier & Marsh, 2020; Brashier, Umanath, Cabeza, & Marsh, 2017; Fazio, Brashier, Payne, & Marsh, 2015). These results illustrate nicely that people prefer to opt for the easiest and most efficient way when judging the truthfulness of statements (e.g., Kahneman, 2011; N. Schwarz, 2004).

Judging truth from feelings

An important key to intuitive assessments of truth is *processing fluency*. Processing fluency describes the ease with which stimuli are processed and it has been shown to directly affect the judgment of statements (for a review, see Oppenheimer, 2008). That is, fluently processed statements are more likely to be judged as true than statements that are difficult to understand and process, because fluently processed information evokes the feeling of ease. Processing fluency is thus typically used as a shortcut when making judgments, and it can be considered to be the scaffolding of truth judgments. This mechanism underlies the illusory truth effect (Unkelbach, 2007; Unkelbach & Stahl, 2009).

The illusory truth effect holds that if people are exposed to a statement repeatedly, they are more likely to judge the statement as true (Hasher, Goldstein, & Toppino, 1977). For instance, the misconception that Vitamin C prevents us

from getting a cold has led many people to stock up on Vitamin C supplements during cold seasons (Douglas & Hemilä, 2005). How do such misleading claims make it into our knowledge base, leading us to the false conclusion that they are true? This illustrated example is attributable to one major factor, that is familiarity. Familiarity can be achieved through constant repetition. The more familiar we are with a particular piece of information, the easier it is for us to process it. We then start to *feel* like it is *true*.

The illusory truth effect was first discussed in the seminal work of Hasher et al. (1977). Participants in their study took part in three separate experimental sessions, where they evaluated the truthfulness of statements on a 7-point scale. The researchers used plausible, yet unfamiliar statements, for the experiments. The topics of the statements ranged from history to political affairs, science, art, and geography. Participants encountered some statements repeatedly, along with other statements that had not been presented before. While some statements were correct (e.g., "Lithium is the lightest of all metals") others were wrong (e.g., "The People's Republic of China was founded in 1947"). The results showed that repeated statements received higher truth ratings compared to non-repeated statements. Most interestingly, this positive effect of repetition affected all statements, that is, statements that were true and statements that were false. The researchers concluded that if listeners are exposed to the same statements repeatedly, they can be swayed to believe that even false statements are true.

A meta-analysis of the illusory truth effect by Dechêne, Stahl, Hansen, and Wänke (2010) pointed out that statements must be in fact ambiguous (e.g., "Nut bread is healthier than potato bread") for the illusion to occur or people must feel uncertain about the statement (Fazio, Rand, & Pennycook, 2019; Unkelbach & Greifeneder, 2018). Although repetition does not provide evidence for truth,

repetition does invoke the feeling of familiarity and thus making the claim "feel" true. However, the illusory truth effect can occur in certain situations even without repetition. Sentences like "Osorno is a city in Chile" were judged more often as true when presented in a high color contrast (e.g., black print on a white background) than when presented in a low color contrast (e.g., yellow print on a white background) (Reber & Schwarz, 1999). That is, black letters on a white background are easier to read and to process than yellow letters on a white background, thus illustrating a pure processing ease effect. Also, rhyming language has been shown to influence truth judgments in the absence of familiarity. Although the phrase "Woes unite enemies" has the same meaning as "Woes unite foes", the latter was perceived as more accurate because of the rhyming of the words woes and foes (McGlone & Tofighbakhsh, 2000). Further variables that may influence processing fluency are neat handwriting, because it is easier to read (Greifeneder et al., 2010), or high video and audio quality for a video talk, since the risk of losing listeners' attention is higher when the quality of broadcasting is poor (Newman & Schwarz, 2018).

The effect of processing fluency even extends to interpersonal evaluations (Lick & Johnson, 2015), that is, who the talker of a statement is, such that trivia statements produced by a foreign-accented talker are judged to be less true than statements produced by a native talker (Lev-Ari & Keysar, 2010). In summary, processing fluency is a seemingly time-efficient and often effective determiner in deciding the truthfulness of statements, but it is not highly sensitive to the source of processing ease and can easily be swayed into inaccurate judgments.

Judging truth from knowledge base

Comparing the incoming input with existing knowledge is conceivably one of the most reliable inferences people can draw when assessing a statement's truthfulness. Information that matches the content retrieved from memory, is accepted and evaluated as true but information that mismatches is rejected and submitted for further analysis. Unlike subjective evaluations (i.e., based on feelings and preferences), objective evaluations (i.e., facts and knowledge retrieved from knowledge base) are more accurate and reliable (Campbell-Kibler, 2010). Even when knowledge is objective and accurate, the feeling of fluency can supersede this knowledge (Fazio et al., 2015; Marsh & Umanath, 2014; Unkelbach & Stahl, 2009).

Even with an accurate knowledge base, processing fluency has still been found to strongly influence judgments about truthfulness. For example, in Fazio et al. (2015) participants were presented with two statements: (1) "Ojos del Salado is the highest mountain in South America" and (2) "The Nile is the longest river in South America". Although both statements are in fact wrong, participants judged the first statement to be more true than the second statement. This pattern was attributed by the authors to the fact that people are more familiar with the Amazon than Aconcagua. In other words, knowledge does not protect against the illusory truth effect. For instance, the acceptance rate (i.e., statements rated as being true) increased when the statement "The Atlantic Ocean is the largest ocean on Earth" was presented repeatedly. Even when participants knew that the statement was wrong (i.e., the Pacific is larger), repetition led to a higher acceptance rate. Although warning participants beforehand that some claims are false, attenuated the size of the repetition effect, meaning that it did not eliminate it (Jalbert, Newman, & Schwarz, 2020).

Similarly, in Erickson and Mattson (1981), participants were informed

beforehand that some of the questions in the experiment can be deceiving and that they are free to reply with "I can't say". Even with this warning and the common knowledge that Noah, not Moses, took the animals on the Ark, participants answered the false premise "How many animals of each kind did *Moses* take on the Ark?" with "two". Unkelbach and Stahl (2009), in contrast, employed trivia statements, that are not deceiving and could theoretically be true or not, such as "Cactuses can procreate via pathogenesis". The authors assumed that knowledge would eliminate the repetition effect for these statements. Fazio et al. (2015) however, found the opposite. Participants read and rated the level of truthfulness of facts varying in public awareness (e.g., "Newton proposed the theory of relativity" and "Bell invented the wireless radio"). Afterward, they were asked to indicate if they had known the facts before the experiment. Irrespective of whether or not participants knew the facts, repetition led to more true ratings for false claims. The repetition effect remained robust in other studies for a wide range of statements covering topics like the animal kingdom, geography, facts about the USA, general science (e.g., Bacon, 1979), consumer trivia statements (e.g., Hawkins & Hoch, 1992), social-political statements (e.g., Arkes, Hackett, & Boehm, 1989). This effect even held when a delay was introduced between the repetitions, ranging from minutes to months (e.g., Begg, Anas, & Farinacci, 1992; A. S. Brown & Nix, 1996).

Taken together, knowing whether a statement is true or not does of course influence people's judgment of the correctness of the statement. However, judgments are additionally still influenced by *processing fluency* and *repetition*, not only when participants do not know when a statement is true or not, but even when they do know its truthfulness. This influence goes to the point where repetition can increase belief even in implausible claims (e.g., "The Earth is a perfect square") (Fazio et al., 2019)) or hold and spread misconceptions (e.g., "The Great Wall of China is visible

from space") (Mitchell, Gottfried, Barthel, & Sumida, 2018).

Judging truth from talker identity

The theoretical framework we presented above is strongly flavored by cognitive and psycholinguistic tradition. In this line of research, variation in speech is often seen as something that listeners need to handle for successful comprehension and which increases mental effort and in turn disrupts processing fluency. By contrast, sociolinguistic tradition sees speech variability more often not as a burden to speech comprehension but as a rich source of information, possibly even facilitating speech comprehension. In the following, we will try to bridge the two traditions of speech perception and social perception, thereby widening the scope of the theoretical framework discussed so far and contributing valuable information to the experiments in **Chapter 6**.

Although it was originally not included as a type of inference in Brashier and Marsh's (2020) model, a fourth inference has emerged from the literature in recent years. This additional element is talker information (see Figure 2.3). It particularly explores the influence of talker information on truth judgments of spoken input. Speech and voice information are conveyed in the same acoustic signal. Thus, it is conceivable that both factors may impact language understanding. For example, while speech carries information about nationality, regional dialect, social status, and educational background, the talker's voice carries information about gender, age, physical appearance, emotional state (e.g., Mack & Munson, 2012), and personality traits (e.g., Baus, McAleer, Marcoux, Belin, & Costa, 2019).

Whenever we encounter new people, we quickly form first impressions and draw social inferences about them (Baus et al., 2019). Importantly, first impressions

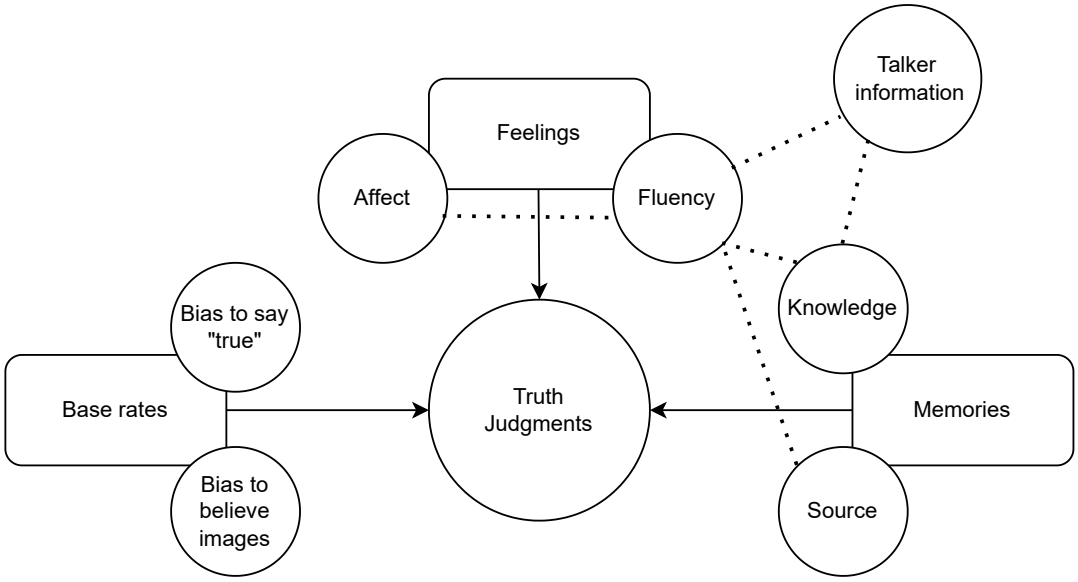

Figure 2.3: This model is a reproduction and adaption from Brashier and Marsh (2020)

are not only made when seeing a person (i.e., based on visual cues) (e.g., Willis & Todorov, 2006) but also when hearing a person speak (i.e., based on auditory cues) (e.g., Baus et al., 2019). This section focuses on the latter. McAleer, Todorov, and Belin (2014) found that people form impressions about a person even from very short bursts of speech. In their study, participants heard various talkers saying the word "Hello!" and they were asked to judge the talkers based on pre-defined personality traits (i.e., trustworthiness, dominance, attractiveness, and warmth). The results obtained showed that participants formed unifying impressions within 300 to 500 milliseconds of hearing the voices of the talkers. Most importantly, trustworthiness and dominance were the traits identified the quickest by participants. The fact that such short stretches of speech allow listeners to form impressions about whether a talker in question is trustworthy or not, indicates that talkers are being evaluated

as soon as their speech signal hits the listeners' eardrums. There is a large body of research examining the perception of distinct indexical cues which demonstrates that people generally rely heavily on voice and speech features when forming first impressions of people (De Groot & Gooty, 2009; Kramer, 1963). The study by Lev-Ari and Keysar (2010), which has been briefly mentioned earlier in this chapter, also demonstrated an effect of talker information on credibility assessment. Spoken trivia statements were evaluated as more true when the accent of the talker was easy to understand Lev-Ari and Keysar (2010), that is a native accent.

While the other studies discussed so far were primarily concerned with the fluency of written information (e.g., readability of statements), Lev-Ari and Keysar (2010) transferred the processing fluency hypothesis to spoken language. In Lev-Ari and Keysar's (2010) native English listeners judged the veracity of trivia statements like "Ants don't sleep". The chosen statements were unlikely to be known as true or false by most participants (in fact, half of them were true and half were not true), and the statements were spoken by either a native or a foreign-accented talker. Lev-Ari and Keysar (2010) found that listeners judged the statements as less true when the same statements were spoken by a foreign-accented talker than when the talker was native (Lev-Ari & Keysar, 2010).

Importantly, the researchers tried to control for negative stereotypical biases toward non-native talkers by telling participants that the statements did not mirror the talkers' own knowledge but that the talkers were merely acting as a messenger repeating statements somebody else has made. Lev-Ari and Keysar (2010) argued that their findings were driven by the fact that foreign-accented speech is harder to understand than native speech. This argument is based on the premise that accented speech is generally more difficult to understand than native speech (Cristia et al., 2012; Floccia, Butler, Goslin, & Ellis, 2009; Munro & Derwing, 1995),

consequently disrupting processing fluency and in turn affecting the credibility of the foreign-accented talker. The researchers concluded that not prejudice but rather segmental and prosodic deviations from the standard norms of the target language negatively impacted processing fluency and truth judgments (Munro & Derwing, 1995).

Although Lev-Ari and Keysar (2010) tried to control for negative stereotypes with their messenger cover story, it is quite certain that listeners noticed the talkers' non-nativeness, which possibly still triggered negative biases regardless of the cover story. The effect of foreign-accented speech on credibility ratings has been tested in different language contexts, with little to partial overlap of the experimental design (Baus et al., 2019; Frances, Costa, & Baus, 2018; Hanzlíková & Skarnitzl, 2017; Podlipsky, Simackova, & Petráž, 2016; Souza & Markman, 2013; Stocker, 2017; M. Wetzel, Zufferey, & Gygax, 2021), and the results have not reached a unifying conclusion about the source of the effect. This leaves open the question of whether the results presented in Lev-Ari and Keysar (2010) can be solely referred to processing fluency. Most intriguing, Boduch-Grabka and Lev-Ari (2021) recently replicated Lev-Ari and Keysar's (2010) findings such that processing fluency can lead individuals to trust information less when it is delivered in a foreign accent. At the same time, their findings showed that discrimination against nonnative talkers can be minimized by means of exposure to foreign accent.

Specifically, the study comprised of three phases: (1) exposure phase, (2) trivia ratings, and (3) comprehension task. Participants were split into two groups. While one group of participants was exposed to Polish-accented speech (i.e., Polish exposure condition), the other group was only exposed to a British accent (i.e., British exposure condition). Overall, findings from Boduch-Grabka and Lev-Ari (2021) showed that sentences, which were produced with a foreign accent, received

lower credibility ratings. That is, participants from the Polish exposure condition gave higher credibility ratings to trivia sentences which were produced with a Polish accent compared to the other group, thus in parallel with their previous findings (Lev-Ari & Keysar, 2010). In addition, results from the third task indicated that exposure to Polish-accented speech improved comprehension when sentences were produced with a Polish accent. In other words, results from the comprehension task indicated that participants from the Polish exposure condition were more accurate in transcribing Polish-accented sentences compared to the participants from the British exposure condition. This is in line with previous studies that showed processing difficulties with foreign-accented speech can be reduced with exposure to the accent (e.g., Bradlow & Bent, 2008; C. Clarke & Garrett, 2004).

Returning to the credibility aspect of the study, the fact that Polish-accented speech received higher credibility ratings reflects the notion that exposure can reduce bias and enhances the processing of the accent. Although Boduch-Grabka and Lev-Ari (2021) succeeded to reproduce Lev-Ari and Keysar (2010), they concluded that processing difficulty might not be the only factor that can lead individuals to trust foreign-accented speech less. Boduch-Grabka and Lev-Ari (2021) explained that their results might have been indeed caused by both prejudice and difficulty in processing fluency and that exposure simply facilitated spoken language comprehension, because even after brief exposure to a talker with deviating pronunciation from the native standard norms, correct identification of words increases (Bradlow & Bent, 2008; C. M. Clarke, 2002; Maye, Aslin, & Tanenhaus, 2008) and they recognize them more readily (C. Clarke & Garrett, 2004).

The potential core for the mixed results in the literature can be due to the fact that the interpretation of *fluency* in combination with social evaluation

is highly subjective. The term fluency can depend strongly on the individual's social status (Weick & Guinote, 2008) and motivation (Freitas, Azizian, Travers, & Berry, 2005). Thus, fluency is inherently a relative concept (for review see, Lick & Johnson, 2015). As such, fluency may have different effects on people with different backgrounds and experiences (Briñol, Petty, & Tormala, 2006). This can be illustrated by the example of the reputation of nonstandard accent in different countries. Generally, non-standard accents are foreign accents spoken by a minority or lower socioeconomic group and are thus associated with negative personal traits. For example, while the Spanish accent is considered a nonstandard accent in the United States, in the United Kingdom it has been shown to positively affect listeners' perception of talkers' educational background, social status, and personal traits like attractiveness (J. Fuertes, Gottdiener, Martin, Gilbert, & Giles, 2012). Moreover, Giles (1970) found that French-accented English received more positive evaluations than Italian or German accents, even superior to English regional accents such as the Birmingham accent.

These issues may explain the mixed results across the studies presented above. Thus, interpretation of the negative effect of foreign-accented speech on truth judgments must be proceeded with caution, but at the same time, it provides room for more interpretation in different areas. Returning to Lev-Ari and Keysar (2010), one caveat of their study is that they did not ask their participants for social evaluations of the talkers they used in their study. Thus, the elicited negative attributions caused by foreign accents might have been driven by other idiosyncracies. Interestingly though, De Meo (2012) found some indications that the results have been caused by suprasegmental deviations from native standard accent on credibility judgment. Irrespective of the strength of foreign accent, low comprehensibility did not affect the credibility of statements but prosodic

features of the spoken message had an increased influence on credibility judgments. This indicates that besides foreign accentedness (Lev-Ari & Keysar, 2010), vocal characteristics can shape credibility judgments. If acoustic characteristics of a talker, such as prosodic features, rather than comprehensibility issues as such, trigger a shift in credibility judgments, then it is possible that foreignness features shape credibility judgments after all (Lev-Ari & Keysar, 2010). Various vocal characteristics have indeed been shown to convey information about personality traits, such as charisma, persuasion, deception, leadership, and also trustworthiness.

There are numerous talker-relevant acoustic properties that listeners can use for their evaluation of a talker. These acoustic properties include formant frequencies (Baumann & Belin, 2010), which can help, for example, to identify a talker's gender (Remez, Fellowes, & Rubin, 1997). Further cues include hoarseness, vowel duration (Murry & Singh, 1980), and shimmer (Kreiman, Gerrat, K., & Berke, 1992). In addition, acoustic characteristics, such as speech rate, can carry information about personality characteristics. For instance, while slow speech rates are often judged as less competent (B. L. Smith, Brown, Strong, & Rencher, 1975) and less trustworthy (Apple, Streeter, & Krauss, 1979), faster speech rates increase judgments of persuasion (Chaiken, 1979; N. Miller, Maruyama, Beaber, & Valone, 1976), competence (R. L. Street, 1984; R. L. Street Jr. & Brady, 1982), charisma (Jiang & Pell, 2017), confidence (Hirschberg & Rosenberg, 2005), and credibility (Duller, LePoire, Aune, & Eloy, 1992). For charisma, this positive effect of speech rate is diminished, however, for very fast speech rates (Duller et al., 1992), possibly due to an excess of vowel reduction and deletion (Niebuhr, Brem, Novák-Tót, & Voße, 2016).

There is still inconsistency among researchers of what types of acoustic features influence social assessment, but the most reliable acoustic feature is

indisputably the fundamental frequency of the voice, also known as F0, or "pitch" (i.e., *highness* or *lowness* of the voice), that is, the rate of vocal fold vibrations (Fitch, 2000). Pitch is the most essential component utilized when assessing a talker's height (Xu, Lee, Wu, Liu, & Birkholz, 2013), physical strength (Sell et al., 2010), social and dominant traits (Tigue, Borak, O'Connor, Schandl, & Feinberg, 2012), and attractiveness (Feinberg, Jones, Little, Burt, & Perrett, 2005). Furthermore, low vocal pitch has been consistently found to convey dominance in many studies testing various contexts such as life-partner choices, selection of political leaders, or determining the most effective voice for business conversations (Belin, Bestelmeyer, Latinus, & Watson, 2011; Feinberg et al., 2005; Jones, Feinberg, DeBruine, Little, & Vukovic, 2010; Klofstadt, Anderson, & S., 2012; McAleer et al., 2014; Rezlescu et al., 2015; Tigue et al., 2012; Tsantani, Belin, Paterson, & McAleer, 2016).

Overall, processing fluency has been framed as a major factor that can influence credibility judgments. Since there is still considerable disagreement among researchers about the relationship between processing fluency and credibility judgment (Baus et al., 2019; Boduch-Grabka & Lev-Ari, 2021; Frances et al., 2018; Hanzlíková & Skarnitzl, 2017; Lev-Ari & Keysar, 2010; Podlipsky et al., 2016; Souza & Markman, 2013; Stocker, 2017), the impact of foreign-accented speech on credibility judgment cannot be generalized and leaves open the possibility that other factors can affect truth judgment.

Non-native accents are often seen as indicative of out-group identity and are most often, though not always, judged less favorably than in-group members. This bias can even extend to perceptions of truthfulness, with non-native talkers sometimes being perceived as less truthful than native talkers. However, prejudice and stereotypes can evoke both negative and positive feelings, depending on the individuals' background and experiences (Briñol et al., 2006; Lick & Johnson, 2015).

Particularly in spoken language, the accent is not the only factor to play a major role in social evaluations but vocal characteristics are one prominent factor which can influence speech perception as well. Thus, the introduction of the talker identity element is of great importance, because it contributes to an additional perspective of the influence of social evaluations, such as credibility judgments.

CHAPTER 3

The present dissertation

Research questions

Previous studies showed that the speech signal not only contains linguistic information but also indexical information (Abercrombie, 1967; Creel & Bregman, 2011; Levi & Pisoni, 2007). During communication, listeners must contend with the speech signal of the *message*, and at the same time, they must contend with information about the *messenger*. Therefore, *who* is talking might matter as much as *what* they are saying. The overall aim of the present dissertation is to study the role of talker information in the comprehension process of spoken language. We particularly concentrate on speech variation that the talker brings into this process. In this dissertation, speech variation specifically entails child speech and to a lesser extent non-native speech. We approached this investigation from three distinct angles:

1. Examine how talker information is processed when speech is coming from an auditory-only signal.
2. Assess the role of talker information coming from audio-visual source since talker information can also be delivered visually.
3. Investigate talker information in the socio-linguistic context; that is, whether talker information has an effect on listeners' attitude.

Experimental methods

The experiments in this thesis make use of a variety of methods conducted largely with native German listeners and partially with native English listeners. The paradigms, tasks, and methods for analyzing their results are briefly described below.

Cross-modal lexical priming

The cross-modal priming method is used in **Chapter 4** to examine the mapping of phonetic information to lexical representations in adult and child speech. Cross-modal priming is a method used to tap into online language processing. This task refers to the presentation of primes and targets across different modalities. For example, primes are presented auditorily and targets are presented visually. Swinney (1979), for instance, is particularly well known for using the cross-modal priming method (Marinis, 2018, for an overview). That is, participants are presented with a word or sentence, followed by a subsequent visual target word. Then participants are required to make a lexical decision on that target word. Typically, participants are asked to answer as quickly and as accurately as possible. Reaction times (= RTs) are measured and compared between identical, related, and unrelated conditions. Results typically show the following pattern: Participants' reaction times are shorter if the word is fully or partially matching to the auditorily presented prime word as opposed to a word that is unrelated to the prime. This pattern is known as facilitation. At the same time, partial mismatching prime-target word pairs can slower down response time, also known as inhibition (e.g., Soto-Faraco, Sebastián-Gallés, & Cutler, 2001).

In this dissertation, the methodology in **Chapter 4** is similar to Swinney's classic task such that participants listened to an auditory stimulus (i.e., the prime) followed by a visual target that can be either a word or a non-word. In contrast to Swinney (1979), primes in this dissertation were German word fragments (e.g., *Para-* form *Parasit*, "parasite") (Friedrich, Felder, Lahiri, & Eulitz, 2013). Prime and target pairs were either (1) matched partially in form (e.g., *Parasit-Parodie*, "parasite-parody"), (2) matched entirely in form (e.g., prime *Para-* form *Parasit*, target *Parodie*), or (3) mismatched completely (e.g., prime *Elo-* from *Eloquenz*,

"eloquence", target *Parodie*). Filler items were added to the experiment, to include word and nonword responses by participants. Participants indicated their lexical decisions via button press based on whether the visually presented string of letters was a real German word or not. Reaction times (RTs) were measured from visual target onset.

Cued-recall task

The cued-recall task was used to examine the impact of face masks on sentence recall in studies in **Chapter 5.1, 5.2, and 5.3**. This procedure is used for testing memory performance and participants are presented with a cue or words that aide in the process of retrieving information stored in memory (Moult, 2011). Some examples of cued recalls are the names of the categories or words that are related in meaning. For instance, the word "bird" may be used as a cue to enhance the retrieval of the word "feather". Cues act as a guide for participants on where to look in the memory, thus making information retrieval more accessible as opposed to not providing them with assistance like in a free-recall task in which no cue is provided. Indeed, Tulving and Pearlstone (1966) showed this phenomenon in their experiment. Participants in the free-recall group remembered fewer words compared to participants in the cued-recall group.

Most importantly, the lack of recall in the free-recall group may not be attributed to the fact that the items were lost in memory but that traces of the items still might have been available in memory storage yet not accessible for retrieval. Therefore, findings of Tulving and Pearlstone's pioneering work showed that retrieval cues aid memory. In this dissertation, participants were presented with video recordings of adult and child talkers producing German sentences with and without a face mask (e.g., *Die Köchin hilft montags armen Kindern,* "the cook

helps on Mondays poor children"). Then participants were presented with cues, to aid memory recall. The cue consisted of the sentence that was presented up to the adverb orthographically on the computer screen (e.g., *Die Köchin hilft montags,* "the cook helps on Mondays"), and participants were asked to type in the missing two final words (e.g., *armen Kindern,* "poor children") on their keyboard. This was done for eight blocks with six sentences each. The analysis refers to the number of correctly remembered words, also known as *memory accuracy.*

Speech intelligibility

An intelligibility task was used to investigate the intelligibility of sentences spoken by a child talker in comparison to an adult talker when the talker's mouth region was covered by a face mask or not in **Chapter 5.2**. In speech intelligibility studies, stimuli recordings are typically embedded in noise, which prevents ceiling performance for words and sentences that are high in lexical frequency (Bent & Bradlow, 2003). Munro and Derwing (1995) described intelligibility as "the extent to which an utterance is actually understood" (p. 291).

In speech intelligibility tests, native listeners, for example, transcribe sentences spoken by foreign-accented talkers. It is generally expected that native listeners perform better when listening to fellow native talkers as opposed to foreign-accented talkers. Indeed, previous research on the perception of foreign-accented speech has continuously demonstrated that native listeners find native talkers more intelligible than non-native talkers, particularly in noisy situations (Bradlow & Bent, 2008; Munro & Derwing, 1995; Smiljanic & Bradlow, 2009). Following this method, sentences, which were taken from **Chapter 5.1** (e.g., *Die Köchin hilft montags armen Kindern,* "the cook helps on Mondays poor children"), were mixed with white noise. Participants were asked to type in the

sentence they had just heard after each sentence presentation. The analysis refers to the number of correctly identified words, also known as *recognition accuracy*.

Rating

A slider scale was used to examine the effects of talker age on credibility judgments in studies in **Chapter 6**. The slider scale method is similar to the traditional Likert scale that employs radio buttons. The Likert scale typically consists of several items which participants need to choose from. The format of responses can be varied. Usually, the response alternatives consist of 5-, 7-, 10-, 11- point format, and participants are required to click on one of the options to express their own opinion. In addition to those numbers, they can also be given different wording levels such as strongly disagree, disagree, neutral, agree, and strongly agree in a 5-point format, while a 10-point format is most often numerical (Roster, Lucianetti, & Albaum, 2015). In contrast to Likert scales, slider scales make use of a continuous rating scale, meaning that they offer participants more response categories and therefore may provide more finely-grained results compared to Likert scales. Furthermore, slider scales also encourage interactive engagement, they may reduce fatigue and non-response, and overall they can create a more pleasing experience for survey takers (e.g., Cook, Heath, Thompson, & Thompson, 2001).

The study presented in **Chapter 6** adopted the experimental methodology of Lev-Ari and Keysar (2010) with the goal to reproduce and extend the scope of Lev-Ari and Keysar (2010) by testing credibility judgments in a different language setting with different talker groups. Following Lev-Ari and Keysar (2010) and keeping the scale as simple and comparable to previous studies as possible (e.g., De Meo, 2012; Hanzlíková & Skarnitzl, 2017; Stocker, 2017), a slider scale was used to measure credibility ratings of participants. Participants actively entered their

credibility ratings with the use of a sliding scale, ranging from 0 to 140 which were invisible to participants. The left end of the scale was labeled with "definitely false", and the right end was labeled with "definitely true". Participants gave their ratings for each trivia statement (e.g., *Ameisen schlafen nicht*, "ants don't sleep") by dragging and dropping the bar of the slider to the desired response position, starting from its default position at the middle of the scale.

Statistical analysis

The program R Core Team (2018) was employed for statistical analysis (versions 3.5.0. to 4.0.5). Linear mixed effects regression models (Baayen, Davidson, & Bates, 2008) were run using the lme4 package (Bates, Kliegl, Vasishth, & Baayen, 2015). Models included both fixed and random effects with random slopes and intercepts (e.g., Baayen, 2008). Mixed models account for an extensive amount of variation in the data, such as individual variation by participants or test items. In addition they can include many fixed and random effects all at once. In contrast to ANOVAs, which need separate analyses, only one analysis can be conducted with mixed models, thus reducing ambiguous interpretations (e.g., Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). Although this method has been considered a standard statistical procedure in psycholinguistics, it is undergoing constant changes, consequently leaving little consistency in conducting statistical analyses.

The current dissertation followed statistical procedures for linear mixed effects regression, which were at that time currently available. For example, the earliest publication (i.e., **Chapter 4**) applied a backward stepwise selection procedure, meaning that the mixed model was a full and complex model and at each step gradually eliminates variables from the regression model, resulting in a reduced model that best explained the data. This procedure ensured that such

models were not overparameterized as recommended by Bates, Kliegl, et al. (2015) and Matuschek et al. (2017). However, statistical analyses have gradually moved away from stepwise regression analysis. Instead of employing the stepwise selection procedure, Schroeder, Sjoquist, and Stephan (2017) state to construct models based on theory, because "without careful thought, stepwise regression analysis can turn into a fishing expedition that is void of theory" (Schroeder et al., 2017, p. 72). Adapting and moving along with the changes in statistical procedures, each chapter contains a detailed description of each individual experiment with slight variations in their statistical analyses. Possible dependent variables in the models that were calculated in the experiments are *reaction times* (lexical decision task), *recall accuracy* (cued-recall), *recognition accuracy* (speech intelligibility task), and *rating* (slider scale).

Outline

Chapter 4 investigates in two cross-modal priming studies whether adult listeners map phonetic information to word representations differently in child speech than in adult speech. To that end, German native listeners were presented with auditory German word fragments (e.g., *Para-* from *Parasit*, "parasite") that mismatched the following visual target word in the second vowel (e.g., *Parodie*, "parody"). Participants gave lexical decisions to a string of letters on the screen if the word was an existing word of German or not. Participants' reaction times were measured.

Chapters 5.1, 5.2, and 5.3 investigate the impact of face masks on higher cognitive processes in both native and non-native listeners. In particular, the experiments put forward the following question: Do native and non-native adult listeners remember words more poorly when sentences are produced by a child

talker compared to an adult talker, with and without a face mask? **Chapter 5.1** investigates the impact of face masks on German native listeners' memory performance. **Chapter 5.2** extended the scope of **Chapter 5.1** with a larger participant group and implemented a speech intelligibility task to examine whether speech with a face mask is harder to understand than speech without a face mask. For this task, white noise was embedded in the stimuli. The experiment in **Chapter 5.3** re-tests the cued-recall task paradigm with non-native adult listeners.

Chapter 6 examines whether adult listeners believe information less when they are produced by children than by adults. Four experiments test the effect of talker age (Experiment 1), gender (Experiment 2 and 3), and foreign accent (Experiment 4) on credibility ratings. Participants judge the degree of credibility for each trivia statement (e.g., Ants don't sleep) using a slider scale labeled with "definitely true" on the right end and "definitely false" on the left end.

Chapter 7 summarizes the results of each experimental chapter and provides a discussion of the main findings of this thesis. Conclusions are drawn on the basis of these findings and implications for existing theories of auditory and audiovisual comprehension of spoken language and possible lines for future research are addressed.

CHAPTER 4

Phonetic-to-lexical mapping in listening to adult and child speech

Experiment 1 of this chapter has been adapted from

Truong, T.L., Schild, U., Friedrich, C. K., and Weber, A. (2019). Phonetic-to-lexical mapping in listening to adult and child speech. In Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia* (pp. 2543-2547).

Abstract

The mapping of phonetic information to lexical representations in adult and child speech was examined using cross-modal priming. Native adult listeners were presented with German word fragments (e.g., *Para-* from *Parasit*, "parasite") that mismatched in the second vowel with a visual target word (e.g., *Parodie*, "parody"). Word fragments were spoken by a female adult talker and a 7-year-old female child. Overall, effects of talker age were elusive, but exploratory analyses in Experiment 1, pointed toward a directional asymmetry in priming for word pairs with the vowels /uː/, /iː/, and /aː/ spoken by the adult talker. This effect was however not replicated in Experiment 2. No priming was found for fragments spoken by the child talker in both Experiments. The results suggest at best a weak sensitivity to the age of the talker for phonetic-to-lexical mapping.

Introduction

Understanding spoken words involves computing, from a continuous speech signal, information that identifies matching words in the mental lexicon (e.g., McQueen, Norris, & Cutler, 1994; Zwitserlood, 1989). This process is also known as phonetic-to-lexical mapping, and in its most economical form, a mismatch between speech input and lexical representation leads to an immediate rejection of the mismatching candidate word. Thus, at the heart of spoken word recognition is the mapping of the speech signal onto matching lexical representations (Norris, 1994). Seminal models of spoken-word recognition assume that two fundamental processes are involved in word recognition: parallel activation and competition, meaning that several matching word candidates are activated in parallel and compete with each other for recognition (e.g., TRACE McClelland & Elman, 1986; Shortlist Norris, 1994). During competition, activation of word candidates that mismatch with the input will be discarded. Specifically, words with a high activation level will penalize candidates with lower activation levels. The best matching word candidate, which completely overlaps with the speech input, will be selected for recognition.

Support for theories of spoken-word recognition has been demonstrated in both priming and eye-tracking studies (e.g., Dahan & Magnuson, 2006). For instance, results of the gating task of Zwitserlood (1989) showed that when Dutch listeners perceived the word fragment *kapit-*, both words like *kapitein* (i.e., captain) and *kapitaal* (i.e., capital) were activated in the mental lexicon. However, when a further segment was attached to the fragment, thus transforming *kapit-* to *kapita-*, the word *kapitein* was no longer relevant and did not serve as a candidate word anymore (Zwitserlood, 1989). This finding was also successfully replicated with monosyllabic spoken word primes such as *buns*. Target words differing in only one

segment neither facilitated nor inhibited the word *guns* (Gaskell & Marslen-Wilson, 2001). Comparable to that, Allopenna, Magnuson, and Tanenhaus (1998) showed prompt deactivation of word pairs that are similar to each other in an eye-tracking study. Participants were presented four objects (i.e., *beaker, beetle, speaker, dolphin*). Note that *speaker* served as the rhyme competitor and *dolphin* served as the unrelated item. Participants were asked to click on one of them after listening to the instruction "Pick up the beaker". The word *beaker* was expected to activate the words that rhyme with the input word, like *speaker*, as well as words overlapping in the initial segments, such as *beetle*. Eye gaze was recorded during the experiment. When participants heard the phrase *pick up the beaker*, the eyes initially fixated on the rhyme competitor *speaker* and the partially matching word *beetle* as well as the target *beaker*, but as soon as the signal favored the word *beaker*, fixation numbers to *speaker* and *beetle* dropped quickly. However, further studies haven shown that lexical activation is not always fully parsimonious since activation of candidate words can be found despite a partial mismatch (e.g., Friedrich, Lahiri, & Eulitz, 2008; Soto-Faraco et al., 2001). Soto-Faraco et al. (2001), for example, obtained evidence that when targets mismatched the primes, they were not immediately discarded but rather inhibited. For example, the Spanish onset fragment *abun-* (from *abundancia*, "abundance") facilitated, that is speeded up, recognition of the visually presented word *abundancia*, whereas partially mismatching words like *abandano* ("abandonment") inhibited, that is slowed down, lexical decision responses. Additional support was provided by Friedrich et al. (2008) in an auditory-visual fragment priming experiment with ERP recordings. Friedrich et al. (2008) replicated Soto-Faraco et al.'s (2001) findings using German words. Event-related potentials and behavioral data were recorded during a lexical decision task. An inhibitory effect was obtained when the prime word partially matched with a target word (e.g., *Anorak-Ananas*, "anorak-pineapple"). In addition to the

behavioral data which showed delayed lexical responses to partially mismatching prime-target pairs, the ERP data suggested that partial mismatching pairs are not entirely excluded from further processing but that the processing system continued to monitor those partial mismatching pairs even after the behavioral response had been made.

Taken together, these findings corroborate the general hypothesis of models of spoken-word recognition that listeners make use of all available acoustic cues for lexical access that can assist in distinguishing between words. Motivated by Friedrich et al.'s (2013) findings, it raised the question of whether this pattern can be found for different types of talkers, such as adults and children.

By and large, models of spoken-word recognition assume that the mapping process from phonetic input to lexical representation is not sensitive to social aspects, such as to which group of individuals the talker of the input belongs to (Weber & Scharenborg, 2012). Recent research on foreign-accented speech suggests, however, that this may indeed influence spoken-word recognition. Foreign-accented talkers typically deviate from the norms of the target language in terms of their pronunciation (Steinlen, 2005). Also because foreign-accented talkers may produce grammatical errors, have an improper choice of words, and may use sentence structures that are different from that of native talkers, foreign-accented talkers may be deemed to be less "reliable" in expressing the intended message. Despite these variations, listeners can adjust their comprehension in line with the properties of the foreign-accented productions such that the same deviations are treated differently depending on the nativeness of the talker (Bosker, Quené, Sanders, & De Jong, 2014; Eisner, Melinger, & Weber, 2013; Lev-Ari, 2015). It has been argued that experience with the source properties can help to adjust the comprehension process from the outset when encountering a (new) foreign-accented talker. For example, Eisner

et al. (2013) investigated whether or not English native listeners can adapt to final devoicing in foreign-accented speech. Specifically, Dutch learners of English typically devoice voiced stop consonants in final word position. So instead of pronouncing the English word *seed* with a final voiced stop consonant /d/, Dutch learners of English typically pronounce it as [siːtʰ]. In three experiments, Eisner et al. (2013) observed that even after just limited exposure, English listeners adjusted and elucidated the devoiced word-final consonants correctly and auditory *seat* primed the visual target *seed*. This provides corroborating evidence that listeners are able to adapt to variations from L2 talkers and adjust the process of spoken-word recognition in line with linguistic evidence.

Foreign-accented talkers are not the only talkers that recurrently deviate in their pronunciation from the standard norms of a native language. Children are "unreliable" talkers too, with a lower linguistic competence than native adult talkers (e.g., Stoel-Gammon & Menn, 2005). Particularly, children's acoustic and linguistic features vary from those of native adult speech (S. Lee et al., 1999) such that their acoustic-phonetic variation is greater and overall comprises higher fundamental frequency (i.e., F0) than native adult speech (B. Smith, Sugarman, & Long, 1983; Tingley & Allen, 1975). Peterson and Barney (1952) measured vowel formant patterns of female, male, and child talkers and found considerable differences among the talker groups. For example, vowel formant frequency averages of children are about 16% higher than that of adults (Kreiman & Sidtis, 2011; but see Hillenbrand, Getty, Clark, & Wheeler, 1995 and Vorperian & Kent, 2007). Overall, children's speech characteristics are largely grounded by the distinct anatomical characteristics (e.g., smaller larynx and shorter vocal folds) and their maturation of speech motor control (e.g., speaking rate, loudness, phonation, pitch range) which gradually meets the phonetic patterns of adult speech as they grow older (Vorperian

& Kent, 2007).

The present research question, therefore, arises whether or not listeners treat variation from child speech differently from adult speech during spoken-word recognition. Listeners can recognize the approximate age of a talker quite easily (e.g., Ptacek & Sander, 1966), and age attributed to a talker has previously been found to shift listeners' perception of vowels that are currently undergoing a chain shift in a language (Drager, 2010) and to influence listeners' interpretation of conceptual messages (Van Den Brink et al., 2010).

The aim of the present study was to investigate if adult native listeners map phonetic information to lexical representations differently when listening to child speech than when listening to adult speech. In two cross-model fragment priming experiments, German listeners heard word onset fragments as primes (e.g., *Para-* from *Parasit*, "parasite") before they had to decide if visually displayed target words (e.g., *Parodie*, "parody") were existing words of German or not. Prime and target words overlapped in onset but mismatched in the vowel of the second syllable (e.g., /a:/, in *Para-* and /o:/ in *Parodie*). Prime words were either produced by a 7-year-old child or by an adult talker. If talker age influences phonetic-to-lexical mapping, then the same mismatches in vowels were predicted to result in different priming effects depending on talker age with mismatches produced by the child being penalized less than mismatches produced by the adult.

Experiment 1

Method

Participants

Thirty-one native listeners of German (21 female), all students from the University of Tübingen (18-30 years old, mean age = 23.5, SD = 3.4) participated in the experiment for monetary compensation. None of them suffered from any hearing disorders, and they all had intact or corrected vision.

Material

Fifty-six German word pairs from Friedrich et al. (2013) were used as experimental items. The full list of items can be found in Appendix A (see Chapter 7). The two words of a pair had the same stress pattern and overlapped segmentally in onset but mismatched in the vowel of the second syllable (e.g. *Parasit-Parodie*, "parasite-parody").

Across word pairs, the mismatching vowels differed in vowel height, backness, and roundedness, and represented the majority of German monophthongs (Wiese, 2000). A total of 16 different vowel mismatches were included. The onset fragments of one word of a pair (e.g., *Para-* from *Parasit*) always served as a prime for the other word (e.g., *Parodie*). Both words of a pair functioned as a fragment prime for the other word in different experimental lists (e.g., *Paro-* also served as a prime for *Parasit*). Taking stress and vowel quality into consideration, onset fragments were never existing German words and only matched up with their carrier word in German.

Since asymmetries in vowel perception (e.g., /oː/-/aː/ being less confusable

than /aː/-/oː/) have been shown to affect lexical activation (e.g., Cutler et al., 2006; Cutler, Weber, Smits, & Cooper, 2004), the direction of vowel mismatch was coded in the present experiment. For more confusable mismatches, the mismatch might be opaque and not preclude (pre-)activation of the target word, while for dissimilar vowels the mismatch might preclude target activation.

Eighty phonotactically legal nonword pairs were selected as filler items, such as *purili* and *tulomen*. In 22 pairs the two onsets overlapped but mismatched in the second vowel, in 22 pairs they were phonologically unrelated, and in 36 pairs they overlapped fully, including the second vowel. The onset fragment of one nonword of a pair served as a prime for the other nonword.

All words and nonwords were recorded by two female native talkers of Standard German who were living in Tübingen at the time of the recording: a 34-year-old adult and a 7-year-old child. Recordings were made in a sound-attenuated room with a high-quality microphone and a sampling rate of 44 kHz. While the adult talker read from orthographic transcriptions, the child was prompted with the adult recordings. Special care was taken that all items were produced as intended. Onset fragments were excised using Praat (Boersma, P., Weenink, D., 2018). The durations of the onset fragments were on average longer in the child voice than in the adult voice (mean child voice = 604 ms; mean adult voice = 555 ms; t = -2.7, $p < 0.008$).

Procedure

The experiment was carried out with Presentation (version 20.1, www.neurobs.com). Before the experiment started, participants signed written informed consent. Participants were seated comfortably in front of a computer screen and wore over-ear headphones (Sennheiser HD 215 II) and were tested individually.

Each trial started with a white fixation cross on a black background with a font size of 40 in the center of the screen. Three hundred milliseconds after its onset, a prime fragment was presented over headphones at a comfortable sound pressure level. The cross was substituted by a target word, a string of letters, with a font size of 25 and uppercase letters. Target words remained on the screen for 300 milliseconds.

Participants were instructed that they would hear a word directly followed by a visual target word. They were asked to indicate whether the string of letters was an existing German word or not. Decisions were indicated by pressing a green button with their dominant hand for "yes" and a red button with the other hand for "no". The subsequent trial began after 2500 milliseconds after a response was given. If they had not pressed any button, the following trial began automatically after 4000 milliseconds after the onset of the target word.

In the related condition, the target word was preceded by the spoken onset fragment of its pair member (e.g., prime *Para-* from *Parasit*, and target *Parodie*). Both pair members served as a target and a prime in a Latin-Square design (e.g., prime *Paro-* from *Parodie*, target *Parasit*). In the unrelated condition, the target word was preceded by the spoken onset fragment of a segmentally unrelated word (e.g., prime *Elo-* from *Eloquenz*, "eloquence", target *Parodie*). All primes used in unrelated trials also served as primes in the related condition (e.g., prime *Elo-*, target *Element*, "element"). Eight experimental lists with the 56 experimental items and the 80 filler items were created.

Each experimental item appeared once in each list, counterbalanced for the role of the target, the relatedness of the prime, and the talker of the prime. The order of item presentation was pseudo-randomized. After the priming task was

completed, participants filled in a short language background questionnaire.

Results

Only trials with correct responses to target words were analyzed (see Figure 4.1). Participants answered on average 84.4% correctly when the primes had been produced by the adult, and 84.0% correctly when the primes had been produced by the child. Thus, neither the task nor the different talkers posed considerable difficulties and performance did not differ for the two talkers.

Reaction times (RTs) faster than 250 ms and slower than 1200 ms were excluded from the analysis since they would not be indicative of the online process of spoken-word recognition (0.1% of the data). R (R Core Team 2018, version 3.5.0) and lme4 (Bates, Maechler, Bolker, & Walker, 2015) were used to perform linear mixed effects analyses on log-normalized RTs.

The full model included *relatedness* (related, unrelated) and *talker age* (adult, child), as well as *direction of vowel mismatch*, *lexical frequency* of the target, and target *word length* as fixed factors. *Participants* and *items* were included as random factors with random slopes. A backward stepwise selection was applied when no model improvement was observed (Bates, Kliegl, et al., 2015). After stepwise selection, the final model showed a facilitatory effect of *relatedness* (b = -0.05, SE = 0.02, t = -2.75, p <.006), faster RTs for primes for the child *talker* (b = -0.04, SE = 0.02, t = -2.44, $p < .02$), an effect of *lexical frequency* (b = -0.02, SE = 0.01, t = -2.7, $p < .007$), and marginal interactions between *direction* and *talker age* (b = 0.05, SE = 0.03, t = 2.36, $p < .02$), between *direction* and *relatedness* (b = 0.06, SE = 0.03, t = 2.37, $p < .02$), and between *talker age, direction*, and *relatedness* (b = -0.04, SE = -0.04, t = -1.04, $p < .3$). Values of the final lmer model are shown in Table 4.1.

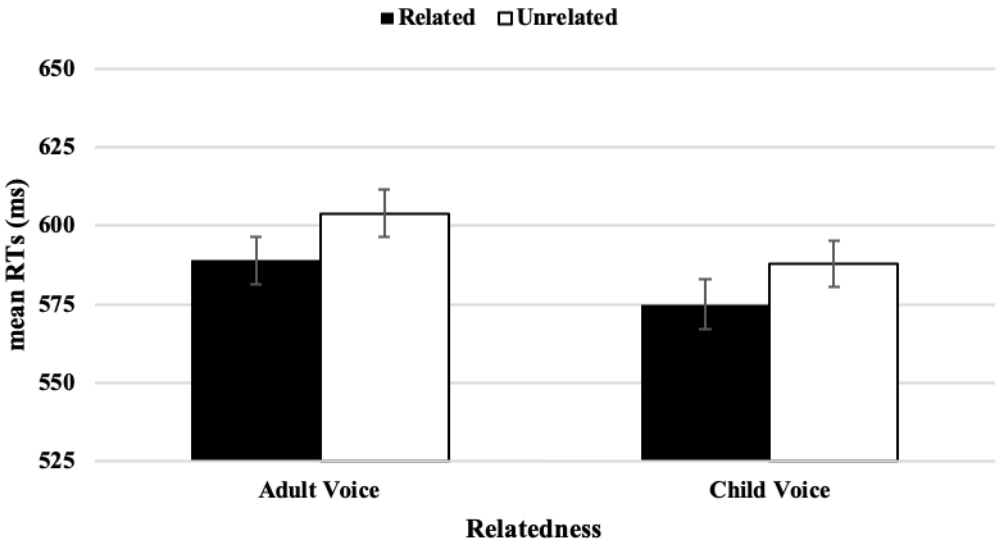

Figure 4.1: Mean RTs (in ms) following related and unrelated primes, presented in adult and child voice. The vertical bars represent standard errors.

Table 4.1: Final model output. Estimates for the best fitting model for the reaction times.

Fixed effects	*Estimate*	*Std. Error*	*t*	*p*	
Intercept	6.393e+00	3.056e-02	209.186	<2e-16	***
Relatedness	-4.986e-02	1.812e-02	-2.752	0.00600	**
Talker	-4.417e-02	1.809e-02	-2.442	0.01473	*
Direction of vowel mismatch	-2.440e-02	2.258e-02	-1.081	0.28046	
Lexical frequency	-2.508e-02	9.224e-03	-2.720	0.00667	**
Relatedness x Talker Age	1.978e-02	2.569e-02	0.770	0.44134	
Relatedness x Direction of vowel mismatch	6.065e-02	2.560e-02	2.369	0.01796	*
Talker Age x Direction of vowel mismatch	4.719e-02	2.548e-02	1.852	0.06425	.
TalkerAge x Direction x Lexical Frequency	-3.787e-02	3.627e-02	-1.044	0.29663	

*Note: *p.05 **.01 *** p.001*

The interactions called for further analyses. Visual inspection suggested, that especially for the adult talker, vowel mismatches in prime-target pairs often affected word recognition differently when the role of prime and target was reversed. For example, while *Para-* numerically facilitated recognition of *Parodie, Paro-* did

not facilitate recognition of *Parasit*. This finding, albeit unexpected, was intriguing because it alludes to important differences in vowel perception which in turn might affect lexical decision. In fact, it is well-attested that vowel discriminability can depend on the direction in which vowels are presented (e.g., Cutler et al., 2004; Repp & Crowder, 1990), and lexical activation has been shown to be affected by these perceptual asymmetries (Cutler et al., 2006; Friedrich et al., 2008; Weber & Cutler, 2004).

The prime-target pairs in the present study comprised 16 different vowel mismatches, and we found in the literature no theoretically-driven predictions about perceptual asymmetries for the complete set of mismatches. However, the Natural Referent Vowel (henceforth, NRV) framework introduced by Polka and Bohn (2011), suggests that there is a universal default bias for the peripheral vowels /uː/, /iː/, and /aː/ which is especially relevant during language development (see also, Schwartz, Abry, Boe, Ménard, & Vallée, 2005). While mature listeners can adjust their initial bias to optimize access to language-specific vowel categories, a privileged fit of the peripheral vowels with human auditory abilities ensures that the bias is also relevant for adult listeners and native contrasts. Using the NRV framework for an exploratory interpretation of the results, the German vowels /uː/, /iː/, and /aː/ are anchor vowels in the present experiment, and a change from an anchor vowel to a non-anchor vowel should be harder to detect than a change in the other direction.

In other words, an anchor vowel in the prime should make the vowel mismatch in the target opaquer, while the vowel mismatch should be more transparent when the anchor vowel occurs in the target. In terms of phonetic-to-lexical mapping, the prediction would be that vowel mismatches that are opaque still prime target word recognition, while vowel mismatches that are transparent do not.

In 41 of our 56 target-prime pairs, an anchor vowel was involved,[1] and for this subset of items the new fixed factor *anchor* coded in a post hoc exploratory analysis if the anchor vowel occurred in the prime or in the target. For the adult talker, an interaction was found between *relatedness* and *anchor* (b = 0.06, SE = 0.03, t = 2.14, $p < .04$), and further analyses showed a facilitatory effect of *relatedness* when the anchor vowel was in the prime (b = -0.07, SE = 0.02, t = -3.25, $p < .002$), and no effect when the anchor vowel was in the target (b = -0.002, SE = 0.02, t = -0.09, $p > .9$). For the child talker, only lexical frequency was significant (b = -3.78, SE = 1.58, t = -2.39, $p < .02$); *relatedness* did not interact with *anchor* (b = -3.82, SE = 3.04, t = -1.25, $p > .2$), and was neither significant when the anchor vowel was in the prime (b = -0.03, SE = 0.02, t = -1.49, $p > .1$). nor when it occurred in the target (b = -0.005, SE = 0.02, t = -0.25, $p > .7$; see Figure 4.2).

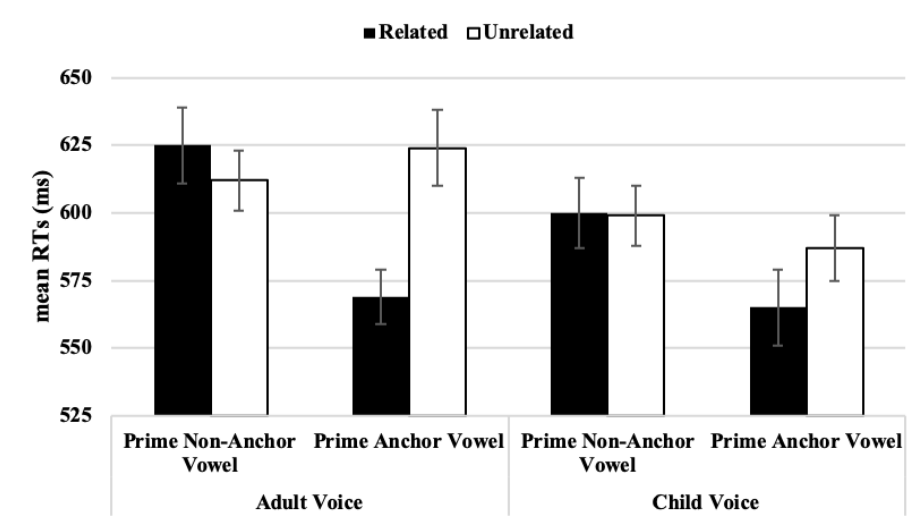

Figure 4.2: Mean RTs (in ms) for the 41 target pairs with the anchor vowels /uː/, /iː/, /aː/, when the anchor vowel occurred in the prime and when it occurred in the target. The vertical bars represent standard errors.

[1]Also German /a/ was considered as an anchor vowel /aː/ as the two phonemes are considered to differ only in duration in German (Strange et al., 2007)

Note that when analyzing the subset of 41 target-prime pairs for both talkers together, the new factor *anchor* did interact significantly with *relatedness* ($b = 0.06$, $SE = 0.03$, $t = 2.14$, $p < .04$) but not with *talker age* ($b = 0.03$, $SE = 0.03$, $t = 1.21$, $p > .2$), thus in fact not licensing the split for the two talkers. For the complete set of 56 target-prime pairs, interactions involving *direction* (rather than *anchor*), *talker age*, and *relatedness* had licensed a split, and further analyses for both talkers showed the exact same pattern of results as was found for the subset with anchor-vowels. Since we found no literature on German vowel confusions that would allow theoretically-driven predictions for all 16 vowel mismatches of the complete set, presenting results based on just the target-prime pairs with anchor vowels seemed appropriate. The pattern of results is however backed up by the analysis of the complete set of target-prime pairs.

The results for the adult voice are in line with the post hoc predictions we derived from the NRV framework (Polka & Bohn, 2011): When the vowel mismatch between prime and target was opaque, the onset fragment of the prime facilitated recognition of the target word (e.g., *Para-* prime *Parodie*); when the vowel mismatch was transparent, there was no priming (e.g., *Paro-* did not prime *Parasit*). However, for the child voice, onset fragments never primed target recognition. Possibly, the vowel space of the child talker was warped, and vowel categories were not distributed as clearly as for the adult talker. Figure 4.3 shows averages for the first two formants at the midpoint of the mismatching vowel in the onset fragments (e.g., [aː] in *Para-* and [o ː] in *Paro-*), separately for the adult voice and the child voice. For each voice, a total of 82 vowels were measured (41 target-prime pairs X 2 members of each pair).

Note that the number of measurements for each vowel varies considerably, since vowel type was not controlled in the experiment (e.g., 25 instances of [a ː] for

each talker and only 3 instances of [ɔ]). As can be expected, the formant values for the child voice were higher, but the overall patterning of vowels in the vowel space seems quite comparable across talkers, certainly with respect to the anchor vowels /uː/, /iː/, /aː/ (see also, S. Lee et al., 1999).

Also note, that overall recognition rates were equally accurate for the two talkers. An alternative explanation for the different patterns is based on listeners' previous experience with the linguistic competence of adult and child talkers. Young children are known to deviate regularly from target norms in their pronunciation (Stoel-Gammon & Menn, 2005), and listeners could take this experience into consideration and hesitate to rely on, for example, vowel information in their interpretation.

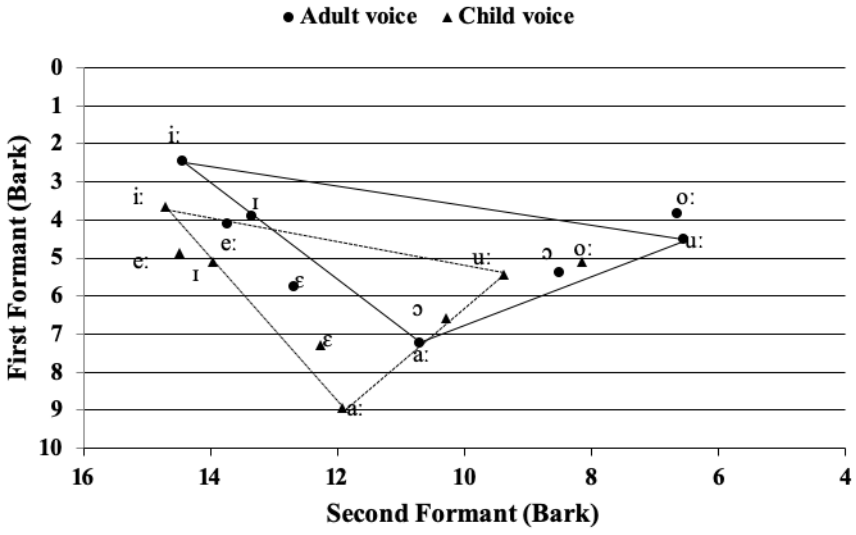

Figure 4.3: Average mid-vowel F1/F2 values (Bark) for all vowels in the subset of 41 target-word pairs with anchor vowels, by the adult talker and by the child talker.

Interim discussion

In the first experiment, we investigated the influence of age of the talker on phonetic-to-lexical mapping. In a first analysis of Experiment 1, no interaction between facilitation for related primes and age of the talker was observed. However, interactions involving age of the talker, relatedness of the prime, and direction of the vowel mismatch warranted further analyses. The NRV framework (Polka & Bohn, 2011) was used to theoretically motivate post hoc predictions for the directionality of vowel mismatches in subsequent analyses (accounting for a subset of 72.3% of the items). When taking directionality into account according to the NRV framework, different priming patterns for the adult voice and the child voice were found. For the adult voice, onset fragments primed target word recognition when they contained anchor vowels /uː/, /iː/, or /aː/ (e.g., *Para-* primed *Parodie*) but not when the anchor vowels occurred in the target (e.g., *Paro-* did not prime *Parasit*). No priming was found for the child voice, neither when the anchor vowels occurred in the fragment primes nor when they occurred in the target.

A comparison of the F1 and F2 values of the vowels produced by the two talkers made it unlikely that the influence of talker age was due to less accurate productions of the child talker. Rather, it seems likely that the phonetic-to-lexical mapping itself was sensitive to the age of talker. One plausible reason for this could be previous experience with children's speech, that often deviates from canonical pronunciations. This experience could set expectations and influence the comprehension process whenever we encounter a (new) child talker. The consequence of experience was such that onset fragments with vowel mismatches of the child talker never facilitated target word recognition. Thus, vowel information in the child voice was never deemed a reliable indicator for the lexical mapping process. Just as well, it could have been that experience led to all vowel mismatches being

accepted as matches for the lexical mapping.

In research on foreign-accented speech, previous experience has indeed been found to make deviations in pronunciation acceptable matches for canonical pronunciations (e.g., Eisner et al., 2013; Trude, Tremblay, & Brown-Schmidt, 2013; Witteman et al., 2013). Note, however, that in most of these studies, experience with specific accents and/or single accent markers have been tested, whereas in Experiment 1 the vowel mismatches comprised a whole range of contrasts. Also, it can be assumed that children vary more between and within talkers in pronunciation than talkers of a specific foreign accent tend to do. Thus, it might be impossible to adapt to anything specific in children's mispronunciations. Taken together, Experiment 1 presented some evidence for the phonetic-to-lexical mapping process being possibly sensitive to the age of the talker. To our knowledge, this is the first time that such an influence has been shown exploratorily for child speech. To test the reliability of this exploratory finding, Experiment 2 is set out to re-examine the effect.

Experiment 2

Experiment 2 aimed to replicate the exploratory finding of Experiment 1, using the same paradigm. According to the NRV framework, a vowel mismatch is easier to detect, when the vowel changes from a more central to a more peripheral vowel (i.e., from an anchor vowel to a non-anchor vowel) than the other way around. In Experiment 2, word pairs from Experiment 1 that consisted of no such change (i.e., anchor vowel to anchor vowel or non-anchor vowel to non-anchor vowel) were excluded from the experimental items.

Based on the exploratory findings in Experiment 1, we expected an

asymmetry in priming effects for lexical items produced by the adult talker compared to no priming for the child talker. This prediction is consistent with the hypothesis that previous experience with the linguistic competence of child and adult talkers can have an influence on the phonetic-to-lexical mapping process.

Method

Participants

Forty-eight native listeners of German (36 female), all students from the University of Tübingen (18-31 years old, mean age = 23.1, SD = 2.8) participated in the experiment and received a small monetary reimbursement. None of them suffered from any hearing disorders, and they all had intact or corrected vision (i.e., contact lenses and glasses).

Material

Forty German word pairs in which one word entailed an anchor vowel and the other did not were taken from Experiment 1. Sixteen further word pairs from Experiment 1 in which both words of the pair entailed an anchor vowel (e.g., *Minister-Minute*, "minister-minute") or no anchor vowel at all (e.g., *Galaxie-Galerie*, "galaxy-gallery") were not used in Experiment 2, making a total of 24 word pairs. The exclusion of some word pairs resulted in a new pairing system for unrelated word pairs. For the current experiment, the vowel mismatches involved an anchor vowel in the prime or in the target. Filler items were the same as in Experiment 1.

Procedure

The procedure was the same as in Experiment 1.

Results

Only trials with correct responses to target words were analyzed. Participants answered on average 84.7% correctly when the primes had been produced by the adult, and 84.0% correctly when the primes had been produced by the child. Similarly to Experiment 1, neither the task nor the different talkers posed considerable difficulties, and performance did not differ for the two talkers (see Figure 4.4). Five participants were excluded from further analyses since they did not meet the criteria, for example, they grew up with more than one language. Only participants who grew up with German as their first language were included in the analysis. As before, reaction times faster than 250 ms and slower than 1200 ms were excluded (0.003%). We used R (R Core Team 2018, version 3.5.0) and lme4 (Bates, Maechler, et al., 2015) to perform a linear mixed effects analysis on log-normalized RTs with *relatedness* (related, unrelated), *age of talker* (adult, child), *anchor vowel* (prime with anchor vowel, target with anchor vowel), as well as *lexical frequency* and *target word length* as fixed factors. The LMER model was built with reaction times (Baayen, 2008) as the dependent measure and fixed factors included *age of the talker*, *relatedness*, and *anchor vowel*. Participants and *items* were included as random factors with random slopes. The results showed a facilitatory effect for *lexical frequency* (b = -0.03, SE = -0.01, t = -3.13, p = .003) and no other significant effects. There were no interactions (all p-levels $>$ 0.1). Thus, in Experiment 2, the exploratory effect of Experiment 1 could not be replicated. While in both Experiments there was no priming when the talker was a child, in Experiment 1 an exploratory analysis found facilitation for primes with anchor vowels for the adult talker, which was not replicated in Experiment 2.

Table 4.2: Full model output. Estimates for the best fitting model for the reaction times when anchor was taken into account.

Fixed effects	*Estimates*	*Std. Error*	*t*	*p*	
Intercept	6.378e+00	2.887e-02	220.946	2e-16	***
Relatedness	1.433e-02	1.976e-02	0.725	0.725	
Talker Age	-2.970e-04	1.994e-02	-0.015	0.98812	
Anchor	-1.438e-02	2.761e-02	-0.521	0.60304	
Lexical frequency	-3.444e-02	1.099e-02	-3.135	0.00247	**
Word length	2.518e-03	7.975e-03	0.316	0.75238	
Relatedness x Talker Age	-2.659e-03	2.830e-02	-0.094	0.92516	
Relatedness x Anchor	8.331e-03	2.833e-02	0.294	0.76874	
Talker Age x Anchor	3.599e-03	2.857e-02	0.126	0.89974	
Relatedness x Talker Age x Anchor	2.279e-02	4.089e-02	0.557	0.57728	

*Note: *p.05 **p.01 ***p.001*

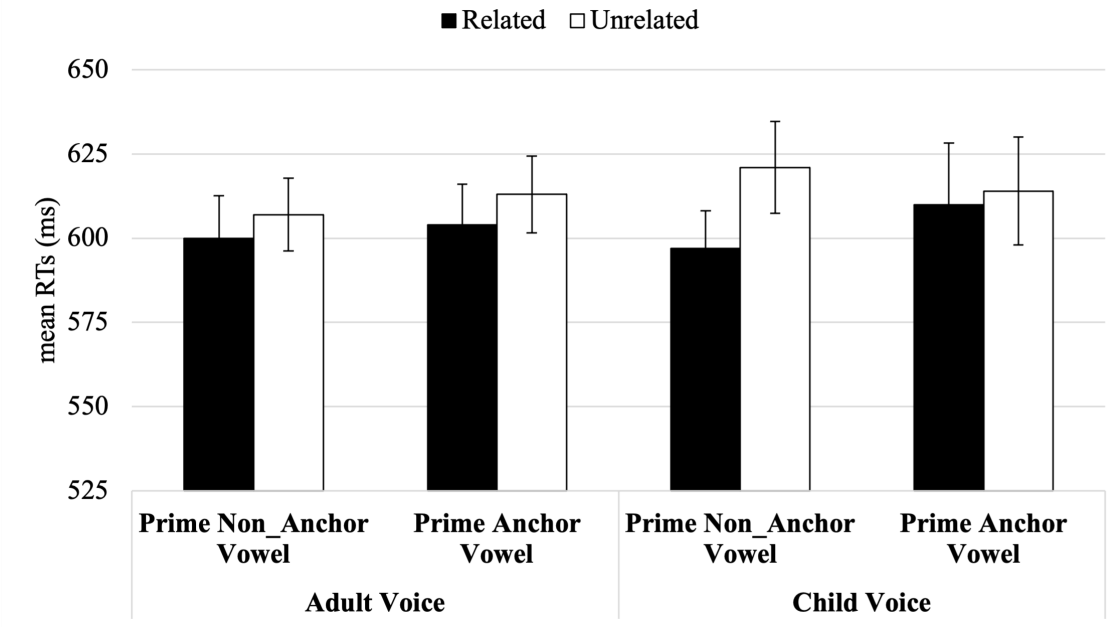

Figure 4.4: Mean RTs (in ms) for the 24 target pairs with the anchor vowels /uː/, /iː/, /aː/, when the anchor vowel occurred in the prime and when it occurred in the target. The vertical bars represent standard errors.

General discussion

The present study investigated if age of the talker can have an impact on the phonetic-to-lexical mapping process. In two experiments with adult German L1 participants, a cross-modal priming paradigm was used in which participants listened to German word fragment primes (e.g., *Para-* from *Parasit*, "parasite") produced by an adult or child talker that mismatched in the second vowel with visual target words (e.g., *Parodie*, "parody"). After listening to a fragment prime, participants made lexical decisions to the visual target via button press, indicating whether they considered the visual string of letters a real word of German or not. Overall, while Experiment 1 found that the phonetic-to-lexical mapping process was being possibly sensitive to the age of the talker, Experiment 2 found no such effect. Specifically, Experiment 1 found facilitation for the adult voice, whereas no priming was found for the child voice. In contrast to that, Experiment 2 found no facilitation for both talkers.

Results of the initial analysis of Experiment 1 showed that the phonetic-to-lexical mapping process might have been influenced by talker age, as indicated by two interactions that involved the age of the talker. The interactions also involved the directionality of the prime-target pairs and motivated subsequent exploratory analyses based on the NRV framework (Polka & Bohn, 2011). The NRV framework postulates directional asymmetries in the discrimination of speech sounds. This framework is based on the notion that spoken language follows a particular order, meaning that earlier events can have an influence on the discrimination or recognition of later events such that discrimination is easier when vowel contrasts are presented in one specific direction (e.g., order /e/-/i/) compared to the reverse direction (e.g., order /i/ - /e/). This can be attributed to the fact that

vowels which are positioned at the corner (periphery) of the vowel space (i.e., /uː/, /iː/, and /aː/) serve as natural referents or perceptual anchors. Anchor vowels that are presented second therefore make vowel discrimination easier (Polka & Bohn, 2011). Hence, discrimination of vowels presented in the order as in /e/-/i/ is easier than the reversed order. Attributing this theory to the phonetic-to-lexical mapping process, prime words that contain an anchor vowel should make the vowel mismatch in the target more opaque, while the vowel mismatch should be more transparent when the anchor vowel is in the target. Using this theory for our post hoc predictions, we assumed that a vowel change from an anchor vowel to a non-anchor vowel would still induce priming of target word recognition, but the reversed vowel change would not. Exploratory analyses indeed confirmed our post hoc prediction. Results showed an impact of talker age on the mapping of phonetic information to lexical representation, meaning that priming only occurred for the adult talker in one direction, but never for the child talker. That is, for the adult talker, anchor vowels that occurred in the onset fragments facilitated priming (e.g., *Para-* primed *Parodie*) but not vice versa (e.g., *Paro-* did not prime *Parasit*).

The present findings differed from Friedrich et al. (2013) in terms of priming, because while the present study obtained facilitatory effects, Friedrich et al. (2013) found inhibitory effects for partially overlapping prime-target pairs. Note that Friedrich et al. (2013) collected behavioral data and ERPs at the same time, which also showed diverging patterns. Particularly, a trend for inhibition was obtained in the behavioral data, thus strengthening Soto-Faraco et al. (2001) findings, but the ERP data showed inhibition for only some of the partially matching prime-target pairs. It is plausible that directionality of the paired prime-target pairs might have also played a role in Friedrich et al. (2013) but was not analyzed as such. In addition, Friedrich et al. (2013) used three groups of item pairings, that is (1)

fully matching prime-target pairs (e.g., prime *Ana-* from *Ananas*, target *Ananas*), (2) partially matching prime-target pairs (e.g., prime *Ana-* form *Ananas*, target *Anorak*), and unrelated pairs (e.g., prime *Ana-* from *Ananas*, target *Eloquenz*). In contrast to Friedrich et al. (2013), experimental items of the present study consisted of only partially and unrelated prime-target pairs, and fully matching pairs were nonword pairs, which served as filler items. This implies that while the present study compared lexical decisions of partial overlapping and unrelated pairs, Friedrich et al. (2013) compared responses between fully overlapping, partially overlapping and unrelated pairs. It is therefore conceivable that in the presence of fully overlapping pairs, partially overlapping pairs contrasted more readily with the mismatch resulting in inhibition. In addition, the directionality of some pairs possibly made the mismatches more prominent since only some of the pairs showed inhibition. However, as directionality was not further investigated in Friedrich et al. (2013), it is at this point merely speculation. However, this possibly explains why the present study did not replicate the inhibition pattern. Particularly, the absence of fully matching prime-target pairs may have led participants to accept partially overlapping prime-target pairs more readily as acceptable matches, thereby causing the facilitatory effect, and this was only the case for some pairs that met the NRV vowel directionality, as proposed by Polka and Bohn (2011).

The exploratory findings of Experiment 1 motivated a re-examination in a subsequent Experiment 2 with a new pool of participants and with only the prime-target pairs from Experiment 1 that contained both a word with an anchor vowel and a word without a non-anchor vowel. This time, no effect of talker age or interaction involving that factor was found, thereby not replicating Experiment 1. Recall that different priming patterns for the adult voice and the child voice were found in Experiment 1. While facilitation was found for the adult voice, no

facilitatory effect was observed for the child voice. That is, neither anchor vowels in the prime nor in the target facilitated priming. The fact that no priming for the child voice was found once again refutes the initial assumption that listeners would be more open and forgiving toward mismatches of the child talker. Instead, listeners were more reluctant to rely on vowel information in their interpretation, possibly because of their previous experience with the linguistic proficiency of child talkers. However, it was unexpected that no facilitation for the adult talker occurred in Experiment 2, because the directional asymmetry, which was found in Experiment 1, is considered to be a robust perceptual phenomenon that has been replicated successfully (e.g., Masapollo, Polka, Molnar, & Ménard, 2017; Masapollo, Polka, & Ménard, 2015; Zhao, Masapollo, Polka, Ménard, & Kuhl, 2019) and across various listener populations. For example, perceptual asymmetries have been found in native and non-native contrasts of infant listeners who were up to 12 months old as well as in adult perception of non-native contrasts (see for a review, Polka & Bohn, 2011). Hence, the absence of an effect in Experiment 2 was surprising.

One possible reason for the absence of directional asymmetries may be attributed to the stimuli used at test. For Experiment 2, those experimental word pairs that did not fulfill the NRV vowel directionality (anchor to anchor; non-anchor to non-anchor) were excluded. Consequently, the reduced set of items might have diminished the priming effect. However, the lack of the perceptual asymmetries could also be explained by differences in tasks used in the present study and earlier studies. The majority of behavioral studies showed support for the NRV framework, but they mostly used a vowel discrimination paradigm with two or three vowel contrasts for their investigation (Masapollo et al., 2017, 2015, 2018), whereas the present study used a larger range of vowel contrasts embedded in existing words of German. Thus, compared to the cross-modal priming task

used in the present study, which requires listeners to map the acoustic input to mental representation stored in the mental lexicon, the tasks and items that are commonly used in the NRV literature differ substantially with regard to task demands and performance requirements. While vowel discrimination studies have shown robust results, favoring the NRV framework, neurophysiological evidence for these asymmetries is less consistent (De Rue, Snijders, & Fikkert, 2021; Polka, Molnar, Zhao, & Masapollo, 2021; Riedinger, Nagels, Werth, & Scharinger, 2021).

For example, Riedinger et al. (2021) used monosyllabic German words containing long vowel contrasts (/uː/, /iː/, and /aː/, /eː/, /yː/) in both electrophysiological (i.e., MMN) and behavioral experiment (i.e., reaction time). While the MMN results showed support for the NRV framework, the reaction time experiment did not show such effect. In the MMN task, stimuli were embedded in a passive oddball design, EEG signals were recorded while participants listened to a sequence of stimuli that were occasionally interrupted by the deviant, and at the same time they were watching a silent movie. In the behavioral experiment, participants took part in an active oddball design. Participants listened to stimuli and pressed a button as soon as they heard the deviant. Results showed easier discrimination from a non-anchor vowel to an anchor vowel and were in line with the NRV hypothesis (Polka & Bohn, 2011) and other behavioral and electrophysiological studies (e.g., Masapollo et al., 2017, 2015; Zhao et al., 2019). For example, the asymmetric pattern of the comparison between /eː/-/aː/ (e.g., *Mehl-Mahl*; 'flour-feast'). Here, /aː/ is an anchor vowel in, and discrimination of /eː/-/aː/ is, therefore, easier and comes with a stronger MMN effect than vice versa. By contrast, data from the reaction time experiment did not match with the MMN data. Thus, the NRV framework failed to comprehensively explain the lack of directional asymmetries in the second part of Riedinger et al. (2021). They,

therefore, cautiously proposed that the directional asymmetry, introduced by the NRV, is dependent on different attention requirements or different processing levels between the electrophysiological and behavioral paradigms, possibly explaining the different results in the two experiments.

To our knowledge, this was the first study to investigate the potential influence of talker age on phonetic-to-lexical mapping processes. While in Experiment 1 an influence of talker age has been observed in an exploratory analysis within the NRV framework, which we interpreted as an effect of previous experience with the linguistic competence of child and adult talkers, this effect was not replicated in Experiment 2. Maybe the methodology used did not suit an investigation with the NRV framework well, or other factors like the random alternation of talkers in the experiment were responsible for the lack of an effect of talker age. In our experiments, trials were not blocked by talkers, but talkers alternated randomly between trials. This design decision could have caused an indirect influence, in the form of a spillover effect. It is evident that listeners can generally differentiate between adults' and children's voices easily. Nonetheless, it is possible that the response to a current trial could have been affected by the previous trial. For example, if lexical decision times did indeed reflect the consideration of talker age in a given trial, for example a trial with an adult talker, this consideration could still be lingering and influencing a subsequent trial, even when the talker has a different age, for example is a child. Future research is needed to clarify the impact of talker groups varying in age and phonetic-to-lexical mapping and task demands.

CHAPTER 5.1

The impact of face masks on the recall of spoken sentences

With the permission of the Acoustical Society of America, this chapter has been reproduced from

Truong, T. L., Beck, S. D., and Weber, A. (2021). The impact of face masks on the recall of spoken sentences. *The Journal of the Acoustical Society of America, 149*(1), 142-144. https://doi.org/10.1121/10.0002951

Abstract

The effect of face-covering masks on listeners' recall of spoken sentences was investigated. Thirty-two German native listeners watched video recordings of a native talker producing German sentences with and without a face mask, and then completed a cued-recall task. Listeners recalled significantly fewer words when the sentences had been spoken with a face mask. This might suggest that face masks increase processing demands, which in turn leaves fewer resources for encoding speech in memory. The result is also informative for policy-makers during the COVID-19 pandemic, regarding the impact of face masks on oral communication.

Introduction

Understanding spoken language requires the translation from speech signal to meaning: phonetic, lexical, and syntactic information must be extracted, and linguistic meaning in sentences must be composed. As adult listeners, we typically carry out these complex mental tasks with astonishing ease and speed. However, processing becomes cognitively more demanding when the speech signal is acoustically degraded or ambiguous (e.g., Ernestus, Baayen, & Schreuder, 2002; Witteman et al., 2013). Increased listening effort in adverse conditions has also been shown to affect higher-level cognitive processing downstream, such as memory encoding. That is, listeners are worse at recognizing which words they have heard before and at recalling exact lexical items when the speech input is degraded, for example, in casual or accented speech or in noisy environments (e.g., Gilbert et al., 2014; Grohe & Weber, 2018; Keerstock & Smiljanic, 2019).

In this study, we examined the effect of wearing a face mask on subsequent recall of spoken sentences. A talker's lip and jaw movements convey linguistic information. For example, lip closure correlates with a bilabial place of articulation for the stop consonants /p/ and /b/, and the openness of the jaw is correlated with the height of vowels (more open jaw for the vowel /a/ and less open jaw for /i/). This visual information is complementary to the auditory signal, and information from both domains is integrated during speech perception (e.g., Jesse & Massaro, 2010). Concealing visual speech information with a mask could therefore result in a decrease in encoding performance. At the same time, mask material could degrade the acoustic signal by dampening it and acting as a low-pass filter. While some studies indeed found effects of various types of mouth and face coverings on speech acoustics (e.g., Corey, Jones, & Singer, 2020; Mendel, Gardino, & Atcherson, 2008),

others found the effects to be negligible (e.g., Llamas, Harrison, Donnelly, & Watt, 2009).

We tested the effect of face masks on memory for spoken language using a cross-modal cued-recall task (see Keerstock & Smiljanic, 2019). Native German listeners watched video recordings of a native talker producing sentences (e.g., *Die Köchin hilft montags armen Kindern*, "The cook helps on Mondays poor children") with and without a face mask. Face masks in public places have been mandatory in many countries during the COVID-19 global pandemic and have become part of our daily lives. There is currently a need to better understand the possible impact of wearing a mask, not only on physical and psychological comfort, but also on verbal communication. Testing the retention of spoken information is one aspect of this.

Methods

Participants

Thirty-two native German listeners between the ages of 20 and 37 years participated in the study (mean: 23.8; 28 females). All participants indicated normal hearing and vision. They were recruited via social media and university email, and electronically signed written informed consent and filled out a brief language background questionnaire. For monetary compensation, participants were given the opportunity to participate in a lottery.

Stimuli

The stimuli consisted of 48 German sentences, modeled after the Oldenburger Satztest (*Oldenburger Satztest: Handbuch und Hintergrundwissen*, 2000). The full list of sentences can be found in Appendix B (see Chapter 7). All sentences

began with a determiner and a noun, followed by a verb, an adverb, an adjective, and a noun. The sentences were not highly predictable in order to reduce the facilitatory influence of context, and to ensure a more thorough processing of the input (Rommers & Federmeier, 2018). All words were of high lexical frequency, and each content word occurred only once in the stimuli. A 22-year-old female native talker of German was video recorded producing all sentences with and without a face mask (see Figure 5.1.1).

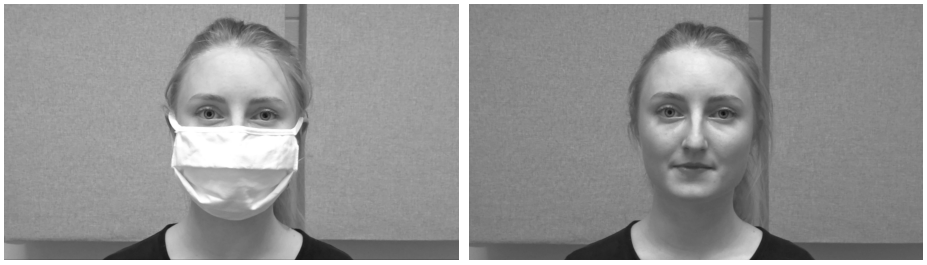

Figure 5.1.1: Representative screenshots for video recordings with and without a face mask. Videos were presented in color in the experiment.

Recordings were made in a sound-attenuated room with a high-quality, stationary RØDE microphone at a sampling rate of 44 kHz and a Sony DSC-Hx90 camera recorder with video resolution parameters set to Full HD 1920 x 1080, which was positioned to capture the talker's head and shoulders. The face mask was made of two layers of fabric: The inner layer consisted of a thin fleece layer, and the outer layer was cotton. The talker was instructed to produce all sentences at a normal speaking rate without hesitations or pauses and to not speak more clearly or loudly when wearing the mask. Unmodified, natural sentence recordings without a mask were on average 3172 ms long and with a mask 3253 ms (t = -1.39). Spectral analysis (RMS power) revealed no difference between sentences with (56.6 dB) and without a face mask (56.7 dB) (t = -0.28).

Procedure

The experiment was implemented with the online software *SurveyGizmoLLC* (surveygizmo.com, 2020). Participants were asked to wear headphones and participated online. The experiment started with two practice sentences and continued with the 48 experimental sentences, divided into eight blocks of six sentences. Sentence order was randomized once, and half of the participants watched the videos in reverse order. The presence of a face mask was blocked, and blocks alternated between the mask and no-mask condition. The order of mask condition was counterbalanced, and sentences were presented with an ISI of 2500 ms. The self-paced cued-recall task followed each block.

For this task, sentences were presented up to the adverb orthographically on the screen (e.g., *Die Köchin hilft montags*, "The cook helps on Mondays"), and participants were asked to type in the two missing final words (e.g., *armen Kindern*, "poor children") on their keyboard. For each participant, there was a total of 96 keywords (2 keywords in each of the 48 sentences) to be recalled. All sentence beginnings of a block were available at once, in the order of block presentation, and participants could choose in which order they typed their responses.

Results

Each recalled keyword was scored by the first author and a research assistant as either correct (1) or incorrect (2) (see Figure 5.1.2). Approximately 70% of all responses that were categorized as incorrect, had been omissions. In the remaining 30% of incorrect responses, a variety of error types was observed: the majority were responses that were unrelated in form and meaning to the keywords (e.g., *schwarze Schuhe*, "black shoes", for *staubige Kissen,* "dusty pillows"); a much

smaller number of responses were closely semantically related (e.g., *Ringe*, "rings", for *Kreise*, "circles"); only few responses were phonetic errors involving a single sound change, that is, a substitution, insertion, or deletion (e.g., *Schweine*, "pigs", for *Steine*, "stones") or typos (e.g., the nonword *Lmpen* for *Lampen*, "lamps").

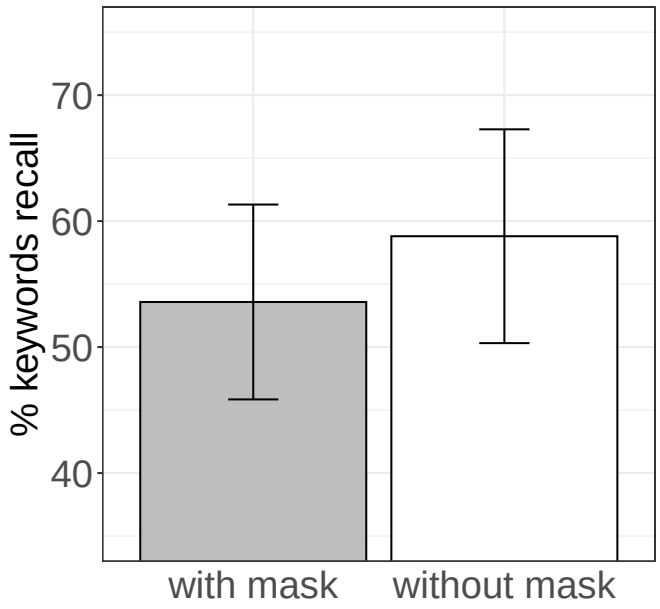

Figure 5.1.2: Average percentage of keywords recalled correctly for sentence recordings with and without a face mask. The vertical bars represent standard errors.

To assess the effect of face masks on listeners' keyword recall, a logistic mixed-effects regression model (Jaeger, 2008) was implemented using the lme4 package (Bates, Kliegl, et al., 2015) in R (R Core Team 2020, version 4.0.2). Accuracy was modeled as binary categorical *keyword recall* (success vs. failure). *Face mask* (mask vs. no mask) and *block* (8 blocks) were entered as fixed effects. To test linear and quadratic effects of block, orthogonal polynomials were used (Mirman, 2017). *Items* and *participants* were included as random crossed effects

(Baayen et al., 2008), with random intercepts and random slopes for both. The analysis showed a difference in keyword recall when the talker was not wearing a mask compared to when she was wearing a mask (b = -0.29, SE = 0.12, t = -2.4, p = .017). There was no significant interaction. The values of the model can be found in Table 5.1.1.

Table 5.1.1: Full output of the LMER model

Fixed effects	*Estimate*	*Std. Errors*	*t*	*p*	
Intercept	0.2801	0.1808	1.549	0.1214	
Linear	0.3004	0.3438	0.874	0.3823	
Quadratic	-0.4675	0.3478	-1.344	0.1789	
Face mask	-0.2890	0.1207	-2.394	0.0167	*
Linear x Face mask	-0.3591	0.3026	-1.186	0.2355	
Quadratic x Face mask	-0.1448	0.3033	-0.477	0.6331	

*Note *p<.05 **<.01 ***<.001*

Conclusion

In a cued-recall experiment, native adult listeners recalled fewer words when the talker had been wearing a face mask than when she had not been wearing one. This result suggests that processing speech produced with a face mask leaves fewer cognitive resources available for storing spoken information in memory. Face masks both conceal visual speech information and can degrade the acoustic signal (e.g., Corey et al., 2020; Mendel et al., 2008). While the present study was not set out to tease apart the reasons for why face masks decrease encoding performance, we have some indications that neither the acoustic signal nor speech perception were affected much by the mask. A lack of a difference in RMS values between the

mask and no-mask condition indicates that, at least spectrally, the two conditions did not differ. In a small post-hoc experiment, we also asked an additional 12 participants to write down the keywords after individual sentences, rather than after blocks of sentences, rendering the task into an assessment of intelligibility. Performance was overall highly correct and did not differ between the mask (98.95% correct) and no mask condition (99.3% correct). Thus at least for a clear speech style, spoken through a cotton mask and recorded in a quiet environment, it seems that missing visual cues rather than decreased intelligibility were the main factor causing a decrease in encoding performance. Future experiments investigating the intelligibility of speech with masks in noise can, however, help to clarify this point.

In order to get a fuller understanding of the impact of face masks on memory for spoken language, different participant groups and talkers must be tested next. For example, non-native listeners and children can be expected to have more difficulties in perceiving spoken language than native adults due to their incomplete mastery of the target language. For these listener groups, removing visual cues with a mask might have an even stronger impeding effect on memory (e.g., Keerstock & Smiljanic, 2018). Also, talkers with varying language experience (e.g., non-natives and children) can deviate noticeably in their pronunciation from the target norms of a language. In such cases, native adult listeners typically rely even more on visual speech cues (e.g., Xie, Yi, & Chandrasekaran, 2014), and concealing these cues with a face mask can be expected to intensify the negative effect on the encoding of spoken information.

For native adult listeners and native speech, the present results already indicate that face masks can impede memory for what has been said. This finding should have implications for communication in various situations, for example, in classrooms and doctor's offices where remembering spoken information is crucial.

CHAPTER 5.2

Intelligibility and recall of sentences spoken by adult and child talkers wearing face masks

With the permission of the Acoustical Society of America, this chapter has been reproduced from

Truong, T. L. & Weber, A. (2021). Intelligibility and recall of sentences spoken by adult and child talkers wearing face masks. *The Journal of the Acoustical Society of America*, 150, 1674-1681. https://doi.org/10.1121/10.0006098

Abstract

With the Covid-19 pandemic, face masks have become part of our daily lives. While face masks are effective in slowing down the spread of the virus, they also make face-to-face communication more challenging. The present study sought to examine the impact of face masks on listeners' intelligibility and recall of sentences produced by one German native adult and one child talker. In the intelligibility task, German native adult listeners watched video clips of either an adult or a child talker producing sentences with and without a face mask. In a cued-recall experiment, another group of German native listeners watched the same video clips and then completed a cued-recall task. The results showed that face masks significantly affected listeners' intelligibility and recall performance, and this effect was equally true for both talkers. The findings here contribute to the fast growing and urgent research regarding the impact of face mask on communication.

Introduction

With the Covid-19 pandemic, communication has changed. In particular, face masks present an additional challenge for listeners, as masks can modify the speech signal and conceal at the same time visual articulatory cues. Recent evidence suggests that, for adult talkers, both intelligibility and recall of what has been said can be negatively influenced by face masks (Smiljanic, Keerstock, Meerman, & Ransom, 2021; Truong, Beck, & Weber, 2021). But what about understanding and recalling what children have said? Children's voices are different from adult voices, and this could affect not only listeners' ability to understand and encode what has been said but also the relevance of visual articulatory information in face-to-face communication. The present study set out to investigate the intelligibility and recall of sentences spoken by a child talker in comparison to an adult talker, when the talkers are wearing a face mask or not.

Visual cues, such as lip and jaw movements, can provide crucial linguistic information about speech sounds (e.g., Campbell, 2008; Summerfield, 1992). For example, lip closure is associated with a bilabial place of articulation as in the stop consonants /p/ and /b/, and the openness of the jaw correlates with the height of vowels (e.g., more open jaw for the vowel /a/ and less open jaw for /i/). Therefore, extracting information from a talker's visible articulators can supplement and complement information about speech sounds that is not included in the auditory signal (Sumby & Pollack, 1954). Indeed, information from both modalities is known to be automatically integrated during speech perception (e.g., Jesse & Massaro, 2010). Face masks, however, constrain access to visual articulatory information, and only auditory information is left for speech perception.

Face masks can potentially also degrade the acoustic signal itself by

affecting the speech directivity and attenuating higher frequencies, which can result in a transmission loss and thus impact comprehension. While face coverings have been found to affect speech acoustics and comprehension in some studies (Corey et al., 2020; Goldin & Weinstein, 2020; Pörschmann, Lübeck, & Arend, 2020), other studies have found no such effects (Llamas et al., 2009).

Generally speaking, any acoustically degraded speech (e.g., background noise or variation in pronunciations) provides listeners with less information, for example, talkers varying considerably in their exact acoustic realization of phonemes and words as well as poor acoustic background conditions. For instance, they can degrade the speech signal and have been found to affect the intelligibility of speech produced by adult talkers (e.g., Ernestus et al., 2002; Peelle, 2018; Wittemann, Weber, & McQueen, 2014). These adverse listening conditions can thus make perceptual processing more effortful.

Given that listeners' cognitive resources are limited, mental resources in adverse listening conditions will be reallocated from memory back to perception (Pichora-Fuller et al., 2016). That is, additional effort can attenuate the listening challenges, but it may come at the expense of cognitive resources that might else be available for memory encoding (see, McCoy et al., 2005; Rabbitt, 1968; Rönnberg et al., 2013). Hence, degraded speech not only causes listeners to be less accurate in word identification, but at the same time it can negatively affect higher-level cognitive processing downstream, such as memory encoding (Pichora-Fuller et al., 2016; Rabbitt, 1968). Worse performance in recognizing previously heard words and recalling them has been found before in noisy conditions for conversational speaking styles and for unfamiliar accents (Gilbert et al., 2014; Grohe & Weber, 2018; Keerstock & Smiljanic, 2019).

However, all previous intelligibility studies on face masks used audio recordings for their investigations, thus not considering how the lack of visual input when seeing a talker with a mask influences speech perception. Recently, Smiljanic et al. (2021) investigated this issue by using video recordings of two talkers (native vs. non-native) producing two speaking styles (clear speech vs. conversational speech) of a cohesive text. The video recordings were presented either in quiet or in the presence of different levels of competing speech noise. Their findings suggest that face masks did not negatively affect intelligibility of conversational native speech when little or no noise was present, but a negative effect emerged for higher noise levels. In comparison, for non-native speech this negative mask-effect emerged already when little noise was present.

Concerning the effect of face masks on memory, **Chapter 5.1** previously found for German that face masks negatively affected memory encoding when sentences of an adult native talker were presented in quiet. That is, listeners recalled fewer words when the talker had been wearing a face mask than when the talker had not been wearing one. Smiljanic et al. (2021) also tested the impact of face mask on subsequent recall in English. While they found no mask-effect in quiet listening condition for an adult native talker, an impact of face masks on memory was observed when sentences were mixed with noise. The difference in findings between the two studies could well be due to methodological differences that include the type of material (dissociated sentences vs. cohesive text) and memory questions (only recall questions vs. various question types). We expand on the findings of **Chapter 5.1** by investigating whether or not child speech produced with a face mask shows similar effects as those for adult speech.

Previous work on intelligibility and memory has almost exclusively investigated adult listeners' ability to understand and recall other adult talkers (but

see Cooper, Fecher, & Johnson, 2020). Here, we investigated word recognition and recall of sentences produced by a child talker in comparison to an adult talker. Children's voices can differ from adult voices on several acoustic and linguistic dimensions, which may impact listeners' ability to understand and encode what has been said by a child. More specifically, children's acoustic and linguistic properties differ from those of native adult speech (S. Lee et al., 1999) such that children's speech is generally characterized by greater acoustic-phonetic variation and by overall higher fundamental frequency (i.e., F0) than native adult speech (B. Smith et al., 1983; Tingley & Allen, 1975). For instance, a comparison of vowel productions by children and adults found that formant frequency averages of children are about 16% higher than that of adults (Kreiman and Sidtis, 2013; but see Hillenbrand, Getty, Clark, and Wheeler, 1995 and Vorperian and Kent, 2007). This difference in F0 is largely due to distinct anatomical characteristics of children which have a smaller larynx and shorter vocal folds. An 8-year-old child's vocal folds are, for example, about 8 mm long, while adult vocal folds are about 12-21 mm long. Because of these distinct physiological features, adult and child voices differ notably in F0 (Kreiman & Sidtis, 2011). At birth, an infant's F0 is at approximately 500 Hz, but by the time the child turns eight, F0 can be as low as 275 Hz, with little difference between boys and girls (Vorperian & Kent, 2007). While F0 remains relatively stable throughout the rest of childhood, children's speech motor control progressively increases until the age of 8-12 years, such that children gain better control over speaking rate, loudness, phonation, and pitch range, gradually meeting adult speech norms (Kent, 1976).

The child talker in the present study was nine years old. Even though pronunciation norms can already approach adult performance when children are five or six years old (i.e., there are only few segmental deviations or mispronunciations,

see Vance et al., 2005), children's voices are still higher than adult voices at that age (Vorperian & Kent, 2007). Since children have a higher F0 compared to adults, it is possible that the child speech produced with a mask is more affected by the mask and generally less intelligible than that of adult talkers, as face masks particularly attenuate higher frequencies. Additionally, limiting access to visual articulatory information through a face mask could further hamper performance especially for child talkers.

The current study investigated how face masks influence intelligibility (Experiment 1) and recall (Experiment 2) of spoken sentences. For the intelligibility experiment, we predicted for the adult and child talkers that a lack of visual cues would affect speech intelligibility negatively, such that recognition is less accurate when the sentences were produced with a face mask than without. This outcome would be in line with Smiljanic et al. (2021) who found that intelligibility was negatively affected by a face mask when native adult conversational speech was presented with a negative SNR. Additionally, it was deemed possible, that the child talker would be less intelligible overall and/or the negative effect of the mask could be enlarged for the child talker.

For the cued-recall experiment, the same recordings were used, but this time sentences were grouped in blocks and presented in quiet. We predicted similar findings to Truong et al. (2021) (i.e., Chapter 5.1), who used the same sentences and presented a subset of participants responding to the adult talker. Based on Truong et al.'s (2021) findings, we predicted that recall rates would be lower for sentences produced with a face mask compared to sentences produced without a face mask for both talkers. Such a result would be in congruence with the effortful hypothesis arguing that listeners must allocate additional cognitive resources when the listening situation is difficult, and this compromises subsequent memory encoding (Peelle,

2018). Additionally, if the child talker was harder to understand than the adult talker in the intelligibility task in Experiment 1, or if not wearing a mask is particularly important for recall when the talker is a child, then the size of the mask-effect in Experiment 1 might be larger for the child talker than the adult talker.

Experiment 1

Methods

Participants

Eighty native German listeners between the ages of 18 and 36 years participated in the study (mean: 22.3, SD = 3.2; 66 females) for a chance to take part in a monetary lottery. All participants reported that German was their first and dominant language. Participants were gathered through social media and university email. None of them reported hearing or vision impairments. Half of the participants watched an experimental version in which all sentences were produced by the female adult talker, and the other half watched a version in which all sentences were produced by the female child talker.

Stimuli

The stimuli consisted of 48 meaningful but not highly predictable sentences, which were modeled after the Oldenburger Satztest (2000). Low predictability had the advantage that listeners could not easily guess individual words correctly without having understood them, since sentence context did not semantically constrain lexical options. The risk of a facilitatory influence of context was therefore relatively low and ensured a more thorough processing of the input (see e.g., Rommers & Federmeier, 2018). The syntactic structure of all sentences was as follows: The

sentences started with a determiner and a noun, followed by a verb, an adverb, an adjective, and a noun (e.g., *Die Köchin hilft montags armen Kindern*, "The cook helps on Mondays poor children"). Each content word occurred only once in the stimuli.

The talkers selected for the experiment were a 22-year-old, female adult, native talker of German and a nine-year-old, female child, native talker of German, who both grew up in the south of Germany, where the experiment was also conducted (see Figure 5.2.1). The talkers were video recorded separately and produced all sentences with and without a face mask. The face mask consisted of two fabric layers: The inner layer was made of thin fleece, and the outer layer was cotton. The talkers were instructed to produce all sentences at a normal speaking rate without hesitations or pauses and to not speak more clearly or loudly when wearing the mask.

Recordings were made in a sound-attenuated room at the LingTüLab of Tübingen University. The talkers repeated the sentences until they were produced without any errors or hesitations. The videos were recorded by using a Sony (Tokyo, Japan) DSC-Hx90 camera with video resolution parameters set to FULL HD 1920x1080, capturing the head and shoulder of the talker (see Figure 5.2.1). Audio was recorded at a sampling rate of 48 kHz with a high-quality microphone placed in front of the talker. Video recordings were segmented using iMovie. Audio was then detached from each segmented video and mixed with noise with a -12 dB SNR in Praat. As in the Bent and Bradlow (2003) intelligibility study, we used white noise, and the SNR level was chosen based on informal pre-testing that yielded an intermediate level of word identification rates for our sentences. While this level of noise can be considered profound, it was deemed necessary to stay clear from a ceiling performance for the chosen short sentences and words with high lexical

frequency. The mixed audio clips were then reattached to the corresponding videos.

The average F0 value of the adult talker was 235.5 Hz, and that of the child talker was 288.7 Hz ($t = 39.35$, $p < 0.01$). Durations for sentences produced by the adult talker without a mask were on average 3255 ms, and with a mask they were 3178 ms ($t = 1.35$ $p = 0.18$). Sentences produced by the child talker without a mask were on average 3997 ms long, and with a mask they were 3928 ms long ($t = 1.13$, $p = 0.26$). While spectral analysis [root mean square (rms) power] of the adult talker revealed no difference between sentences with (56.6 dB) and without a face mask (56.7 dB) ($t = 0.28$, $p = 0.77$), rms power for the child talker revealed a small but significant difference between sentences with (60.31 dB) and without a face mask (61.27 dB) ($t = 4.2$, $p < 0.001$).

Figure 5.2.1: Representative screenshots for video recordings of both adult and child talker with and without a face mask. Videos were presented in color in the experiment.

Procedure

The experiment was administered with the online software *SurveyGizmoLLC*, now called *Alchemer* (alchemer.com, 2020). Participants were asked to wear headphones and participated online. We had emphasized in the instructions to use headphones and take part in the experiment on a computer, laptop, or tablet. Participants furthermore indicated after the experiment the type of device and headphones they had used. Prior to the experiment, they electronically signed written informed consent and were informed that they would listen to sentences mixed with noise.

The experiment started with a practice trial and continued with the 48 experimental sentences recorded either with or without a face mask. During the practice trial, participants were asked to adjust the volume level to a comfortable listening level at which they could understand the sentences best and to keep it the same for the entire experiment. Mask condition was counterbalanced, and sentence order was randomized once, with half of the participants watching the video clips in the reverse order. After each video clip, a prompt with empty boxes for the words appeared on the screen and participants were asked to type in the sentence they had just heard. They were asked to write down as many words as they had understood and to leave the box empty if they had not understood a word. The whole experiment lasted approximately 20-30 minutes. After the experimental session was completed, participants filled out a brief language background questionnaire and were asked about technical problems of which none were reported.

Results

For the purpose of the analysis, the initial determiner and noun (e.g., *die Köchin*, "The cook") of the sentences were considered as one keyword, resulting in a total

of five keywords for each sentence.[1] The maximum number of correct keywords for each participant was therefore 240 (5 keywords x 48 sentences). Each keyword was scored as either correct (1) or incorrect (0) (see Figure 5.2.2). Scoring was done by T.L.T. and a research assistant. For any remaining uncertainties, A.W. was consulted. Overall, 51.8% of the keywords were identified as correct and 48.2% as incorrect.[2] Correct identification responses included identical matches with the intended word forms (97.8% of the correct responses for the adult talker, and 96.6% for the child talker), and what we categorized as typing errors (e.g., nonwords with mixed up letter order like *orndet* for *ordnet*, "orders", word forms with a minimal segmental difference that can be used in free variation in German like *gern/gerne*, "gladly") (1.2% of the correct responses for the adult talker and 1.8% for the child talker).

To assess the effect of face masks on listeners' keyword recognition accuracy, a logistic mixed-effects regression model (Jaeger, 2008) was incorporated using the lme4 package (Bates, Kliegl, et al., 2015) in R (R Core Team 2021, version 4.0.5) with *Correct keyword recognition* (Success vs. Failure) as the dichotomous dependent variable. The model included *face mask* (mask vs. no mask) and *talker* (adult vs. child) as independent variables, and *face mask* x *talker* as an interaction term. To account for additional variation, fixed factors of *sentence duration* and *rms power* were also included in the model. *Items* and *participants* were included as random crossed effects (Baayen et al., 2008), with random intercepts and random slopes. The values of the lmer model can be found in Table 5.2.1.

[1]This was done because it was considered unlikely that participants would make a gender error between the determiner and noun. And indeed, none had made this mistake in the data.

[2]Of all incorrectly identified keywords, the majority had been omissions, that is, blank keyword boxes (70.0% for the adult talker; 77.3% for the child talker).

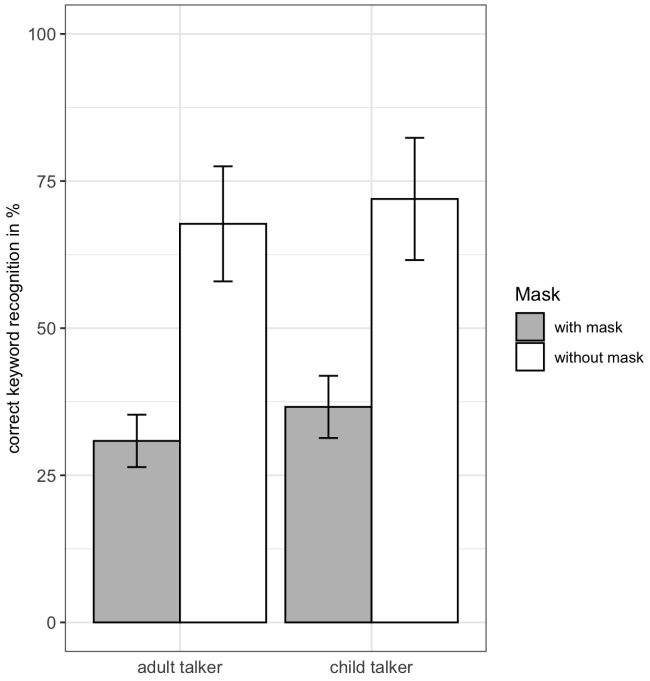

Figure 5.2.2: Average intelligibility scores for the adult and child talker in the conditions with and without face mask. The vertical bar represents standard errors.

The analysis showed a significant main effect of face mask (b = -1.70, SE = 0.1, t = -16.3, $p < .001$) and of rms power (b = 51.15, SE = 12.28, t = 4.2, $p < .001$) on listeners' keyword recognition accuracy. Listeners recognized considerably fewer keywords accurately when the talkers had been wearing a mask (adult talker 31% correct; child talker 37% correct) compared to when the talkers had not been wearing a mask (adult talker 68% correct; child talker 72% correct). No main effect of talker was found (b = -0.05, SE = 0.15, t = -0.3, p = 0.7), indicating that word recognition was comparable for both talkers. There was also no significant interaction between *face mask* and *talker* (b = -0.16, SE = 0.11, t = -1.5, p = .14).

Table 5.2.1: Full output of the LMER model

Fixed effects	*Estimate*	*Std. Errors*	*t*	*p*	
Intercept	-0.06944	0.43211	-0.161	0.872	
Face mask	-1.70015	0.10458	-16.256	<2e-16	***
Talker	-0.04802	0.14751	-0.326	0.745	
Sentence Duration	-0.19701	0.12545	-1.570	0.116	
RMS	51.15275	12.28217	4.165	3.12e-05	***
Face mask x Talker	-0.16733	0.10969	-1.525	0.127	

*Note *p<.05 **<.01 ***<.001*

The results thus suggest that intelligibility was considerably hampered when talkers were wearing a face mask, and this was equally true for the adult and child talker. This leaves open the question if the face masks produced an additional listening effort which came at the expense of memory encoding. To test for this possibility, we tested a new group of adult listeners using a cued-recall paradigm in Experiment 1.

Experiment 2

Methods

Participants

Eighty native German listeners between the ages of 19 and 56 years participated in the study (mean; 23.6; SD = 5; 63 females) for a chance to take part in a lottery.[3] All participants reported that German was their first and dominant language. Participants were recruited through social media and university email. Two participants had to be excluded from further analyses since they did not follow the instructions. None of the participants reported hearing or vision impairments, and none had participated in Experiment 1. As in Experiment 1, half of the

[3]A subset of 32 participants listening to the adult speaker had been reported in **Chapter 5.1**

participants watched the videos produced by the adult talker, and the other half watched the videos produced by the child talker.

Stimuli

Sentence recordings were identical to Experiment 1, but this time, the sentences were presented without noise. The 48 experimental sentences were divided into eight blocks of six sentences each. Sentence order was randomized once, and half of the participants watched the videos in the reverse order. The presence of a face mask was blocked, and blocks alternated between the mask and no-mask condition. The order of mask condition was counterbalanced, and sentences were presented with an interstimulus interval (ISI) of 2500 ms.

Procedure

As in Experiment 1, participants took the online experiment on *Alchemer*. Participants were instructed to wear headphones and take part in the experiment on a computer, laptop, or tablet. Furthermore, participants indicated after the experiment the type of device and headphones they had used. Before the experiment started, participants digitally signed written informed consent.

The experiment began with two practice trials during which participants could adjust the volume. After the practice trials, participants were asked not to change the volume for the 48 experimental sentences. The self-paced cued-recall task followed immediately after a block. For this task, sentences were presented up to the adverb orthographically on the screen (e.g., *Die Köchin hilft montags*, "The cook helps on Mondays"), and participants were asked to type in the missing two final words (e.g., armen Kindern, "poor children") on their keyboard. Block length and cue length were determined based on informal pre-tests that yielded an intermediate level of recall rates. The recall cues were shown after each video block,

in the order of block presentation, and participants could fill in their responses in any order. Then participants could press a button to initiate the next video block of six sentences. The whole experiment lasted approximately 20-30 minutes. After the experiment, participants filled out a short language background questionnaire and were asked about technical problems of which none were reported.

Results

Scoring was again done by T.L.T. and a research assistant. For any remaining uncertainties A.W. was consulted. Each correctly recalled word received the score correct (1), while incorrectly recalled words or unrecalled words received the score incorrect (0) (see Figure 5.2.3).

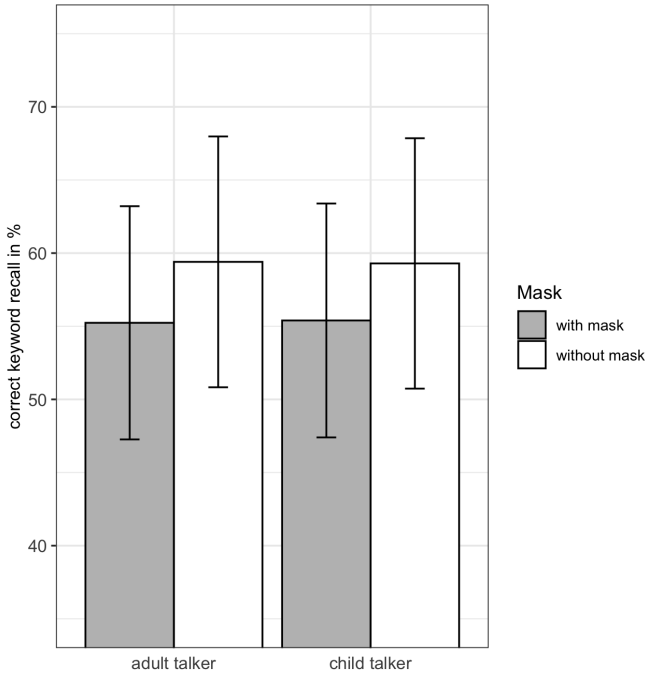

Figure 5.2.3: Average keyword recall scores for the adult and child talker in condition with and without face mask. The vertical bar represents standard errors.

There were two keywords for each of the 48 sentences, making a total of 96 keywords to be recalled per participant. Overall, 57.3% of the words were recalled correctly and 42.7% incorrectly. Descriptive analyses of the incorrectly recalled words showed that most incorrectly recalled words had been complete omissions of a keyword (68% for the adult talker and 73% for the child talker). The remaining incorrect responses consisted of a variety of error types. Some responses were unrelated in form and in meaning to the intended words (e.g., *schwarze Schuhe*, "black shoes," for *staubige Kissen*, "dusty pillows"), fewer responses were closely semantically related (e.g., *Ringe*, "rings," for *Kreise*, "circles"), and a small number of responses consisted of phonetic errors involving a single sound change, like substitution, insertion, or deletion (e.g., *Schweine*, "pigs," for *Steine*, "stones") or typos (e.g., the nonword *Lmpen* for *Lampen*, "lamps").

Next, a logistic mixed-effects regression model with the lme4 package in R (R Core Team 2021, version 4.0.5) was employed to assess the effect of face masks on listeners' correctly recalled keywords (Bates, Kliegl, et al., 2015). *Keyword recall* (success vs. failure) was the dichotomous dependent variable, and *talker age* (adult vs. child), *face mask* (mask vs. no mask), and *block* (8 blocks) were the independent variables; *face mask* x *talker* x *block* was added as an interaction term. To test linear and quadratic effects of block, orthogonal polynomials were used (Mirman, 2017). The same fixed effects and random intercepts were included as in Experiment 1. The values of the lmer model output are displayed in Table 5.2.2.

Table 5.2.2: Full output of the LMER model

Fixed effects	*Estimate*	*Std. Errors*	*t*	*p*	
Intercept	2.29695	0.67148	3.421	0.000625	***
Linear	0.29061	0.25672	1.132	0.257626	
Quadratic	-0.45767	0.27898	-1.641	0.100900	***
Face mask	-0.30180	0.08407	-3.590	0.000331	
Talker	-0.54577	0.28052	-1.946	0.051704	.
Sentence Duration	-0.36209	0.14667	-2.469	0.013556	*
RMS	-37.73710	17.37923	-2.171	0.029902	*
Linear x Face mask	-0.28000	0.17889	-1.565	0.117540	
Quadratic x Face mask	-0.11714	0.18072	-0.648	0.516866	
Linear x Talker	0.28436	0.29942	0.950	0.342262	
Quadratic x Talker	0.07903	0.22118	0.357	0.720868	
Face mask x Talker	0.07066	0.15374	0.460	0.645783	
Linear x Face mask x Talker	-0.03198	0.31653	-0.101	0.919529	
Quadratic x Face mask x Talker	0.53551	0.32259	1.660	0.096911	.

*Note *p.05 **.01 ***.001*

The analysis showed a significant effect of *face mask*, with listeners recalling fewer words when the talkers were wearing a mask (adult talker 55.2%; child talker 55.4%) compared to when the talkers were not wearing a mask (adult talker 59.4%; child talker 59.3%) (b = -0.30, SE = 0.08, t = -3.5, p = .0003). There was a marginal effect for *talker age* (b = -0.54, SE = 0.28, t = -1.9, p = .05), indicating a trend for a better recall rate for the child talker than for the adult talker. Recall performance was better for shorter sentence durations than for longer ones as the main effect for *sentence duration* showed (b = -0.36, SE = 0.15, t = -2.5, p = .01). Further, a main effect for *rms* (b = -37.7, SE = 17.4, t = -2.2, p = .03) was found, indicating that sentence recordings with less rms power were recalled better than sentences with higher rms power. There was no significant effect for *block*, and there were no interactions (all p-levels < .1).

This result suggests that processing was easier when visual articulatory cues were available than when they were not present, and this availability left more

cognitive resources for successful memory encoding. Overall, participants' recall performance was not worse for the child talker than for the adult talker. It thus appears that listening to the child's voice did not negatively affect recall.

General discussion

The current study expanded on the results from **Chapter 5.1** which found that face masks can significantly impede the recall of sentences spoken by a native adult. To broaden the scope of **Chapter 5.1** findings, we investigated the impact of face masks on speech intelligibility and memory for adult and child speech. Adult listeners watched video clips of either an adult or a child talker producing sentences with and without a face mask embedded in noise (Experiment 1, intelligibility task) or in quiet (Experiment 2, cued-recall task). For both the intelligibility task and the recall task, it was found that performance was worse when the talkers were wearing a face mask than when there was no mask.

Interestingly, the response patterns to the child talker did not differ substantially from the responses to the adult talker. The child talker in the current study was nine years old and contrasted in her average F0 from the adult speaker by 53.2 Hz. It was easy to notice that she was a child, and there were various reasons why talker age could have affected the results and impeded responses to the child talker. Firstly, it was possible that white noise may have masked the high frequencies of the child talker more effectively than that of the adult talker. Secondly, children at that age can still vary more in their pronunciation than adults, and their formant values are higher across the board (Hillenbrand et al., 1995). It is also conceivable that listeners previously experienced children as talkers who regularly deviate from canonical pronunciations and this experience could have been taken into

consideration during comprehension of children's speech (see e.g., **Chapter 5.1**). However, neither recognition rates in Experiment 1, nor recall rates in Experiment 2 were lower for the child talker than for the adult talker. Likewise, the size of the mask-effect did not differ for the two talkers, even though listeners may depend more heavily on visual speech cues in difficult listening situations (Xie et al., 2014), and concealing these cues with a face mask could have been particularly detrimental for the child talker.

The only difference between the two talkers was marginally better recall rates for the child talker than the adult talker in Experiment 1. This difference could be related to the overall longer sentence durations of the child talker which possibly enhanced memory encoding. Howsoever, the present data clearly indicate that in terms of a mask-effect, intelligibility and recall responses did not differ for the two talkers. There was thus never a disadvantage for the child talker. Since this is the first study we know of on the intelligibility and memory for child talkers, we do not know what would happen with younger talkers. The current talker was nine years old, and even though her pronunciation probably still varied more than an adult's pronunciation, there were no clear mispronunciations as one would find for younger children. Also, younger children have even higher average F0s. It is thus still possible that differences in intelligibility and recall, as well as a modulation of the mask-effect, would emerge for talkers younger than nine or in different communicative settings.

This brings us to the question of why face masks impeded recognition and recall in the current study? Face masks can change the acoustic signal, but also hide visual articulatory information. While the current study was not set out to tease apart these two possible sources for mask-effects, we can still speculate on the primary reason for the observed impediment on recognition and recall. Previous work has shown that face masks have the potential to change the acoustic signal;

observed changes range from negligible to substantial and depend on both the mask material and microphone position (e.g., Bottalico, Murgia, Puglisi, Astolfi, & Kirk, 2020; Corey et al., 2020; Magee et al., 2020). In the current study, spectral analyses showed no difference in rms values between the mask and no-mask condition for the adult talker, and a small (< 1 dB) but significant difference for the child talker. Despite the rms differences for the child talker, responses to the child talker were seemingly not less accurate. If the acoustic signal itself would have been strongly affected by the mask, it should also be harder to understand the audio recordings in quiet.

We therefore presented in a post hoc test an additional 12 participants with the audio recordings in quiet and asked them to write down the final two keywords after each sentence. Word recognition rates were overall very high and did not differ between the mask (adult 97.9% correct keywords; child 99.3% correct keywords) and no-mask condition (adult 97.4% correct keywords; child 98.6% correct keywords). This ceiling effect in combination with the small differences in rms, make it unlikely that the signal itself was changed dramatically by the mask, and that the missing visual cues due to the face masks were potentially the primary reason for the decrease in performance.

While our findings from Experiment 1 are in line with the intelligibility results of Smiljanic et al. (2021), the results of Experiment 2 are at first glance in contrast with their results. In our recall experiment, for which the sentences were presented in quiet, we found a negative impact of face masks for both the adult talker (in line with **Chapter 5.1**) and the child talker. Smiljanic et al. (2021), however, found a negative mask-effect for an adult talker only when noise was added with a negative SNR, but not when sentences were presented in quiet. There are, however, several methodological differences across the two studies which could

well explain the difference in findings. The two most important ones are arguably the employed materials and the type of memory task. While the current study used dissociated sentences with low predictability, to avoid facilitatory influences of context, Smiljanic et al. (2021) used sentences from a coherent text, which made the listening environment more naturalistic. Encoding of cohesive information is, however, easier than encoding of dissociated information (Black & Bern, 1981), and this alleviation through cohesion possibly prevented a mask-effect when listening in quiet in Smiljanic et al. (2021). Also, in their memory task, Smiljanic et al. (2021) included different question types, ranging from fill in the blank, and true/false, to close questions. These questions might well tap into different memory processes and could explain the difference in findings with the present study which only tested recall memory with a fill in the blank task.

In summary, this study examined the effect of face masks on intelligibility and recall produced by adult and child talkers, and it makes an important contribution to the field's current understanding of the impact of face mask on speech perception. First, we found that face masks impede intelligibility and recall for both adult and child talkers equally. This finding should have implications in various communication situations, such as in classrooms, where information has to be understood and retained. Second, we established that intelligibility and recall were not worse for the child talker, certainly highlighting an encouraging observation that masked child speech is not disadvantaged more than adult speech in face-to-face communication.

To our knowledge, the present work provided the first investigation of intelligibility and recall of sentences produced by adult and child talkers wearing face masks, laying a solid foundation for future research examining how face masks influence speech understanding for various talker and listener groups. Wearing a face

mask is an essential means to slow the spread of Covid-19, and to further advance our understanding of the potential impact of face masks on communication is one step toward a better understanding of the impact of the pandemic.

CHAPTER 5.3

L2 recall of sentences spoken by adult and child talkers wearing face masks

Abstract

During the Covid-19 pandemic, concerns have been raised about the impact of face masks on communication for native listeners when perceiving speech. Research findings have repeatedly shown since then that, for native listeners, speech without a face mask is easier to understand and easier to remember. Thus, native listeners experience an audiovisual benefit when signals from both modalities are available. However, it is unclear whether non-native listeners experience a similar benefit when listening to speech produced with a face mask. Non-native listeners of German watched video recordings of a native German adult talker and a native German child talker producing German sentences with and without a face mask. Subsequently, they completed a cued-recall task. Findings showed no significant differences in the mask conditions (face mask vs. no mask) or in the talker age conditions (adult vs. child). The findings indicated that non-native listeners, in contrast to native listeners, did not gain an audiovisual benefit when information from both modalities was available. Non-native listeners in the present study had a medium level of proficiency, and it is possible that only listeners with a higher language proficiency in the second language can benefit from an audiovisual context.

Introduction

Understanding speech can be challenging for non-native listeners (henceforth, L2) especially when the listening environment is degraded (e.g., a noisy main street, the holiday office party, a busy restaurant) (Mattys, Davis, Bradlow, & Scott, 2012). Previous studies have examined how native adult listeners overcome such challenging listening conditions and have found that they can utilize information from both the auditory and visual modality in order to mitigate the effect of a noisy environment (e.g., Campbell, 2008; Jesse & Janse, 2012; Massaro, 1987; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby & Pollack, 1954; Summerfield, 1992). Visual speech information, such as lip and jaw movements, can contribute crucial phonological information about speech sounds (e.g., Campbell, 2008; Summerfield, 1992) and help their comprehension (Navarra & Soto-Faraco, 2007; Sumby & Pollack, 1954). For example, while closed lips indicate a bilabial place of articulation (e.g., /p/ and /b/), an open jaw indicates vowel height (e.g., more open jaw for the vowel /a/ and less open jaw for /i/). Thus, visual information can supplement and complement information about speech sounds that is not included in the auditory signal itself (Massaro, 1987; Sumby & Pollack, 1954). The investigation of audiovisual speech perception has primarily focused on native talkers and listeners (Jesse & Janse, 2012; Ross et al., 2007; Sumby & Pollack, 1954), which leaves audiovisual speech perception for L2 listeners an understudied area.

Some studies, however, have indeed observed already that for certain tasks, L2 listeners can make use of visual information to enhance the perception of speech. These studies showed, for example, that the presence of visual information helped English participants to a more accurate perception of foreign accents such as for French (Reisberg et al., 1987), Korean (Davis & Kim, 2004), Irish and Spanish

(Erdener & Burnham, 2005). Therefore, visual speech information can facilitate the perception of L2 listeners (Massaro, 1998), and they might in fact pay more attention to the information that is conveyed by visual articulatory movements to compensate for their poorer comprehension skills in the non-native language (Drijvers & Özyürek, 2020).

Theories of the acquisition of L2 auditory sounds in fact show similarities to the perception of L2 visual information. L2 auditory sound perception theories (e.g., the speech learning model (SLM) Flege, 1995; the perceptual assimilation model (PAM), Best, 1995) state that L2 listeners have more difficulties perceiving L2 sound contrasts relative to L1 listeners because of the differences in sound inventories of the native and non-native language (Best, 1995; Flege, 1995). Not only do L2 listeners have to develop L2 speech categories, they also have to develop L2 visual categories which are also known as *visemes*. Visemes are the visual pendant to phonemes. While phonemes underlie the representations of speech sounds (Wells, 1977), visemes categorize phonemes based on their visual distinctiveness (Fisher, 1968). In this respect, the pattern for L2 auditory perception may also be applied to L2 visual speech perception, because L2 listeners must learn to associate visual cues in the L2 that are not present in their L1. Therefore, L2 listeners need to attune to visual cues to establish new L2 categories (Hazan et al., 2006; Hazan, Sennema, Iba, & Faulkner, 2005). Given the fact that L2 listeners are typically less proficient than L1 listeners, L2 listeners are overall less efficient at using visual information compared to L1 listeners (Drijvers & Özyürek, 2020). Nonetheless, the acquisition of visual information can be enhanced through language experience, as experience may play a crucial factor in modulating the utilization of audiovisual cues in L2 speech. This has been shown on the sound level (Navarra & Soto-Faraco, 2007; Wang, Behne, & Jiang, 2008) and the sentence level (Xie et al., 2014).

For example, Navarra and Soto-Faraco (2007) study presented the Catalan vowel contrast /ɛ/-/e/ in audio-only or audiovisual conditions to Spanish-dominant bilinguals and Catalan-dominant bilinguals. In the audio-only condition, Catalan-dominant bilinguals could unsurprisingly distinguish between the vowel contrast given that they were already highly experienced with the vowel contrast as they grew up with it. However, Spanish-dominant bilinguals could not discriminate the Catalan vowel contrast in the audio-only condition, but when visual articulatory information was included, both participant groups were able to perceive the contrast. These results indicated that the sensitivity to visual information was correlated to the amount of experience in the L2. Additionally, Wang et al. (2008), for example, showed an improvement of English phoneme identification by Mandarin Chinese listeners when visual information was available. Specifically, the experimental items consisted of three distinct English fricatives: interdentals, labiodentals, and alveolars. While interdentals are non-existent in Mandarin Chinese, the other two are relatively common in both languages. As expected, the size of the audiovisual benefit was smaller for the Mandarin Chinese listeners compared to L1 English listeners, because visual cues from the second language may differ from the first language. Based on this explanation, the researchers argued that the smaller size of the audiovisual benefit for L2 listeners can be explained by their lack of ability to integrate visual and auditory cues in their second language. Wang et al. (2008) additionally demonstrated that longer exposure to English enhanced Mandarin listeners' identification of English phonemes in the audiovisual condition. Specifically, they tested Mandarin Chinese listeners who had been living in Canada for 2 years (short LOR) or 10 years (long LOR). Their results showed that the long LOR group outperformed the short LOR group, emphasizing the importance of linguistic expertise in audiovisual speech perception.

These findings, therefore, suggested the potential role of linguistic expertise in modulating the extent of audiovisual benefit in L2 listeners. A further study that reinforced this notion was conducted by Xie et al. (2014), who found that the language background of the talker and listener can interact to modulate the audiovisual benefit for the intelligibility of speech. Xie et al. (2014) investigated whether visual information facilitates speech perception in noise using a sentence transcription task. Participants listened to sentences (e.g., "The gray mouse ate the cheese"), which were mixed with noise. After each sentence, participants were asked to write down what they had heard. Results showed an audiovisual benefit for Korean L2 listeners of English when audiovisual information of the English talker was available. In other words, Korean participants with higher English proficiency levels showed higher accuracy in speech recognition of native English speech.

Since the Covid-19 pandemic, face masks can present an additional challenge for L2 listeners. Overall, results so far suggest that the difficulties listeners encounter when listening to speech produced with a face mask are likely to stem from both the acoustic degradation of the speech signal and the lack of visual information of the talker's mouth movements. Face masks can impede the acoustic properties of the speech signal such that higher frequencies are particularly negatively affected since face masks function similarly to a low-pass filter (Bottalico et al., 2020; Corey et al., 2020). Specifically, recent studies showed varying attenuation effects for different types of face masks. For instance, cloth face masks attenuate the sound by 3-4 dB and the N95 face masks attenuate the sound by 12 dB (Goldin & Weinstein, 2020). In addition to that, face masks create a visual barrier to the mouth region of a talker's face, thus restricting access to visual articulatory information that can be particularly helpful in language comprehension when listening conditions are adverse (e.g., noisy background) (Campbell, 2008; Summerfield, 1992).

The pandemic has promoted multiple researchers to investigate the effect of different face masks on speech intelligibility with noise levels ranging from moderate to high levels of noise. Previous studies, for example, showed that all types of face masks increased listening effort and reduced native adults' correct identification of words and sentences under noisy listening conditions (for details see, e.g. Bottalico, Murgia, Puglisi, Astolfi, & Kirk, 2020; V. A. Brown, Van Engen, & Peelle, 2021; Randazzo, Koenig, & Priefer, 2020). Besides the negative impact face masks on speech intelligibility, masks can also have an impact on native adults' recognition memory, even in quiet listening conditions (Truong & Weber, 2021). For instance, participants in Truong and Weber (2021) remembered fewer words when adult and child talkers were wearing a mask compared to when they were not wearing one (Smiljanic et al., 2021; Truong et al., 2021; Truong & Weber, 2021). This effect has also been reproduced for speech produced by non-native talkers (Smiljanic et al., 2021).

In line with the Framework for Understanding Effortful Listening (Pichora-Fuller et al., 2016), the Effortfulness Hypothesis (McCoy et al., 2005), and the Ease of Language Understanding (Rönnberg et al., 2013), the above mentioned findings clearly demonstrate that face masks influence language comprehension such that listening effort increases, thus invoking higher cognitive load. More cognitive resources are utilized for speech comprehension when listening conditions are adverse (i.e., speech produced with a face mask), consequently fewer resources are available for retaining information in working memory (McCoy et al., 2005; Peelle, 2018; Pichora-Fuller et al., 2016; Rönnberg et al., 2013), which is fueled by limited cognitive resources (Just & Carpenter, 1992). In contrast, when speech is presented in an optimal listening environment (e.g., speech produced without face mask in a quiet environment), processing effort is low which leaves more cognitive resources

available for memory encoding of spoken information.

With regard to L2 memory, Just and Carpenter's (1992) *capacity theory* posits that L1 listeners have more cognitive resources available when listening to spoken language compared to L2 listeners for the reason that L1 listeners have a higher command of the language and more linguistic knowledge. Bearing that in mind, L1 listeners usually process spoken language automatically and efficiently, with little conscious attention to individual words. Contrarily, L2 listeners have typically less linguistic knowledge, leading to less automatized and efficient processing of the speech input, because L2 listeners need to allocate their attention more to individual words and sounds. This can come at the expense of remembering less well what they hear, given the limited capacity in working memory and the difficulty to process all information within a certain amount of time (Vandergrift, 2007). Hence, to compensate for their lack of L2 knowledge, L2 listeners might make use of visual information conveyed by the lips of the talker. As mentioned earlier, previous work has indeed demonstrated that visual articulatory cues can improve L2 language learning and comprehension (Davis & Kim, 2004; Hazan et al., 2006; Wang et al., 2008).

To our knowledge, the effects of face masks on L2 processing have not been examined. Note that the above mentioned face mask studies have focused on the perception of masked speech from a native adult listeners' perspective (V. A. Brown et al., 2021; Randazzo et al., 2020; Smiljanic et al., 2021; Truong et al., 2021; Truong & Weber, 2021) and native children (J. Schwarz et al., 2022). This leaves open the question of how face masks affect L2 listeners.

The present study, addressed this research gap with a similar methodology used in Truong et al. (2021) to capture non-native participants' memory performance

of sentences produced by a child talker and an adult talker when the talkers were wearing a face mask or not. Recall that in Truong et al. (2021) and Truong and Weber (2021) native adults' memory performance was worse when the talkers were wearing a face mask compared to when they were not wearing one, and this was equally true for both adult and child talkers.

We predicted a similar pattern for non-native adults, possibly, however, with an overall worse performance for the child talker since children's acoustic and linguistic characteristics deviate from adult norms. Child speech, therefore, may have an impact on L2 listeners' ability to decode and encode what has been said by a child. Children's speech is described to be overall higher in fundamental frequency (F0) compared to native adult speech (B. Smith et al., 1983; Tingley & Allen, 1975), because children's anatomical characteristics are different from adults such that their larynx and vocal folds are smaller in size and shorter in length. For example, the vocal folds of an eight-year-old child are about 8 mm long, whereas the vocal folds of an adult are approximately 12-21 mm long. These physiological differences alone make the voices of children and adults notably different in F0 (Kreiman & Sidtis, 2011).

Even though the child talker did not pose a particular listening challenge for L1 listeners in Truong et al. (2021) (Chapter 5.1) and Truong and Weber (2021) (Chapter 5.2), child speech may still pose a challenge for L2 listeners due to their lower competence in the second language. As children's voices are characterized by higher F0, it is also conceivable that L2 listeners have a harder time understanding masked child speech than masked adult speech. Since children's voices are higher than adult voices (Vorperian & Kent, 2007) and face masks attenuate those frequencies more strongly, child talkers could be less intelligible than adult talkers for L2 listeners. In addition, restricting access to visual articulatory

information through a face mask could further increase listening effort for L2 listeners as they cannot utilize the visual articulatory information to enhance comprehension, which in turn may affect memory encoding.

Taken together, the current study built on the experiments in Chapter 5.1 and 5.2. L2 listeners of German were presented with the same stimuli and using the design of Truong et al. (2021). The goal was to examine the effect of wearing a face mask on L2 listeners of German on subsequent recall of sentences spoken by a child talker and an adult talker.

Experiment

Methods

Stimuli

The materials were identical to those of Chapter 5.1 and 5.2.

Participants

Eighty native English listeners between the ages of 18 and 61 years participated in the study (mean; 30.9; SD = 11.8; 52 females and 1 undisclosed). Participants were recruited and paid via Prolific. All participants indicated that English was their first and dominant language and German their second language. All participants reported normal or corrected to normal vision (i.e., glasses and contact lenses). With the exception of two, participants reported no hearing impairments.[1] Four participants were excluded from further analyses because they did not complete the experiment.

[1]One participant reported having a hearing aid and the other uncorrected hearing impairment. Performance did not differ from the other participants.

We used the pre-screening function of Prolific, which ensured that only participants who registered themselves with an intermediate or advanced level of German were allowed to participate. Participants' average current daily use of German ranged from 0-90%; three reported currently living in Germany, forty reported having lived in Germany before (min: two months, max: seven years). Additionally, German language proficiency was measured via self-report on a scale from 1 (very poor) to 7 (very good) for all four modalities (i.e., writing, listening, speaking, reading). The average score was 4.87 (SD = 1.35). Half of the participants watched the videos produced by the adult talker, and the other half watched the videos produced by the child talker.

Procedure

The experiment was created and hosted on the Gorilla Experiment Builder platform (www.gorilla.sc) (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2019). Participants were recruited through Prolific. The online experiment consisted of four phases: (1) ethics and consent, (2) a headphone screening test, (3) the experiment, and (4) a language background questionnaire.

Ethics and consent Prior to the experiment, participants signed a written informed consent. Then they were given instructions in English, asking them to take part in the experiment in a quiet room, to eliminate distractions (e.g., turn off smartphones), and to wear a pair of headphones (e.g., over-the-ear or in-ear headphones were recommended) during the entire experiment.

Headphone screening test A headphone-screening task (i.e., 3AFC paradigm) was implemented prior to the actual experiment (Woods, Siegel, Traer, & McDermott, 2017). This task enabled high accuracy in performance when using headphones compared to using loudspeakers. Participants heard three intervals of

randomly ordered white noise with equal frequency and duration, but one interval contained a Huggin's Pitch tone. Listeners were asked to identify which of the three tones contained the Hugging' Pitch. There were in total six trials and listeners were required to detect at least 5 out of 6 correctly in order to proceed to the experiment. Those who passed the test were immediately directed to the experiment, while those who failed were automatically rejected. This task simultaneously checked the Autoplay function on the browser. If Autoplay did not work, the task also showed instructions on how to enable it. In addition, participation was only permitted when doing the experiment on a laptop or a computer with browsers such as Firefox, Google Chrome, and Safari. Participants were automatically rejected if the technical requirements were not met. Stimuli were loaded prior to the experimental session to prevent loading issues during trials. None of the participants reported technical problems.

Procedure of the experiment Although the layout differed from Truong et al. (2021), the experimental procedure was identical. After the experiment, participants were asked to fill in a language background questionnaire, including questions about their German language proficiency in writing, listening, speaking, and reading.

Results

As in Truong et al. (2021), keyword scoring was done by T.L.T. and a research assistant. For any remaining uncertainties, A.W. was consulted. Each correctly recalled word received the score correct (1) and each erroneously recalled word received the score incorrect (0) (see Figure 5.3.1). Some of the verbatim scoring rules used in Truong et al. (2021) needed to be adjusted for the present study. For example, typing errors for umlauts with two letters were accepted as correct (e.g., *Veogel*, "birds", for *Vögel/ Voegel*), because this error might have occurred

due to the different usage of keyboard layout. That is, US or UK keyboards use the QWERTY keyboard layout, which does not include umlauts like the German QWERTZ keyboards. There were two keywords for each of the 48 sentences, making a total of 96 keywords to be recalled per participant (see Figure 5.3.1).

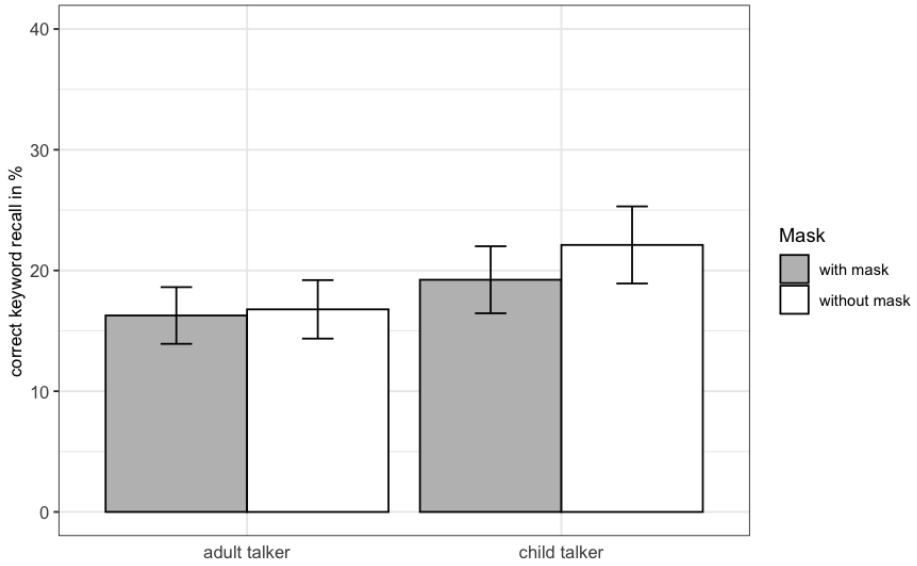

Figure 5.3.1: Average keyword recall scores for the adult and child talker in condition with and without face mask of L2 listeners. The vertical bar represents standard errors.

Overall, only 18,7% of the words were recalled correctly and 81.3% incorrectly. Descriptive analyses of the incorrectly recalled words showed that most incorrectly recalled words had been complete omissions of a keyword (49.4% for the adult talker and 46.1% for the child talker). The remaining incorrect responses consisted of a variety of error types. Some responses were unrelated in form and in meaning to the intended words (e.g., *schwarze Schuhe*, "black shoes," for *staubige Kissen*, "dusty pillows"), phonetic errors involving a single sound change, like substitution, insertion, umlaut mistakes, or deletion (e.g., *Bieren*, "biers," for *Beeren*, "berries"), fewer responses were closely semantically related (e.g., *Dackel*,

"Dachshund," for *Hunde*, "dogs") and an even smaller number of responses consisted of typos (e.g., the nonword *schawrze* for *schwarze*, "black"). The values of the lmer model output are displayed in Table 5.3.1.

Table 5.3.1: Full output of the LMER model

Fixed effects	*Estimate*	*Std. Errors*	*t*	*p*	
Intercept	-8.12129	2.15773	-3.764	0.000167	***
Linear	-0.23873	0.42149	-0.566	0.571128	
Quadratic	-0.61831	0.41388	-1.494	0.135188	
Face mask	-0.10419	0.09724	-1.071	0.283967	
Talker	-1.00025	0.92914	-1.077	0.281688	
Language proficiency	0.21065	0.09520	2.213	0.026911	*
Sentence Duration	-0.08275	0.23758	-0.348	0.727632	
RMS	13.84378	24.60437	0.563	0.573670	
Linear x Face mask	-0.08150	0.25671	-0.317	0.750879	
Quadratic x Face mask	0.03346	0.24691	0.135	0.892217	
Linear x Talker	0.24546	0.31533	0.778	0.436313	
Quadratic x Talker	-0.18395	0.30109	-0.611	0.541221	
Face mask x Talker	0.30377	0.18050	1.683	0.092394	.
Linear x Face mask x Talker	0.28034	0.48569	0.577	0.563797	
Quadratic x Face mask x Talker	-0.29765	0.47093	-0.632	0.527347	

*Note *p.05 **.01 ***.001*

For the statistical analysis, a logistic mixed-effects regression model with the lme4 package in R Core Team (2021) (version 4.0.5) was employed to assess the effect of face masks on listeners' correctly recalled keywords (Bates, Kliegl, et al., 2015). *Keyword recall* (success vs. failure) was the dichotomous dependent variable, and *talker age* (adult vs. child), *face mask* (mask vs. no mask), and *block* (8 blocks) were the independent variables; *face mask* x *talker* x *block* was included as an interaction term. To test linear and quadratic effects of block, orthogonal polynomials were used (Mirman, 2017). Additionally, to account for additional variation, fixed factors of *sentence duration*, *rms power*, and *language proficiency*

were also included in the model. The German proficiency factor consisted of the sum of self-reported ratings (i.e., from 1 the lowest to 7 the highest) in the category of writing, listening, speaking, reading. The analysis showed a significant effect for *language proficiency* ($b = 0.21$, $SE = 0.1$, $t = 2.2$, $p = .03$), showing that participants with higher language proficiency performed better than those participants with lower language proficiency. There were no significant effects for face mask (mask vs. no mask) (b = -0.1, SE = 0.1, t = -1.1, p = 0.3) and for talker age (adult vs. child) (b = -1.0, SE = 1, t = -1.1, p = 0.3). There were no interactions (all p-levels < 0.1).

The findings suggest that neither the covering of visual articulatory cues with a face mask nor listening to a child's voice did negatively influence recall performance. However, it is important to mention that the overall performance was very low and could be considered at floor. Overall, correct scores were around 16.5% for the adult talker and 20.7% for the child talker.

Discussion

The current study aimed to get a better understanding of the impact of face masks on memory for spoken language since face mask studies predominantly focused on native listeners (V. A. Brown et al., 2021; Randazzo et al., 2020; Smiljanic et al., 2021; Truong et al., 2021; Truong & Weber, 2021). Specifically, the present study examined whether memory recall performance of L2 listeners is negatively affected by face masks as was shown for L1 listeners (Truong et al., 2021), using the same experimental design as in Truong et al. (2021). This time, native English listeners watched video recordings of either a native German adult or a native German child talker producing sentences with and without a face mask in quiet listening conditions. Results of the present study showed no influence of face masks or talker

age on recall. This is in contrast to Truong et al. (2021), who found for native adult participants that face masks significantly decreased the recall of sentences for both adult and child talkers equally. While in principle, this could imply that L2 listeners are not using visual speech cues for encoding, the overall low performance leaves open the possibility that no difference between conditions was found, because performance was at floor. Interestingly, the present data showed an effect of proficiency with higher proficient participants recalling more words than lower-proficient participants. Thus, L2's proficiency in German mediated performance, including possibly the null-effect in the talker age condition (adult vs. child) and face mask condition (face mask vs. no face mask).

As for the effect of talker age, the initial hypothesis was that L2 listeners have more difficulties when perceiving the child talker, because children vary in their pronunciation and acoustic features from the standard norms of native adult speech (Hillenbrand & Clark, 2009; Vorperian & Kent, 2007). Even though L1 listeners showed no decrease in performance for the child talker in Truong and Weber (2021), L2 listeners' incomplete mastery of the German language could make it harder for them to understand a child that deviates in the acoustic realization from the adult norms. Present findings, however, showed no difference in response patterns, suggesting that L2 listeners did not experience any difficulties when perceiving child speech compared to adult speech. At first glance, this result is in line with Truong et al. (2021) and Truong and Weber (2021), who found no difference between the child and adult talkers for native adult participants. However, interpretation must again be proceeded with caution, because recall performance in the present experiment was overall low (16.5% correct for the adult talker and 20.7% for the child talker), reaching an apparent floor effect. It can be ruled out that the floor effect was caused by a low quality of the speech stimuli, since the same stimuli as in Truong et al.

(2021) and Truong and Weber (2021) were used. Hence, factors other than the stimuli quality caused this floor effect.

There are to date no published journal articles that report on the perception of child speech by L2 listeners, but a poster was recently presented by Konopka and colleagues at the online ISBPAC Symposium in 2021 gave first insights into the processing of adult and child speech by L1 and L2 listeners.[2] Konopka et al. (2021) examined the memory performance for adult and child speech by L1 and L2 adult listeners and tested whether memory depends on the linguistic proficiency of the listeners. English L1 and L2 listeners participated in the experiment which consisted of (1) a study phase and (2) a test phase, consisting of two blocks. During the study phase, participants studied pictures with recorded descriptions produced by either an adult or a child talker who were both native talkers of English. In the following test phase, participants were presented with new (unstudied items) and old items (studied items from the previous phase). Participants then indicated whether the items presented were new or old items. Overall, descriptions produced by the child talker were remembered poorly compared to that of adult talkers. Most intriguing, a difference in performance was noticeable for the L2 listeners, meaning that they improved recalling items correctly only for adult speech in the second block - not for child speech. This suggests that L2 listeners possibly processed adult speech more carefully than child speech (i.e., superficially). Interestingly, L2 listeners with higher linguistic proficiency in their L2 showed greater improvement, indicating that memory in L2 was modulated by L2s' linguistic proficiency. The results showed that memory performance of L2 listeners is more sensitive to talker identity and cognitive load than L1 memory and that this effect is modulated by L2 listeners' language

[2]This poster, with the title "Native and non-native listeners differ in their memory for adult and child speech", was presented at the third International Symposium on Bilingual and L2 Processing in Adults and Children.

proficiency.

Intrigued by this result, it gave reason to suspect that differences in recall would have emerged for both adult and child talker if the participants of the present study were more proficient in German. It appeared that our L2 listeners' language competence was too low to trigger any sensitivities in the talker age condition, possibly explaining why the L2 listeners did not recall fewer words when the talker was a child compared to when the talker was an adult. Based on this premise, it is rather likely that the potential cause for the absence of the face mask effect may also be attributed to the low proficiency level of the L2 listeners.

The null-effect in the face mask condition was rather unexpected since previous research has consistently shown that face masks hide visual speech information and can negatively affect the acoustic signal (Bottalico et al., 2020; Corey et al., 2020; Mendel et al., 2008). The combination of degradation of the speech signal and the absence of visual information of the talker's mouth movements creates an adverse listening situation. Listeners must then reallocate cognitive resources from memory back to perception, which in turn comes at a cost in terms of poor memory performance, as the working memory is fueled by limited cognitive resources (Peelle, 2018). Hence, while covering visual speech information with a face mask can further reinforce the impeding effect on memory recall, the presence of visual speech information is subject to believe to enhance recognition and the recalling of words. This effect has indeed been shown for native and non-native speech (Randazzo et al., 2020; Smiljanic et al., 2021; Truong et al., 2021; Truong & Weber, 2021) in younger and older native listeners (V. A. Brown et al., 2021).

Based on those previous findings, a similar pattern of recall performance for L2 listeners was predicted. In order to make up for the lack of L2 competence,

the prediction was such that L2 listeners make use of visual information to enhance perception (Davis & Kim, 2004; Erdener & Burnham, 2005; Reisberg et al., 1987). Specifically, the following pattern was expected: worse recall performance when the talker had been wearing a mask than when there was no mask. Nonetheless, we expected an overall poorer performance compared to L1 listeners (Truong et al., 2021; Truong & Weber, 2021), due to their incomplete mastery of the German language. However, this is not what was found. Instead, the results showed an overall floor effect, and this effect may be attributed to the low proficiency level of German of our L2 participants. Our L2 listeners' German language skills were apparently not proficient enough to associate L2 visual categories with the speech input.[3]

Generally, the extraction of information from both auditory and visual modalities has been repeatedly shown to enhance speech perception particularly in challenging listening conditions relative to auditory-only listening conditions for native listeners. Similar to L1 listeners, L2 listeners also utilize visual cues and therefore also experience a significant boost in speech perception. However, previous audiovisual studies indicated that language proficiency can modulate the usage of audiovisual information in speech recognition (Wang et al., 2008; Xie et al., 2014), reinforcing the notion that the higher the proficiency, the more efficiently listeners are able to extract information from visual cues (Drijvers & Özyürek, 2020). Hence, higher proficiency might therefore be required to optimally make use of the enhancement provided by visual speech information, and higher language

[3]Note that participants were recruited via Prolific. Given the restricted conditions due to the pandemic, we specifically made use of the pre-screening option in which only participants who reported being fluent in German were allowed to participate in this experiment. Although we ensured to make these restrictions, online testing still permits less control of the experimental situation, thus allowing for more variability in the data and unwanted and uncontrolled deviation in the experimental procedure, which may confound the results. This might also be one of the major downsides of online research.

proficiency leads to greater automatization of its processing, which leads to smaller processing costs in comprehension of the stimuli. This in turn makes a larger amount of resources available that can be employed in the retainment of the information in working memory. According to this, participants of the present study were unable to couple the phonological cues with the visual cues due to their low-language proficiency. This came at the expense of worse recall performance in both mask and no mask condition, suggesting a possible correlation between working memory and L2 language proficiency.

In fact, previous research studies have investigated the question of whether working memory performance varies as a function of L2 listeners' language competence. For example, Service, Simola, Metsänheimo, and Maury (2010) presented auditory sentences accompanied by pictures, and Finnish-English bilinguals were asked to memorize the last two words. Results showed significantly lower working memory spans for L2 listeners than for L1 listeners. Investigating the same issue, Van den Noort, Bosch, and Hugdahl (2006) examined working memory operations of multilingual talkers using the reading span task. Their participants were native (L1) Dutch talkers who spoke fluent German (L2) and less-fluent Norwegian language (L3). Participants performed best in their L1 as compared to their L2 and L3, and their performance in L2 was better than in their L3. Hence, the answer to the question of whether working memory performance is affected by language proficiency is clearly affirmative (Vejnovic, Milin, & Zdravković, 2010). Taken together, advanced linguistic expertise can enhance the retainment of information in working memory, explaining that L2 comprehension and working memory capacity are mediated by L2 language proficiency. Given that the working memory operates on limited cognitive resources (Just & Carpenter, 1992), we propose that the following happened in the present study: The low language

competence in the L2 led to higher processing cost of the stimuli which in turn required extra processing demand in comprehension, leaving no resources for retaining information in the working memory.

In summary, the present findings were to our knowledge the first to investigate L2 processing of masked speech. Although the results showed no impact of face mask on higher-level cognitive processing downstream, such as memory encoding, the results of the present study can be largely explained by the low language proficiency skills of our L2 listeners, thus reinforcing previous findings stating that L2 language proficiency plays a crucial role in extracting visual information from the lips of the talker. Future research could re-test this design with highly proficient L2 listeners, as this would create a complete picture of the impact of face mask going beyond the native context and illustrating the impact of face masks cross-culturally. Furthermore, it should be borne in mind that the usage of information from the lips is different cross-culturally. For example, while the McGurk effect was consistently found in Western countries, the magnitude of the McGurk effect was weak to non-existent in Asian countries like Japan (Sekiyama & Tohkura, 1993) and China (Sekiyama, 1997), suggesting that Japanese and Chinese listeners make much less use of visual information in speech comprehension. These research efforts, in turn, can further elucidate our present results and also inform debates on the impact of face mask and the role of visual information as well as linguistic and cultural experience in second-language listening comprehension.

CHAPTER 6

Trust issues: The talker age effect on credibility

Experiment 1 of this Chapter has been adapted from

Truong, T. L. and Weber, A. (2020). Trust issues: The effect of speaker age on credibility. In R. Hörnig, S. von Wietersheim, A. Konietzko, and S. Featherston (eds.) (2022), Proceedings of Linguistic Evidence 2020: Linguistic Theory Enriched by Experimental Data (pp. 351-361). Tübingen: University of Tübingen.

Abstract

Foreign-accented speech and child speech both deviate typically from the standard norms of adult native pronunciation. For foreign-accented speech, prior research has shown that English participants believed information less when it is produced in a foreign accent rather than a native accent, presumably because foreign-accented speech is harder to understand (i.e., processed less fluently).

In the present study, the effects of talker age, gender, and foreign accent were investigated in German. Native German participants were asked to judge the truthfulness of 48 German trivia statements spoken by one male adult talker and one child talker (Experiment 1), one female adult talker and one child talker of German (Experiment 2), four female adult talkers and four child talkers (Experiment 3), and four foreign-accented female adult talkers and four child talkers (Experiment 4). Generally, the results suggest that talker age does not influence credibility ratings in all cases, but listeners' voice preferences mediated the relationship between credibility judgments and talker age for female listeners (Experiment 1-3), such that female listeners rate statements from male adult talkers and female children as more trustworthy than statements from female talkers.

Although no direct evidence for an influence of *processing fluency* was found in the comparison of foreign-accented speech and child speech, foreign accents do not seem to have a detrimental effect on credibility judgments in the German context (Experiment 4).

Introduction

Throughout our lives, how we talk and sound affects how we are perceived and judged by others. That is, whenever we speak, we are being evaluated, and the credibility of what we say is being weighted (Ferguson & Zayas, 2009). Importantly, credibility not only depends on *what* we say but also on *how* we say it. The *how* includes, for example, the nativeness of our pronunciation, such that trivia statements made by foreign-accented talkers have been rated as less true than the same statements made by native talkers (Lev-Ari & Keysar, 2010).

The literature provides at least two explanations for the effects of foreign-accentedness on credibility. Negative attitudes toward foreign-accented talkers are possibly being promoted by in-group biases and not by the accent as such, which can serve as a marker for the biases (e.g., Dixon, Mahoney, & Cocks, 2002). Consequently, people with foreign-accented pronunciations often have to face stigmatization, social ostracism, or unfair jurisdiction (e.g., Dixon et al., 2002). It is generally acknowledged in sociolinguistics that people with a foreign accent are commonly judged as inferior (Edwards, 1999; Gluszek & Dovidio, 2010; Munro, Derwing, & Satō, 2006) in terms of intelligence, educational background, prestige, kindness, attractiveness, and trustworthiness (Anderson et al., 2007; J. N. Fuertes, Potere, & Ramirez, 2002; Lev-Ari & Keysar, 2010; Lindemann, 2003). As a consequence, eyewitnesses with foreign accents are, for example, perceived as less credible than those with native accents (Frumkin & Stone, 2020). Alternatively, credibility depends on how easily listeners can process the linguistic signal that deviates from the norms of the target language (i.e., foreign-accented speech; Lev-Ari and Keysar, 2010).

Given that foreign-accented speech typically deviates from the standard

norms, it can conceivably inhibit *processing fluency*, which in turn may have a potential impact on listeners' credibility judgments (Oppenheimer, 2008; Unkelbach, 2006). The term processing fluency can be broadly described as the ease of stimulus processing. For example, if speech is easy to understand, it is perceived as not only more pleasurable (Reber, Schwarz, & Winkielman, 2004), familiar (Whittlesea, Jacoby, & Girard, 1990), and less risky (Song & Schwarz, 2009) but also as more truthful (Reber & Schwarz, 1999). For example, rhyming language is known to be easier to process, and indeed it has been found that although the phrase "Woes unite enemies" has the same meaning as in "Woes unite foes", the latter is perceived as more accurate because of the rhyming of the words (McGlone & Tofighbakhsh, 2000).

One of the most compelling studies that found a relation between foreign-accentedness and credibility was conducted by Lev-Ari and Keysar (2010). Lev-Ari and Keysar (2010) argued that processing difficulties were the driving factor for more negative credibility ratings in foreign-accented speech. They tested three types of accents with different degrees of accent strength (native accents: English; mild foreign-accented accent: Polish, Turkish, and Austrian-German; heavy foreign-accented accents: Korean, Turkish, and Italian). Native English listeners were asked to judge the veracity of trivia statements like "Ants don't sleep" on a 14 cm long scale with the left pole marked with "definitely false" and the right pole marked with "definitely true". In an attempt to control negative stereotypical biases toward foreign-accented talkers, participants were told that the foreign-accented talkers solely acted as messengers of the statements, reciting statements which were provided by the experimenter. Thus, the statements would not reflect the talker's educational background, for example. The results showed that native English listeners judged trivia statements less often as true when the statements were spoken

by a foreign-accented talker than when the talker was native.

Interestingly, this effect was true for both mildly and heavily accented speech. When participants were made aware of the difficulty, mildly accented speech was rated as true as native accented speech. However, the negative effect remained for heavily accented speech. The authors concluded from this that not so much prejudice but rather segmental and prosodic deviations from the standard norms of the target language had a negative impact on processing fluency (Munro & Derwing, 1995), which in turn impacted listeners' credibility judgments. Their findings propelled further research investigating credibility judgments from different perspectives, in different language contexts, using different experimental methods.

A number of these studies had actually difficulties replicating Lev-Ari and Keysar's (2010) original finding. For example, Souza and Markman (2013) failed to replicate Lev-Ari and Keysar's (2010). Initially, Souza and Markman (2013) embedded the speech signal in noise at different signal-to-noise ratios and also added speech babble noise to the recordings of their native talkers. The hypothesis was that if indeed processing difficulties led to Lev-Ari and Keysar's (2010) findings, then noise should have a similar effect as the foreignness of the talker. Their findings, however, showed that noise variation did not negatively affect the credibility ratings of native talkers. Next, Souza and Markman (2013) used an experimental design very similar to Lev-Ari and Keysar (2010). Even though these two experiments overlapped now largely in terms of methodology, they failed to replicate Lev-Ari and Keysar (2010), and their participants did not rate statements produced by foreign-accented talker as less truthful than statements produced by native talkers. Stocker (2017) also attempted to reproduce Lev-Ari and Keysar's (2010) results - this time with a greater pool of participants and in two languages, French (n = 194) and German (n = 184). They too did not observe an effect of non-nativeness of the talker

on credibility. In addition, De Meo (2012) also failed to replicate the difference in credibility between native and foreign-accented accents, even though they increased the number of participants to 300. Podlipsky et al. (2016) tested three groups of participants: (1) native participants, (2) foreign-accented participants matching the L1 of the foreign-accented talkers, and (3) foreign-accented participants mismatching the L1 of the foreign-accented talkers. Findings again failed to support the processing fluency hypothesis. While foreign-accented participants indeed gave native statements higher credibility ratings, native participants showed no inclination to judge foreign-accented statements as less true than native statements. Podlipsky et al. (2016), however, did observe a "moderate correlation between comprehensibility and credibility of foreign-accented utterances" (p. 30). It must be noted, however, that the experimental design in Podlipsky et al. (2016) distinctively differed from the one Lev-Ari and Keysar (2010) used. For example, Lev-Ari and Keysar (2010) designed a within-subject and within-item design, meaning that both talkers and statements were fully crossed across conditions. This method ensured that the actual credibility of the statements themselves did not influence the judgments of the participants. Since Podlipsky et al. (2016) distributed statements and talkers across conditions, the actual truthfulness of statements and talker conditions may have been confounded, because a statement was always attributed to a specific talker condition. Frances et al. (2018) explored Lev-Ari and Keysar's (2010) research question with regional rather than foreign-accented accents. They compared credibility ratings for various regional accents (i.e., non-standard) and local accents (i.e., standard). Although this design tested a different type of accent, it is still relevant for Lev-Ari and Keysar's (2010), because the unfamiliar regional accents were not that different from the foreign-accented speech in Lev-Ari and Keysar (2010) as the unfamiliar regional accents and the foreign-accents are both non-standard conditions. In line with Lev-Ari and Keysar (2010), Frances et al.

(2018) predicted that regional accents will receive lower credibility ratings due to lower processing fluency (see also, Floccia et al., 2009). Their results, however, also showed no support for the processing fluency hypothesis. In fact, statements with a stronger accent were even considered to be more credible than others. Finally, M. Wetzel et al. (2021) tested whether or not familiarity can adjust the impact of foreign-accentedness on credibility. The researchers compared the credibility of French talkers, both with a familiar and unfamiliar native accent and with a familiar and unfamiliar foreign accent. Again, no corroborative findings were observed, that is, no effect for either foreignness or familiarity was found. However, it is difficult to draw conclusions from this since M. Wetzel et al. (2021) only employed one talker per accent condition.

The above mentioned studies make it clear that the impact of foreign-accented speech on credibility ratings has been examined in different language contexts, with little to partial methodological overlap and mixed results (Baus et al., 2019; Frances et al., 2018; Hanzlíková & Skarnitzl, 2017; Podlipsky et al., 2016; Souza & Markman, 2013; Stocker, 2017; M. Wetzel et al., 2021). Crucially, the mixed findings point out that the results presented in Lev-Ari and Keysar (2010) cannot be exclusively referred as processing fluency.

In light of the numerous conflicting findings, Boduch-Grabka and Lev-Ari's (2021) results recently strengthened Lev-Ari and Keysar (2010) such that processing fluency can make individuals trust information less when it is delivered in a foreign accent, at the same time their findings showed that discrimination against foreign-accented talkers can be reduced by means of exposure to foreign accents. Even though Boduch-Grabka and Lev-Ari (2021) accomplished a successful replication of Lev-Ari and Keysar's original finding, they concluded that processing difficulty might not be the major factor that can make individuals

distrust foreign-accented speech. Boduch-Grabka and Lev-Ari (2021) argued that their results might have been indeed generated by both prejudice and difficulty in processing fluency and that exposure merely facilitated spoken language comprehension, because even after short exposure to a talker with deviating pronunciation from the native standard norms, correct identification of words increases (Bradlow & Bent, 2008; C. M. Clarke, 2002; Maye et al., 2008) as well as comprehension ease (C. Clarke & Garrett, 2004). But what about other types of deviations from the standard norms of native adult speech, like for example children's speech?

Children's speech is interesting in this regard since attitudes toward children are very likely to be more positive than toward foreign-accented talkers, but both varieties of speech deviate from the native adult norms of the target language and hence could be more difficult to process. Compared to foreign-accented talkers, judges perceive children as more honest than adult witnesses, despite their limited memory capacities and verbal skills, which make them appear less reliable than adults (Bala, Ramakrishnan, Lindsay, & Lee, 2005). Unlike toward foreign-accented talkers, listeners typically have a positive attitude toward children, although the reliability of what they say might be seen to be lower than that of adult talkers. Similar to foreign-accented speech, acoustic and linguistic properties of children's speech are distinct from those of native adult speech (S. Lee et al., 1999). Generally, there can be differences between adult and child talkers based on deviations in pronunciation from the adult norm (e.g., pronouncing "cows" as "tows") as well as differences that are caused by distinct physical characteristics between the two groups of talkers. For example, while the average vocal folds of the adult are about 12-21 mm long, the vocal folds of an 8-year-old have grown to approximately 8 mm in length. The fundamental frequency (i.e., F0) of an infant's voice is at birth

around 500 Hz. As the larynx grows with age, F0 drops to about 275 Hz by the age of eight, with little difference between boys and girls. More specifically, boys have typically lower formant frequencies than girls (Vorperian & Kent, 2007). While F0 remains quite stable throughout childhood, at about 2.5 octaves, the variability decreases progressively until the age of 8-12 years (Kent, 1976). Not surprisingly, speech performance thus becomes more adult-like as children grow older.

The present study concentrates on children's speech at the age of seven. Although pronunciation norms typically start to approximate adult performance by seven years of age, it is likely that they do not fully align with that of adult talkers yet. Furthermore, children's speech is generally characterized by greater acoustic-phonetic variation than is native adult speech (B. Smith et al., 1983; Tingley & Allen, 1975) which might consequently impact processing fluency negatively, too. For example, children's speech displays "higher pitch and formant frequencies, longer segmental durations, and greater temporal and spectral variability" (S. Lee et al., 1999, p. 1455). Thus, the principle question at issue here concerns whether credibility judgment is being affected by talker age.

The present work aims at investigating the effects of talker age on credibility judgments. Given that child speech is generally characterized by greater-acoustic-phonetic variation than is adult speech (e.g., S. Lee et al., 1999), which can cause processing difficulties, we expected lower credibility ratings for statements spoken by a child talker than by an adult talker.

Experiment 1

Method

Speech material

Forty-five trivia statements were taken from Lev-Ari and Keysar (2010) and translated from English into German. The majority of the trivia statements were about the animal kingdom (e.g., *Ameisen schlafen nicht*, "ants don't sleep"). Nine of the statements were replaced by new statements because their German translation did not work well. For example, "The original name for butterfly was flutterby." was replaced since the German word for "butterfly" (i.e., *Schmetterling*, does not entail the embedded words "flutter" and "by"). The selected trivia statements were statements for which the correct answer is typically not known; thus judgments on a scale were not likely to be fixed on the endpoints of a credibility scale. This was necessary to allow differences in judgment to emerge when the same statements were produced by different talkers. The complete list of trivia statements can be found in Appendix C (see Chapter 7).

In order to have occasional trials in which participants could be sure of the truth value, 14 filler statements were added (e.g., *Brokkoli ist ungesund*, "broccoli is unhealthy"). Two experimental lists with the 45 statements and 15 filler statements were created. Each experimental sentence appeared once in each list, half of which were true and half of which were false, counterbalanced for the age of the talker. The order of sentence presentation and talker was pseudo-randomized. Each trial began with two practice sentences.

Talker selection and recording

The sixty trivia statements were recorded by one male adult talker of German (age 54) and one female child talker of German (age 7), both living in Tübingen, in the south of Germany. Neither talker had reported any speech impediment. Before the recording session, the talkers had time to get acquainted with the list of statements in order to prevent any disfluencies when reading.

The recording session took place in a sound-attenuated room with a high-quality microphone and a sampling rate of 44 kHz. Both talkers were recorded separately. The child talker, however, was recorded together with her mother. While the mother read from orthographic transcription, the child was prompted to repeat after her reading.[1] Special care was taken that all sentences were produced as intended and without disfluencies.

The adult talker had an average F0 of 224.14 Hz and the child talker had an average F0 of 287.98 Hz. The difference in F0 between these two talkers was significant ($t(118)$ = -2.51, $p < .02$). The correctness of the statements was not shared with the talkers, so that their speech was not affected. The main purpose of the study was revealed to the talkers only after the recording session.

Participants

Forty-two native listeners of German (23 female), between 19 and 34 years old (mean age = 23.5, SD = 3.8), participated in the experiment for monetary compensation. All participants were students at the University of Tübingen. None of them suffered from any hearing disorders, and they all had intact or corrected-to-normal vision. The procedures for the present experiment were approved by the DGFS

[1]Elementary school starts at the age of six in Germany, and by the age of seven reading aloud unprompted is typically still less fluent than in adults. Repeating after an auditory prompt ensured that the child talker produced the sentences naturally and fluently.

(Deutsche Gesellschaft für Sprachwissenschaft) ethics committee for the Chair of Psycholinguistics and Applied Linguistics at the University of Tübingen.

Procedure

The experiment was carried out with Presentation (version 20.1, www.neurobs.com). Before the experiment started, participants signed written informed consent. Participants were seated comfortably in front of a computer screen and wore over-ear headphones (Sennheiser HD 215 II) and were tested individually.

The experiment was controlled with Excel Visual Basic (version 16.0.11328.20362). Participants wore over-ear headphones and were tested individually. Instructions were presented on the computer screen. The experiment started with two practice sentences, followed by the 60 trivia statements. Each statement was presented once. Participants entered their truth judgments with the use of a sliding scale. The sliding scale was similar to the one used by Lev-Ari and Keysar (2010); the left end of the scale was labeled with "definitely false", and the right end was labeled with "definitely true".

Participants evaluated the level of credibility for each sentence by dragging the slider bar, starting from its default position at the scale's center, until it reached the desired answer position. Although not visible to participants, the positions on the scale ranged underlyingly from 0 to 140, with higher numbers indicating higher perceived credibility. Furthermore, participants were asked to genuinely try to assess the veracity of each statement and to use the full scale.

In addition, participants were asked to indicate after each truth judgment, whether they had heard the statement before and already knew the correct answer. Three possible answer options were given: "yes", "no,", and "unsure". After the experiment was completed, participants filled in a short language background

questionnaire.

Results

R (R Core Team 2018, version 3.5.0) and lme4 (Bates, Maechler, et al., 2015) were used to perform linear mixed effects analyses on listeners' perceived truth judgments (see Figure 6.1). Only statements that were unknown to the participant were analyzed (73.3 % of the items), since we expected that known statements were less likely to be affected by talker attributes,[2] and the pattern of results did not significantly change. Two participants had to be excluded since they did not follow the instructions. The initial model included *talker age* (adult, child) and *participants* and *items* as random variables with random slopes (see Table 6.1).

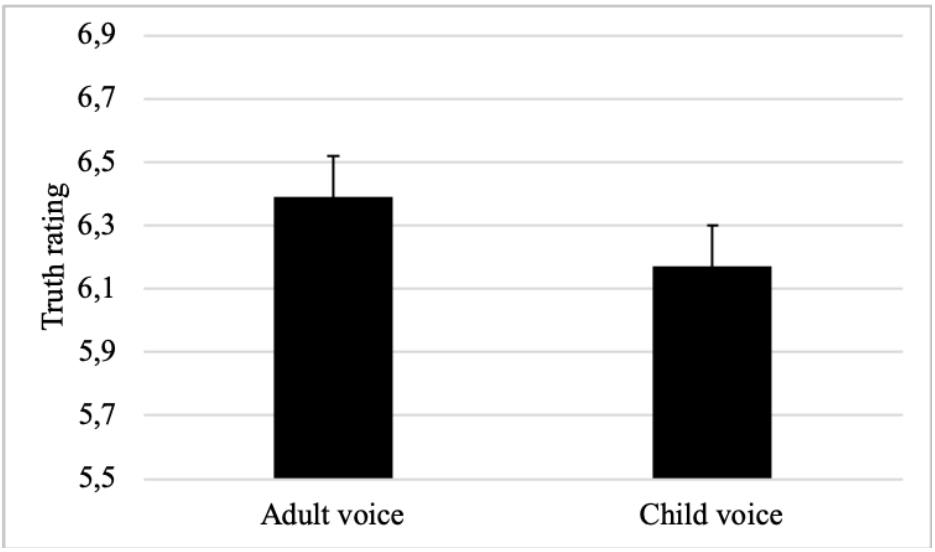

Figure 6.1: Truth ratings as a function of age of talker voice (male adult, female child). The y-axis indicates the credibility ratings from *definitely false* to *definitely true.* Higher numbers indicate higher perceived credibility.

[2]Note that Lev-Ari and Keysar (2010) had not found any evidence for an effect of *knowledge*. Identical to their statistical analysis, we additionally tested all statements and included the interaction of *age of talker* (i.e., adult, child) and *knowledge* (i.e., yes, no, unsure) to the model. Similarly to Lev-Ari and Keysar (2010), *knowledge* did not improve the model.

Table 6.1: Experiment 1 initial LMER model

Fixed effects	*Estimate*	*Std. Errors*	*t*	*p*
Talker Age	-0.2293	0.1723	-1.331	0.184

*Note *p<.05 **<.01 ***<.001*

Although descriptively, ratings for the child talker were somewhat lower than for the adult talker, the analysis showed only a non-significant trend of *talker age* (b = -0.23, SE = 0.17, t = -1.3, $p > .1$), suggesting that overall credibility ratings were not significantly affected by *talker age*. When we further looked at the data descriptively, we noted, however, that participants varied in their response patterns. Specifically, female participants displayed a different pattern from male participants. Based on this observation, we decided to conduct an exploratory analysis (see Figure 6.2). We now grouped the data based on the gender of the participants (i.e., male and female listeners).

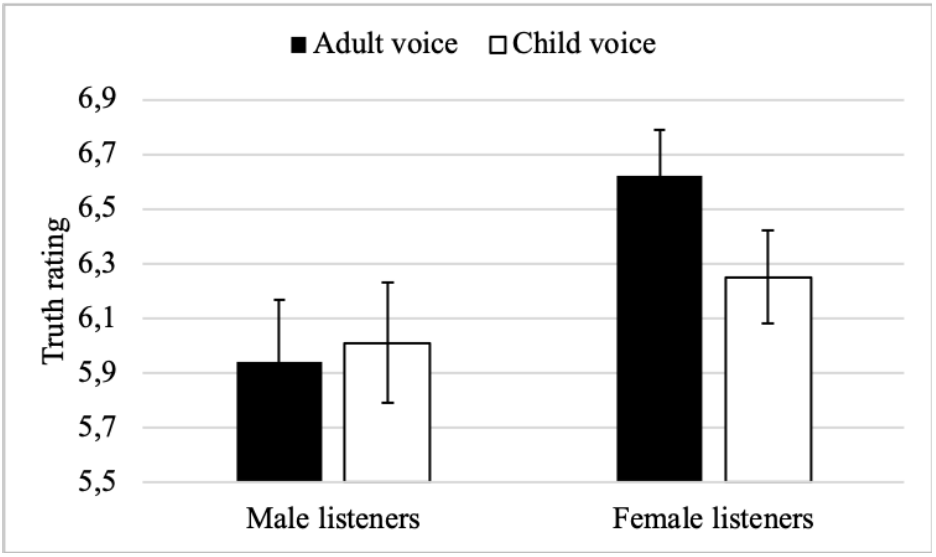

Figure 6.2: Truth ratings of male and female listeners for adult and child voices. The y-axis indicates the truth ratings from *definitely false* to *definitely true*. Higher numbers indicate higher perceived credibility.

Three participants had to be excluded from the analysis since they did not

indicate their gender in the questionnaire. When gender of the participants was considered in the analysis, there was a marginally significant effect of *age of talker* (b = -0.11, SE = 0.06, t = - 1.82, $p < .07$) for female participants.[3] For male listeners, no indication of a voice effect was found (b = -0.06, SE = 0.05, t = -1.20, $p > .2$). The direction of the effect for female listeners was such that statements made by the child talker were judged to be less trustworthy.

Interim discussion

While our initial hypothesis that credibility ratings are affected by talker age, was not confirmed, subsequent exploratory analyses indicated a different pattern of ratings for female and male listeners. Specifically, when taking listeners' gender into account, female listeners judged sentences spoken by the male adult talker as more credible than sentences spoken by the female child talker. No such effect for the male listeners was found. These exploratory findings suggest that factors other than processing difficulty in child speech might have an impact on truth judgments. Since the exploratory findings were only marginally significant, interpretation has to be approached with caution, and at this point it is rather speculative. But it leaves open the possibility that listeners' voice preferences modified the results and that processing fluency played only a rather minor role in credibility judgments in Experiment 1.

A growing literature has demonstrated that the talker's first impression is inferred not only from visual cues but also from auditory cues (Rezlescu et al., 2015; Zuckerman & Miyake, 1993). Listeners particularly rely on auditory cues when it comes to identifying dominance and trustworthiness impressions (Rezlescu et al.,

[3]We refrained from a Bonferroni correction, as this might lead to committing a Type II error (i.e., false negative) (Winter, 2019). However, if we were to compute a pairwise comparison, using the *emmeans* (Lenth, 2018), the p-value would change from <.067 to < .068.

2015). Several empirical studies have demonstrated that F0, a key acoustical feature, (i.e. "highness" or "lowness" of the voice) and its corresponding harmonics (Fitch, 2000) influence judgments of people's personality traits (Belin et al., 2011; McAleer et al., 2014; Tsantani et al., 2016). While women typically have smaller vocal folds that vibrate at a higher rate, thus having a higher-pitched voice, men by contrast typically have larger vocal folds which vibrate at a lower rate, thus providing a lower-pitched voice. Generally, individuals with lower voices, are perceived as taller (Xu et al., 2013), physically stronger (Sell et al., 2010), socially more dominant (Tigue et al., 2012), and more attractive (Feinberg et al., 2005).

In simple terms, voices which are preferred are perceived as more attractive and could thus be more trustworthy. This particularly motivated us to conduct Experiment 2, as it allows us to widen the scope of our research question and thus lets us go beyond the notion of processing fluency.

Experiment 2

Experiment 2 used the same trivia sentences as in Experiment 1, but this time we substituted the male adult talker with a female adult talker and tested only female participants. Recall that in Experiment 1 the talkers not only differed in age (adult vs. child) but also in gender (male vs. female). Our prediction for Experiment 2 was that if the credibility effect is caused by the gender of the talker, we should find no difference in credibility ratings between the female adult talker and the female child talker. However, if the effect is caused by talker age, then we should find a difference.

Method

Participants

Forty female native listeners of German, between 18 and 27 years old (mean age = 21.43; SD = 2.1), participated in the experiment for a small monetary compensation. All participants were students at the University of Tübingen and had no reported visual or hearing impairments.

Material

The trivia sentences and the recordings of the child talker were the same as in Experiment 1. However, the male adult talker was replaced by a native German female adult talker (age 38). Similar to the recording session of Experiment 1, the female adult talker was recorded separately (and not together with the child talker). The female adult talker had an average F0 of 191.16 Hz. The F0 difference between the female adult talker and the child talker was again significant (t = -5.59, p = 1.473e-07).

Procedure

The procedure was identical to the procedure in Experiment 1.

Results

Similar to Experiment 1, R and lme4 were used to perform linear mixed effects analyses on listeners' perceived truthfulness (see Figure 6.3). Only statements that were unknown to the participant were included in the analysis (68.38 % of the items). The result showed a marginally significant effect for talker voice (b = 0.28, SE = 0.16, t = 1.7, p = 0.08). The values of the initial model can be found in Table 6.2.

Table 6.2: Experiment 2 initial LMER model

Fixed effects	*Estimate*	*Std. Errors*	*t*	*p*	
Talker Age	0.2884	0.1692	1.705	0.0885	.

*Note *p.05 **.01 ***.001*

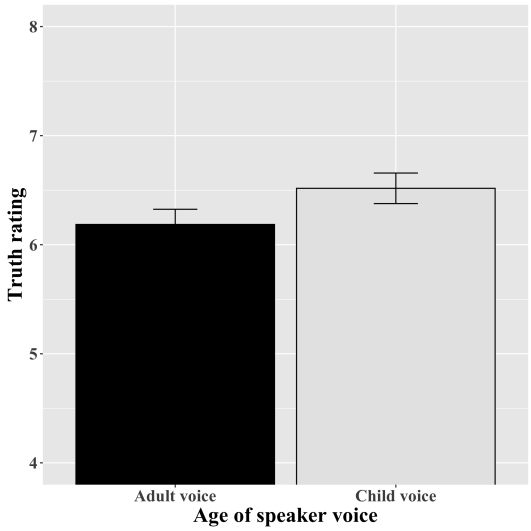

Figure 6.3: Truth ratings as a function of age of talker voice (female adult, female child). The y-axis indicates the truth ratings from "definitely false" to "definitely true". Higher numbers indicate higher perceived truthfulness.

Surprisingly, and contrary to Experiment 1, the results showed that female listeners judged trivia statements now as less true when the sentences were produced by the female adult talker than when the talker was a female child. We then collapsed the data from the female participants in Experiment 1 with that of Experiment 2 and assessed whether listeners' ratings from female participants differed across experiments. A linear mixed effects analysis was conducted, with *ratings* as the dependent variable, and *experiment version* (Experiment 1 vs. Experiment 2) *talker age* as fixed effects. *Participant* and *items* were included as random factors with random slopes. We found a significant interaction between *experiment version* and *talker age* ($b = 0.62$, $SE = 0.28$, $t = 2.26$, $p = 0.02$) (see Table 6.3), indicating the different patterns between Experiments 1 and 2. That is, while female listeners

rated the male talker as more credible than the child talker, female listeners also rated the female talker as less credible compared to the child talker. The pitch range between the male and female adult talkers was not significantly different (t = -1.34, p = 0.18).

Table 6.3: Experiment 2 LMER model of combined data from Experiments 1 and 2

Fixed effects	*Estimate*	*Std. Errors*	*t*	*p*	
Talker Age	-0.3151	0.2154	-1.463	0.144	
Experiment Version	-0.3211	0.2785	-1.153	0.252	
Talker Age x Experiment Version	0.6223	0.2754	2.260	0.024	*

*Note *p.05 **.01 ***.001*

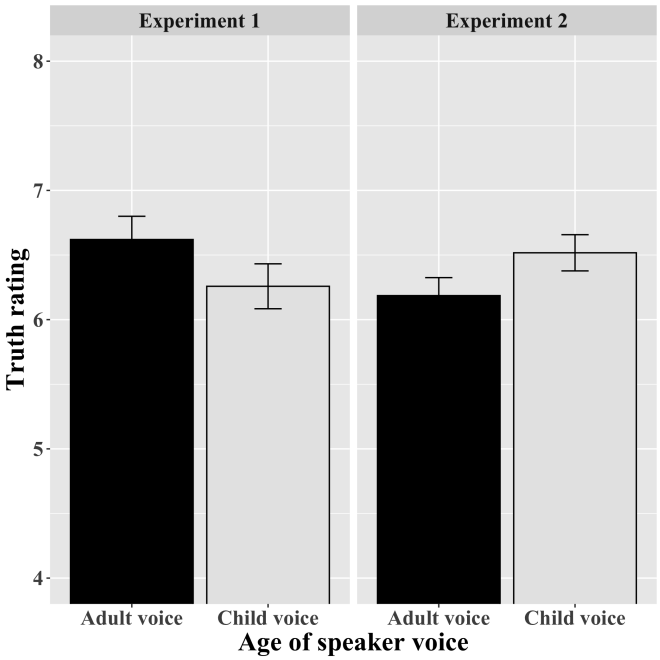

Figure 6.4: Truth ratings from Experiments 1 and 2 as a function of age of talker voice (adult, child). The y axis indicates the truth ratings from *definitely false* to *definitely true.* Higher numbers indicate higher perceived truthfulness.

Interim discussion

Recall that the exploratory findings of Experiment 1 showed higher credibility ratings for the male talker than for the child talker by female listeners. When combining the data from Experiments 1 and 2, results demonstrated that female listeners judged the female talker as less credible than the child talker in Experiment 2, strengthening the notion that the effect found in Experiment 1 was caused by factors (e.g., attitudes) other than inhibition of processing fluency. Since Experiment 2 used the same recordings from the child talker that were used in Experiment 1, it is rather unlikely that the results were due to the pronunciation of the child talker. Participants also reported not having any issues understanding the adult and child talkers. So, this finding rather indicates that the results of Experiment 1 were not primarily driven by processing fluency but that attitudes may play a more significant role in credibility judgments. Yet, we cannot draw strong conclusions at this point since we only used a single talker for each condition (adult vs. child). It is still possible, that some uncontrolled voice characteristics of the talkers, for example, rather than their age, influenced the results in Experiment 2. To ensure that age of the talker rather than individual voice characteristics influence truth judgments, Experiment 3 investigated if the findings of Experiment 2 remain robust with multiple talkers.

Experiment 3

A new group of participants listened to the same trivia sentences as in Experiments 1 and 2. As in Experiment 2, talker age was compared, while gender was kept constant. In contrast to Experiment 2, however, we increased the number of talkers in Experiment 3.

Method

Participants

Twenty-seven female listeners of German, between 20 and 33 years old (mean age = 25.22; SD = 3.4), participated in the experiment for a small monetary compensation. All participants were students at the University of Tübingen and had no reported visual or hearing impairments.

Materials

All trivia statements were the same as the trivia statements used in Experiment 1. Speech materials of the female adult talker and the female child talker of Experiment 2 were included in Experiment 3. In addition, three more female adult talkers and three more children were recorded for the present experiment. Overall, four female adult talkers (in their mid-40s) and four female child talkers (two 11-year-olds and two 7-year-olds, mean age = 14.5) were used as talkers. All talkers were native Germans who were living in the Tübingen area at the time of the recordings. The recording procedures were the same as in Experiments 1 and 2. The children were prompted to repeat the sentences after their mother had read from orthographic transcriptions. The voice pitch of all talkers was measured. Adult talkers had an average F0 of 227.31 Hz; child talkers had an average F0 of 244.01 Hz. The average F0 difference between these two groups of talkers (i.e., adults vs. children) was significant (t = -1.83, p = 0.06).

Procedure

The procedure was identical to the procedure in Experiments 1 and 2.

Results

As in Experiments 1 and 2, R and lme4 were used to perform linear mixed effects analyses on listeners' perceived truth (see Figure 6.5). Only statements that were unknown to the participant were analyzed (70.6% of the items). The results showed a significant effect for talker age ($b = 0.37$, $SE = 0.18$, $t = 1.98$, $p = .04$). The values of the initial model are displayed in Table 6.4.

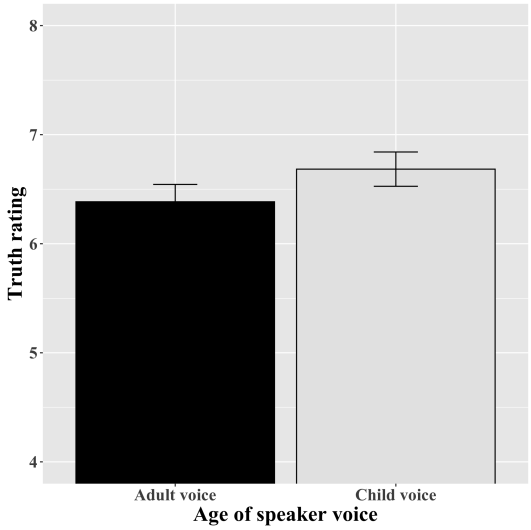

Figure 6.5: Truth ratings as a function of age of talker voice (female adults, female children). The y axis indicates the truth ratings from *definitely false* to *definitely true*. Higher numbers indicate higher perceived truthfulness.

Table 6.4: Experiment 3 initial LMER model

Fixed effects	*Estimate*	*Std. Errors*	*t*	*p*	
Talker Age	0.376	0.1893	1.986	0.0473	*

*Note *p.05 **.01 ***.001*

Interim discussion

In line with Experiment 2, findings of Experiment 3 showed that female listeners indeed found trivia statements spoken by female adult talkers as less true compared to female child talkers. This replication of the pattern of Experiment 2 suggests that the findings are not caused by a specific individual talker, but are indeed talker independent and show an effect of talker age.

Overall, the results of Experiments 1 to 3 strengthened the claim that credibility is most likely not mediated by processing fluency since it can be assumed that processing fluency was not reliably different across the different tested talkers. However, to fully understand the roles of processing fluency and talker age, a comparison with foreign-accented talkers is needed. The direct comparison between child speech and foreign-accented speech was made in Experiment 4.

Experiment 4

Experiment 4 asks once again the question if credibility judgments are influenced by processing fluency. For this matter, we used the same child talkers as in Experiment 3, but substituted the four female adult talkers of Experiment 3 with four non-native German female talkers. We predicted the following outcome: If processing fluency affects credibility ratings, we should find a difference in credibility ratings such that statements made by the foreign-accented talkers are rated less credible than those of the child talkers. This prediction can be attributed to our earlier findings. We started out with the hypothesis that child speech is similar in many aspects to foreign-accented speech and that, in line with previous studies, statements by child talkers should therefore be rated as less credible than statements by native adult talkers.

Findings of Experiment 1 through 3, however, made it clear that the pronunciation of our child talkers was in fact highly intelligible and did not deviate too much from adults' norms, since none of the participants reported any comprehension difficulties. Hence, processing fluency is less likely to be the primary reason for the results. Foreign-accented speech, however, deviates more noticeably from the standard norms of native pronunciation. We, therefore, expected that listeners will find it harder to understand foreign-accented speech than child speech, which in turn should lower the credibility ratings for foreign-accented talkers, as suggested by Boduch-Grabka and Lev-Ari (2021) and Lev-Ari and Keysar (2010).

Method

Participants

Sixty-two native listeners of German, between 18 and 49 years old (mean age = 24.1, SD = 6.0) participated in Experiment 4. All participants were female and were recruited via Prolific. They reported no visual or hearing impairments. Fifteen participants in our participant pool turned out to be bilingual, meaning that they grew up with more languages than just German, and therefore did not meet the participant requirements. As a result, they were excluded from further analyses.

Materials

The trivia statements were the same ones used in the previous experiments. Also, the same recordings of the four child talkers in Experiment 3 were used again. For Experiment 4, however, four female non-native adult talkers of German were recorded. The total number of talkers in Experiment 4 were: four non-native adult talkers (mean age = 41) and four child talkers. All talkers were female.

The language background of the non-native adult talkers comprised of

Arabic, Chinese, Spanish, and Russian. The recording procedures of the adult talkers were the same as in the previous experiments. Special care was taken to ensure that the non-native talkers produced the materials without any hesitations, pauses, or mispronunciations.

The adult talkers had an average F0 of 367.34 Hz; child talkers had an average F0 of 244.01 Hz. F0 difference between these groups of talkers (i.e., adults vs. children) was significantly different ($t = 12.54$, $p < .001$).

Procedure

Due to the Covid-19 pandemic and the restrictions we had on using our in-person lab, we switched to online testing for Experiment 4. Experiment 4, was implemented and run on Gorilla, which is an experiment builder software and host of online research studies (www.gorilla.sc) (Anwyl-Irvine et al., 2019).

The online experiment comprised of four phases: (1) ethics and consent, (2) headphone screening test, (3) the experiment, and (4) language background questionnaire. Before the experiment started, participants first provided informed consent. After that, they were provided with detailed instructions asking them to sit in a quiet room, to eliminate distractions, and to wear a pair of headphones (e.g., over-the-ear or in-ear headphones) during the entire experiment. All experimental materials were loaded prior to the start of the experimental session to guarantee no loading issues during the experiment. To further ensure that participants completed the experiment with headphones, a headphone-screening task was incorporated prior to the actual experiment Woods et al. (2017).

For this task, we used a 3AFC paradigm, meaning that participants heard three intervals of randomly ordered white noise with equal frequency and duration, but one interval contained a Huggin's Pitch tone that was played in a randomized

order. The task of the listeners was to detect which of the three tones contained the Hugging' Pitch. This test comprised of six trials and listeners needed at least five correct responses out of six to pass the headphone test in order to proceed with the experiment. Listeners who passed, were immediately directed to the experiment but those who failed were automatically rejected from the experimental trials. The goal of the headphone screening test was to establish ideal listening conditions when using headphones compared to in-built computer loudspeakers.

In addition, participation was only permitted using a laptop or computer with the browsers Firefox, Google Chrome, and Safari, since the correct functioning of our experiment with these browsers and hard devices was pre-tested by our research assistants. Participants were automatically rejected if the technical requirements were not met. Despite the difference in layout and setup, the experimental procedure was identical to the in-lab versions (Experiments 1-3). Participants clicked on the play button to listen to the stimuli. Then they were asked to rate the credibility on a sliding scale and indicate whether or not they knew the answer to this phrase before. In addition, participants could also indicate whether or not they understood the talkers. After the rating task was completed, participants filled in a short language background questionnaire, including questions about their prior experience with Arabic, Chinese, Spanish, and Russian.

Results

As in the previous experiments, R Core Team (2021) (version 4.0.5) and lme4 (Bates, Maechler, et al., 2015) were used to perform linear mixed effects analyses on listeners' perceived truthfulness. Only statements that were unknown to the participants were

included in the analysis (82.1% of the items).

Table 6.5: Experiment 4 LMER model

Fixed effects	*Estimate*	*Std. Errors*	*t*	*p*	
Talker Age	-0.38625	0.24349	-1.586	0.1132	
Language Familiarity	0.30603	0.13206	2.317	0.0329	*
Target Language Contact	-0.09124	0.08435	-1.082	0.2945	
Contact Frequency	-0.10854	0.10170	-1.067	0.3012	

Note *p.05 **.01 ***.001

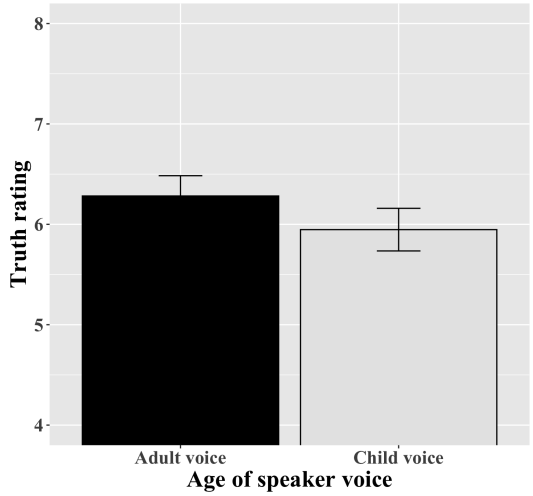

Figure 6.6: Truth ratings as a function of age of talker voice (non-native female adults, native female children). The y axis indicates the truth ratings from *definitely false* to *definitely true.* Higher numbers indicate higher perceived truthfulness.

The model included *talker age* as an independent variable. In addition, to account for prior experience with the talkers' first language (i.e., being familiar with the talkers' L1 Arabic, Chinese, Spanish, and Russian), we included *language familiarity*, *target language contact* through friends, and *contact frequency* as a fixed factor. Prior experience consisted of the sum of self-reported ratings (i.e., from 1 the lowest to 7 the highest) in each category. As before, *participants* and *items* were included as random variables with random slopes. The analysis showed no significant effect for *talker age* (b = -0.39, SE = 0.24, t = -1.59, p = 0.11), meaning accentedness of the non-native German talkers did not influence credibility ratings. We found no difference for *target language contact* (b = -0.09, SE = 0.08, t = -1.1, p = 0.29) nor for *contact frequency* (b = -0.1, SE = 0.1, t = -1.1, p = 0.3).

Interestingly, there was a main effect for *language familiarity*, meaning that the participants who were familiar with the talkers' L1, credibility ratings were positively affected (b = 0.31, SE = 0.13, t = 2.32, p = 0.03). This indicates that being familiar with the language of the foreign-accented talker shows a positive effect, which can result in higher credibility ratings.[4] The values of the initial model are displayed in Table 6.5.

[4]We conducted an analysis that included the full participant pool, including the bilinguals to see whether null effect for talker age was due to lack of power. The null effect of *talker age* remained the same (b = -0.39, SE = 0.24, t = -1.6, p = 0.11). Additionally, no main effect of *language familiarity* was found (b = 0.18, SE = 0.1, t = 1.8, p = 0.08), which could be due to the variation that bilinguals introduce to the data

General discussion

We will first provide a summary of all experiments which is then followed by a discussion of the findings. Lastly, we will then dive into the key message that goes beyond the interpretation of the impact of processing fluency on credibility ratings: That is, credibility judgments can also be formed by individuals' voice preference.

Summary of the experiments

The present study further explored the idea that the difficulty associated with child speech potentially affects truth judgments as was previously shown for foreign-accented speech in Lev-Ari and Keysar (2010). We present findings from four experiments, which adopted the methodology from Lev-Ari and Keysar (2010).

In the experiments, we compared credibility judgments of trivia statements spoken by one male adult talker of German and one female child talker of German (Experiment 1), one female adult talker of German and one female child talker of German (Experiment 2), four female adult talkers of German and four child talkers of German (Experiment 3), and finally four non-native female adult talkers of German and four female child talkers (Experiment 4). Analogous to foreign-accented speech, children's speech typically deviates to some extent from native adult speech in terms of acoustic and linguistic properties (S. Lee et al., 1999). This variation could make speech processing conceivably more difficult, which in turn could influence credibility ratings, if processing fluency has a direct influence on trustworthiness. On the other hand, as adults, we have accrued extensive knowledge of the world through experience. Possibly this experience would make it more likely for everyone to trust the credibility of statements more when spoken by an adult talker than by a child talker. However, the observed pattern of results was more complex than that,

and did not offer much support for the initial hypothesis of lower credibility ratings for child talkers.

Findings from Experiments 1 to 4 showed overall a varied pattern of differences in credibility judgments for statements by adult and child talkers. What we did not find was a confirmation of our initial hypothesis that credibility ratings for the adult talkers will be higher than for the child talkers. In Experiment 1, a comparison of statements made by a male adult talker and a female child talker, the first analysis did not reveal the pattern we initially hypothesized.

That is, overall, participants did not judge trivia statements significantly different when the statements were spoken by a child talker or an adult. Only when taking participants' gender into account, different patterns seemed to emerge in an exploratory post hoc analysis. This analysis indicated a different credibility judgment pattern for male and female participants. While there was no effect for male participants, female participants gave lower credibility judgments for the female child talker than for the male adult talker. In Experiment 2, we used the same trivia sentences but substituted the male adult talker with a female adult talker and tested only female participants, who had shown an effect in Experiment 1. Female participants did show again a significant difference in Experiment 2, but contrary to the exploratory results of Experiment 1 with the male talker, female listeners judged sentences produced by the female adult talker as less, and not more, credible compared to sentences spoken by the female child talker. This pattern was then successfully replicated in Experiment 3, in which we used multiple female adult and child talkers. The finding that adult talkers can be rated as less credible than child talkers, challenges the hypothesis made by Lev-Ari and Keysar in 2010 for foreign-accented talkers, who argued that ratings are lower when processing fluency is reduced. It is unlikely that the female adult talkers in Experiments 2 and 3 were

harder to process than the child talkers. If anything, it should have been the other way around. We tested the processing fluency hypothesis once more in Experiment 4 with a direct comparison of child and foreign-accented talkers. In Experiment 4, we did not obtain evidence for negative credibility ratings for foreign-accented talkers. We used four different foreign accents (i.e., Arabic, Chinese, Spanish, and Russian), but only one talker per accent. No significant difference in credibility ratings was found.

Discussion of the findings

Although the experimental design of Experiments 1 to 4, was very similar to the method used in Lev-Ari and Keysar (2010), it should be borne in mind that the context of the present study did differ in some important aspects from that of Lev-Ari and Keysar's (2010) study. The present results suggest that their findings do not generalize to an effect of talker age. Additionally, one major factor that makes the present work distinct from Lev-Ari and Keysar (2010) is that the study tested the hypothesis in a German context. Previous studies, which examined that same hypothesis in different language contexts, also found no evidence supporting (Lev-Ari & Keysar, 2010)'s findings (e.g., Baus et al., 2019; Frances et al., 2018; Hanzlíková & Skarnitzl, 2017; Podlipsky et al., 2016; Souza & Markman, 2013; Stocker, 2017). It is worth pointing out that the aforementioned studies were no full replications of Lev-Ari and Keysar (2010) since different foreign accents in different language contexts were tested, possibly explaining the differences in results.

Recently, Boduch-Grabka and Lev-Ari (2021) strengthened the results of Lev-Ari and Keysar (2010). Even though Boduch-Grabka and Lev-Ari (2021) found that a foreign accent can lead individuals to distrust information more when it is delivered by a foreign-accented talker, they also found that this bias against

foreign accent can be minimized by exposing participants to the respective foreign accent. Boduch-Grabka and Lev-Ari (2021) explained that increased familiarity with a foreign accent can facilitate language comprehension and in turn can influence attitudes such that participants believe information more when it is produced in a foreign accent. However, it is important to note that attitudes toward foreign accents may vary across countries, meaning that they are not always negative, because foreign accents can also evoke positive reactions in listeners (Dewaele & McCloskey, 2015). For instance, while the Spanish accent is considered a nonstandard accent in the United States and is connected with negative traits, it has been shown to positively affect listeners' perception of talkers' educational background, social status, and personal traits like attractiveness in the United Kingdom (J. Fuertes et al., 2012). Moreover, Giles (1970) found that French-accented English received more positive evaluations than Italian or German accents, even superior to English regional accents such as the Birmingham accent. Based on different stereotypes in different countries, the effect shown by Lev-Ari and Keysar (2010) appears not to be easily generalized to different language contexts.

Although findings of Experiment 4 do not corroborate the claim that processing difficulties associated with foreign-accented speech downgrade credibility judgments, it is still intriguing in this regard since it seems that participants experienced no processing difficulties when sentences were produced by foreign-accented talkers, suggesting that the intelligibility of our foreign-accented talkers was in fact relatively high. In other words, comprehensibility was not compromised to trigger lower credibility ratings. In fact, none of the participants reported having difficulties understanding the talkers. This pattern is indeed in line with De Meo (2012), saying that credibility is more directly linked to comprehensibility rather than the level of accentedness.

In addition, Lev-Ari and Keysar (2010) also found no difference in ratings between mildly and heavily accented speech. That is, stronger accents did not evoke stronger effects in credibility ratings. When looking at our participants' profiles, we saw that the majority have markedly multilingual profiles, that is they reported having competencies in more than two languages which they obtained in German high schools. Also, many of the participants claimed knowledge of the accent languages investigated, particularly Spanish. This suggests that familiarity made participants immune to the influence of foreign accent on credibility, possibly explaining the null effect. This could also explain the main effect of *language familiarity*, because participants who reported being familiar with the talkers' accents gave higher credibility ratings compared to those who were not familiar with the accents - in line with Boduch-Grabka and Lev-Ari (2021).

Similarly, speech intelligibility of the child talker was not further manipulated in this study. The size of the larynx of children as well as the ongoing process of language acquisition make child speech different from adult speech. While the former has an influence on the general height of the voice (F0) and also results in adjusted vowel formants (e.g., Hillenbrand et al., 1995; Peterson & Barney, 1952; Vorperian & Kent, 2007), the latter typically results in segmental speech production that deviates (e.g., sound substitutions, omissions, and additions) from native adult talkers. Especially this latter aspect bears similarities in type of deviations to foreign-accented speech. The child talkers in Experiments 1 to 4 did differ in F0 from the adult talkers, but had produced fluent sentences without obvious mispronunciations.

In fact, none of the participants reported any comprehension difficulties when listening to both adult and child talkers. Our child talkers were relatively advanced in their speech production skills, as they were between the ages of seven

and eleven. Hence, the pronunciation of the child talkers did not deviate noticeably from the native norms and apparently not enough to cause disruptions in processing fluency. Thus, it remains a possibility that in future studies a negative effect on credibility emerges for younger children with lower accuracy in their speech production skills. Interestingly, De Meo (2012) pointed out that in their study with foreign-accent speech, the strength of the foreign accent in terms of segmental deviations did not influence the perceived credibility of statements but that prosodic characteristics of an utterance had a greater influence on credibility judgments.

Based on this premise, it is conceivable that we would have found the same results had we used younger talkers in our experiments, at least as long as their speech was still intelligible. Although speech from much younger children unquestionably would vary more distinctly in terms of acoustical features and pronunciations, it would bear the risk of being incomprehensible. For the present study, it was essential that listeners understood what was being said. Otherwise low credibility ratings could have been simply due to statements being not intelligible. What remains as a source of influence on credibility ratings in our experiments were the voices of the talkers and the attitudes toward the talkers, implying that factors other than processing fluency could have induced credibility ratings.

The influence of voice on credibility

Just by listening to the voice of a talker, listeners can determine a range of information about the talker quite quickly and effortlessly (McAleer et al., 2014). The voice contains a vast spectrum of information about the talker such as gender, age, height and weight, ethnic background, and emotional states (Latinus & Belin, 2011). This clearly illustrates that the voice is not only a medium for oral communication, but that it conveys important information about the talker.

Listeners automatically perceive and decode this information and it guides them in their evaluation of the talker.

It has long been known, for example, that attractiveness of a person is typically associated with honesty, commonly related to the "'what is beautiful is good"' stereotype (Dion, Berscheid, & Walster, 1972). Dion et al. (1972) was the first study that documented the so-called attractiveness "halo" effect (Nisbett & Wilson, 1977; C. G. Wetzel, Wilson, & Kort, 1981). The "halo" effect entails the concept that listeners who assess a talker positively, might implicitly attribute this judgment to further favorable traits like trustworthiness. Whereas, talkers showing negative attractiveness are attributed to more negative traits. Dion et al. (1972), for example, asked participants to rate the physical attractiveness of faces. Prior to the experiment, the target pictures had been selected and rated by a different group of judges, categorizing them into low, medium, and high physical attractiveness. Results indicated that participants attributed positive personality characteristics to physically attractive faces.

This "halo" effect, however, is not solely limited to physical attractiveness, but it can be attributed to vocal attractiveness, too. For example, attractive voices are characterized with positive attributes such as intelligence, kindness, and trustworthiness (Hughes & Miller, 2016; Zuckerman & Miyake, 1993). Therefore, voices which are preferred are perceived as more attractive and thus more trustworthy. Several studies have found that attractive voices have been characterized with positive attributes such as dominance, accomplishment, and likeability (Zuckerman & Driver, 1989), better performances at work (DeGroot & Kluemper, 2007; DeGroot & Motowidlo, 1999), higher persuasion skills as political leaders (e.g., Chaiken, 1979; DeGroot, Aime, Johnson, & Kluemper, 2011; Surawski & Ossoff, 2006; Tigue et al., 2012), socially more competent (Burgoon, T., & Pfau,

1990), as well as more intelligent, kind, and trustworthy (Hughes & Miller, 2016; Surawski & Ossoff, 2006; Zuckerman & Miyake, 1993). Taken together, voices which are preferred are perceived as more attractive; this positive attitude toward them can result in higher trustworthiness.

If attitudes also play a role in credibility ratings, then our initial finding in Experiment 1 in which we compared a male adult talker with a female child talker could be explained with children's speech lowering processing fluency and therefore lowering credibility ratings while at the same time positive attitudes toward children having a positive effect on the ratings, despite the fact that children can be considered as less knowledgeable than adults. These two factors might have canceled each other out, possibly explaining the null effect in the initial analysis of Experiment 1. However, the exploratory analysis showed a different pattern when taking listeners' gender into account. For male listeners, no difference in truth judgments was obtained between the adult and child talker, but for female listeners, truth judgments were lower for the child talker than for the male talker. A difference between adult and child talkers was also found in Experiment 2 with female adult talkers and multiple talkers in Experiment 3, albeit in the other direction with higher ratings for the child talkers. Overall, findings therefore rather point to listeners' voice preferences playing a role in credibility judgments.

Moreover, the findings of Experiments 1-3 suggest that it was not exclusively vocal pitch that influenced the credibility ratings. If the results in Experiment 1 had been mainly caused by lower-pitch preference for the male adult talker, then the same pattern should have been found for female adult talkers in Experiment 2 and 3 as well, since they also had lower-pitch than the children. However, in Experiments 2 and 3, participants rated the child talkers higher than the female adults. It should be noted that the two talkers in Experiment 1 differed

not only in age (adult vs. child) but also in gender (male adult vs. female child). Based on Experiment 1 alone, we can thus not exclude the possibility that the gender of the talker rather than age caused the talker effect for female participants. Sexual dimorphism can be present in children as young as four years of age. Also typically, boys, at the age of seven or eight, have somewhat lower formant frequencies in comparison to girls (Vorperian & Kent, 2007). If gender rather than age of the talker influenced credibility ratings, there should be no difference in ratings between a female adult talker and a female child. This is, however, not what was found in Experiment 2, making it less likely that the findings were driven by the gender of the talkers rather than their age.

When putting the exploratory findings of Experiment 1 together with the results of Experiment 2, it comes to light that both results rather lend support to the voice preference hypothesis, but it is not strictly about the influence of vocal pitch on credibility ratings, it is more about liking voices in general. If vocal pitch played a greater role in credibility judgments, then child talkers should have received lower credibility ratings than female adult talkers. The fact that female listeners rated female adult talkers as lower than children, makes it appear as if overall female listeners like men's voices most, followed by children's and then their own voices: Men's voices possibly because of their low pitch (e.g., Collins, 2000; Rezlescu et al., 2015) and for children we can only speculate that likeability had a positive effect on the ratings for women. Since male participants were only tested in Experiment 1, we do not have the complete picture of the effect of voice preferences across gender yet. Possibly the null effect for male participants in Experiment 1, was a tug of war between the likeability of the child talker and a voice preference for low-pitch voices that is less pronounced than for women. To what extent voice preferences influence ratings for male participants is a topic that will need to be investigated in future

research.

For the present four experiments, we interpret the results as not being influenced by processing fluency but an indirect relationship between credibility judgment and talker age, possibly mediated by voice preference - at least in the native language context of German.

CHAPTER 7

General Discussion

Abstract

This chapter provides a summary of the major findings of this dissertation and their interpretation, answering the overarching research question: What is the role of talker information in spoken language comprehension? The findings are then related to theories of auditory and audiovisual comprehension of spoken language and its non-direct impact on socio-linguistic aspects. The dissertation ends with an outlook on possible directions for future research.

Summary of the results

In general terms, this dissertation set out to take a closer look at the role of talker information in speech comprehension. Speech consists of a multifaceted array of information, and it entails information both about the message (i.e., linguistic information) and the messenger (i.e., indexical information). More specifically, this dissertation focused on the effect of child speech on spoken language comprehension. While the role of talker information has been investigated from various angles, using for example talkers that vary in gender or language background, no previous research has looked at child speech.

Similar to child speech, foreign-accented talkers, are known to deviate in their pronunciation from the standard norms of the target language (Bosker et al., 2014; Eisner et al., 2013; Lev-Ari, 2015). Previous studies on foreign-accented speech found that variation (e.g., in pronunciation or grammatical errors) in the speech signal can have a negative effect on sentence comprehension (Goslin et al., 2012), and listeners utilize non-natives' idiosyncrasies to adapt to them in order to understand what has been said (Hanulíková et al., 2012). Thus, opposing traditional theories of spoken-word recognition (e.g., McClelland & Elman, 1986; Norris, 1994), research on foreign-accented speech suggested that social aspects are indeed relevant for understanding speech (Bosker et al., 2014; Eisner et al., 2013; Lev-Ari, 2015). Thus, at least in the case of foreign-accented talkers, the non-nativeness of talkers has been found to modulate the comprehension process, thereby illustrating the importance of talker information in spoken language comprehension. We wanted to broaden our understanding of the role of talker information by focusing on child speech and to a lesser extent on non-native speech. In order to address this general topic, the experiments described in this dissertation approached talker information from three different angles, using several methods:

1. To examine talker information in an audio-only speech scenario with cross-modal priming.

2. To investigate the role of talker information in an audio-visual speech scenario with cued-recall and a speech intelligibility test.

3. To assess talker information in the socio-linguistic context, that is, the effect of talker information on listeners' trustworthiness in a rating task.

The aim of **Chapter 4** was to investigate the role of talker information in an audio-only speech scenario. More specifically, Experiment 4.1 looked at the influence of age of the talker on phonetic-to-lexical mapping. Previous research had found that listeners adapt to and improve their understanding of foreign-accented productions such that the same deviations are handled differently depending on the nativeness of the talker (Bosker et al., 2014; Eisner et al., 2013; Lev-Ari, 2015), but these studies solely looked at deviations in pronunciation of foreign-accented talkers. Child speech is similar in this regard because children's pronunciation also typically deviates from the standard norms of native adult speech. We tested native German adult listeners in a cross-modal priming experiment. Specifically, L1 German participants listened to German word fragment primes (e.g., *Para-* from *Parasit*, "parasite") that mismatched in the second vowel with visual target words (e.g., *Parodie*, "parody") and were produced by an adult or child talker. After listening to a fragment prime, participants made lexical decisions to the visual target word via button press, indicating whether they considered the visual string of letters a real word of German or not. Initially, we did not find a talker age effect on phonetic-to-lexical mapping. However, when following the NRV framework (Polka & Bohn, 2011), which proposes directional asymmetries in the discrimination of speech sounds, subsequent exploratory analyses showed different

facilitatory priming patterns for the adult talker and child talker. While priming was found for the adult talker, no priming was found for the child talker. For the adult voice in particular, participants showed facilitatory priming such that onset fragments primed target word recognition when the anchor vowels /uː/, /iː/, or /aː/ occurred in the prime (e.g., *Para-* primed *Parodie*). By contrast, no priming occurred when the anchor vowels appeared in the target (e.g., *Paro-* did not prime *Parasit*). Since no facilitatory effect was observed for the child talker, the findings suggested that prior experience with the linguistic competence of child speech might have caused this effect. We provided two explanations for this interpretation. First, vowel information of the child was never considered a reliable marker for the lexical mapping process. Second, previous experience with child speech might have led participants to tolerate all vowel mismatches for the phonetic-to-lexical mapping processes. This is in alignment with research on foreign-accented speech because previous experiences have been found to make deviations in pronunciation acceptable matches for canonical pronunciations (e.g., Eisner et al., 2013; Friedrich et al., 2013; Trude et al., 2013).

To test the reliability of the exploratory findings found in Experiment 4.1, we attempted to replicate the results in Experiment 4.2, using the same paradigm. The stimuli, however, had to be adjusted: word pairs from Experiment 4.1 that contained no such change (i.e., anchor vowel to anchor vowel or non-anchor vowel to non-anchor vowel) were excluded from the experimental items. The procedure was furthermore identical. Contrary to Experiment 4.1, the results of Experiment 4.2 showed that the exploratory effect could not be reproduced when only word pairs that included a change in anchoring were used. We offered two possible explanations for this null-effect. First, it is possible that generally the methodology did not suit an investigation with the NRV framework very well (Riedinger et al., 2021). Second,

our talkers appeared in random order and were not presented in blocks, thus causing a potential spillover effect between talkers, which might have been the key factor for the lack of talker age effect.

Taken together, the experiments in Chapter 4 addressed the role of talker information in an audio-only speech scenario and found some evidence for an indirect effect of talker age on phonetic-to-lexical mapping. That is, exploratory analysis in Experiment 4.1 found an influence of talker age on phonetic-to-lexical mapping processes within the NRV framework. We interpreted this result as part of an effect of previous experience with the linguistic competence of child speech. Although we were not able to replicate the findings in Experiment 4.2, we are not convinced that the null-effect provides a substantial argument that talker age does not play a role in the mapping of phonetic-to-lexical representations. Rather, the diverging results in both experiments might be subject to the experimental design, as suggested by previous investigations which also found conflicting results depending on the method employed (De Rue et al., 2021; Polka et al., 2021; Riedinger et al., 2021).

In **Chapters 5.1** through **5.3**, we further investigated the role of talker information in an audiovisual speech scenario, thus addressing the examination from the second angle mentioned above. Previous research has shown that visual articulatory information, such as lip and jaw movements, provide important phonological information about speech sounds (e.g., Campbell, 2008; Summerfield, 1992), which can aid spoken language comprehension (Navarra & Soto-Faraco, 2007; Sumby & Pollack, 1954). In light of the Covid-19 pandemic, we wanted to investigate whether the lack of visual information can impede our memory since the wearing of face masks has become part of our daily lives, consequently making face-to-face communication more challenging. Due to the pandemic, we conducted the series of experiments online and used a cued-recall task. Participants were shown

video recordings of a native German adult talker, producing six sentences per block (e.g., *Die Köchin hilft montags armen Kindern*, "The cook helps on Mondays poor children") with and without a face mask. The task of the participants consisted of memorizing the final two words for each sentence. After presenting a video block, participants were asked to type in the two missing final words (e.g., *armen Kindern*, "poor children") on their keyboard.

Chapter 5.1 served as the starting point for investigating whether participants remember words less well when having no access to the mouth region of the talker (i.e., movements of the lips and the jaw). Native German participants completed a cued-recall task and showed that it was much harder to remember words when the mouth region was covered by a face mask. This is in line with the Framework for Understanding Effortful Listening (Pichora-Fuller et al., 2016), the Effortfulness Hypothesis (McCoy et al., 2005), and the Ease of Language Understanding (Rönnberg et al., 2013), stating that increased listening effort results in higher cognitive processing load (Peelle, 2018), which in turn depletes mental resources reserved for cognitive functions such as memory encoding (Rabbitt, 1991). In addition, the findings were in line with recent face mask studies (Bottalico et al., 2020; V. A. Brown et al., 2021; Corey et al., 2020; Randazzo et al., 2020; Smiljanic et al., 2021).

Chapter 5.2 used the same experimental design as Chapter 5.1, but this time we extended the scope by adding a child talker to the stimuli and increasing participant numbers. In addition to that, Chapter 5.2 implemented an intelligibility task to measure whether or not speech produced with a face mask is less intelligible than without a face mask. For this task, white noise was embedded to the video. Two major findings emerged from this study. Firstly, the analysis of correctly recalled items showed again that participants remembered fewer words when the speaker

had been wearing a face mask than when she had not been wearing one, and this was equally true for both adult and child talker, thus supporting the results in Chapter 5.1. Secondly, results from the speech intelligibility task confirmed that speech produced with a face mask was harder to understand compared to speech produced without a face mask.

After having established that native German participants have difficulties understanding and storing words when produced with a face mask, **Chapter 5.3** explored the further impact of face masks on memory for non-native participants (L2). Thus, we tested non-native German participants for this experiment. Previous research has shown that L2 listeners can make use of visual cues to improve speech perception (Massaro, 1998) to make up for the generally lower comprehension skills in their L2 language (Drijvers & Özyürek, 2017), and this has been demonstrated for various language contexts (Davis & Kim, 2004; Erdener & Burnham, 2005; Reisberg et al., 1987). Our results, however, showed no effect of visual cues and just a generally poor performance across conditions, resulting possibly in a floor effect. Very likely, this floor effect has been driven by the low language proficiency of our L2 participants. Although the results of Chapters 5.1 and 5.2 were not replicated for L2 participants, they nevertheless support previous studies that state that language proficiency is crucial when perceiving audiovisual speech (Wang et al., 2008; Xie et al., 2014). In this sense, the results strengthened the possibility that advanced linguistic expertise can enhance the extraction of visual speech information for subsequent higher cognitive performance, especially when it is in the non-native language (Vejnovic et al., 2010).

Overall, Chapters 5.1 to 5.3 explored talker information from the second angle mentioned above. Although our findings did not provide evidence for a talker effect in an audiovisual setup, the results were encouraging in the sense that masked

child speech was not disadvantaged more strongly than masked adult speech in face-to-face communication. Our findings, therefore, provide a better understanding of the impact of the pandemic and its potential implications of face masks in various communicative situations like in classrooms and a doctor's appointment, where recalling spoken information is important.

In summary, Chapters 4 to 5.3 found mixed results of talker age in the linguistic context. While talker information played an indirect role in the audio-only scenario (Chapters 4), no effect was found in the audio-visual scenario (Chapter 5.1 to 5.3). We then asked the question of whether an indirect linguistic consequence of talker age effect was possible. This question was posed in Chapter 6, in which we investigated the role of talker information in the social context of credibility judgments, which led us to the investigation of talker information from the third angle.

Chapter 6 examined whether talker information can have an influence on social evaluations like credibility judgments. More specifically, we were interested in whether listeners would believe information less when it is delivered by a child talker than by an adult talker. Previous research conducted by Lev-Ari and Keysar (2010) and Boduch-Grabka and Lev-Ari (2021) found that people believed information less when produced by a foreign-accented talker than when produced by a native talker. The explanation was that foreign-accented speech was harder to understand and thus processed less fluently than native speech, which in turn could impact listeners' credibility judgments (e.g., Oppenheimer, 2008; N. Schwarz, 2004; Unkelbach, 2006). In the four-part experiment, we wanted to see whether children's speech is also associated with reduced credibility, as was shown for foreign-accented speech (Boduch-Grabka & Lev-Ari, 2021; Lev-Ari & Keysar, 2010). To this end, both German native adult and child talkers were recorded and produced German

trivia sentences like *Ameisen schlafen nicht* (i.e., "Ants don't sleep"). Only trivia statements for which the correctness was typically not known by participants were selected. This design choice was essential to allow differences in credibility ratings to emerge when the same sentences were spoken by different talkers. Filler sentences were added and participants could be certain of their truth value (e.g., *Brokkoli ist ungesund*, "broccoli is unhealthy"). Participants gave their truth judgments by using a slider scale, with the left end of the scale marked with "definitely false", and the right end marked with "definitely true".

Experiment 6.1 served as a baseline to see whether native German participants find statements made by a child less credible than by an adult. In this experiment, a male adult talker and a female child talker, who were both native Germans, produced the trivia sentences. Native German listeners judged the truthfulness of trivia statements and initially showed no difference in truthfulness ratings. When we looked closer at the data, we found that female listeners' voice preferences affected the relation between credibility judgments and talker age. Female participants rated sentences spoken by the male adult talker as more credible than sentences spoken by the female child talker. Voice studies in fact suggest that an attractive voice is typically associated with trustworthiness (Sell et al., 2010; Tigue et al., 2012; Xu et al., 2013).

This motivated us to conduct **Experiment 6.2** in which participants listened to the same statements, but this time the adult talker was replaced by a female adult talker. Therefore, Experiment 6.2 did not only take a closer look at the relationship between voice preference and credibility but was also extending the scope that allowed us to go beyond the concept of processing fluency. Contrary to Experiment 6.1, the results of Experiment 6.2 demonstrated that female participants judged trivia statements now as less true when the sentences were spoken by the

female adult talker compared to when the talker was a female child. With the goal to assess whether listeners' ratings from female participants differed across experiments, we then merged the data from Experiments 6.1 and 6.2. Results demonstrated that female listeners judged the female talker as less credible than the child talker in Experiment 6.2, corroborating the notion that the effect found in Experiment 6.1 was caused by factors (e.g., attitudes) other than inhibition of processing fluency.

Building up on this, **Experiment 6.3** replicated the results with multiple adult and child talkers (four female adults and four female children), to rule out the possibility that some uncontrolled voice characteristics of the individual talkers were the main cause for the results in Experiment 6.2. The pattern of results of Experiment 6.3 remained robust. Nonetheless, no direct link between talker age and processing fluency had been observed so far.

This motivated us to challenge the hypothesis one more time: Are credibility judgments influenced by processing fluency? **Experiment 6.4** used the same children as in Experiment 6.3 but replaced the four female adult talkers with four non-native female adult talkers (i.e., Arabic, Chinese, Spanish, and Russian). Once again, no direct evidence for an influence of processing fluency was found in the comparison of foreign-accented speech and child speech. Finally, we interpreted the findings in Chapter 6 as not being affected by processing fluency but that the relationship between credibility judgment and talker age was possibly mediated by voice preference, thus illustrating an indirect effect of talker information on social evaluations.

Talker identity in the perspective of theories of speech comprehension

As described in the introduction of this thesis, speech carries both linguistic and indexical information. Specifically, while linguistic information entails information about the content of an utterance like phonological, syntactic, and semantic information (Levi & Pisoni, 2007), indexical properties entail information about the talker like age, gender, and language background. Therefore, talkers introduce substantial variability in the acoustic realizations of speech (Abercrombie, 1967). Although indexical information presents significant variability in the speech input, it does not automatically decrease its intelligibility, but it may increase cognitive demands for processing (Aydelott & Bates, 2004). Our findings support this notion, because no participant in any of our studies reported having any problems understanding the child and adult speech. As such, speech variation introduced by child talkers did not affect the quality and intelligibility of the auditory input. However, it might have increased cognitive demands which in turn could have influenced speech processing. Overall, we have obtained mixed findings on the role of talker information on speech comprehension, indicating that the influence of child speech variation on adult listening is complex.

To the best of our knowledge, we were the first to investigate the effects of indexical information in children's speech, but the influence of indexical information for spoken-language comprehension has previously been investigated, for example, for non-native speech. Research has shown for instance that listeners' familiarity with foreign-accented speech and the likelihood of pronunciation deviations have modified processing (Hanulíková & Weber, 2012). Specifically, native listeners were found in an ERP study to be more forgiving of grammatical errors produced by

non-native talkers than by native talkers (Hanulíková et al., 2012), and they have been found to relax their vowel categories to accept deviating forms more willingly for non-native talkers than for native talkers (Hay, Nolan, & Drager, 2006). This might be due to the fact that non-native speech, that is, speech produced by second language talkers, is even more variable, both within talkers and across talkers (Nissen et al., 2007; Wade et al., 2007) than children's speech. Non-native talkers are often recognizable by an accent in pronunciation or by grammatical errors, as it is difficult to achieve native-like proficiency in a second language. The typical deviations from the target norms of a language are mostly due to interference from the talkers' native language (Bent & Bradlow, 2003). While there are clear findings for native listeners to include knowledge about the non-nativeness of talkers during spoken language comprehension, our results merely indicated hints of a possible impact of native child speech on adult listening. Most intriguing, our findings cannot be strictly explained by models of spoken-language comprehension. Therefore, the results presented in this thesis have important implications for traditional models of speech comprehension.

There is a long-standing debate between the proponents of abstractionist and episodic models of speech comprehension (see also Chapter 2) on the role of indexical information. An immediate influence of indexical information on comprehension challenges abstractionist accounts of spoken-language comprehension (McClelland & Elman, 1986; Norris, 1994), because they assume that lexical entries, which are represented in memory in abstract phonological codes, are the most relevant information for word identification (Luce & Lyons, 1998). Hence, variation introduced by factors like the identity of the talker is treated as irrelevant information for the initial comprehension process and is discarded early in the encoding process. In other words, listeners discard any information that entails

surface acoustic variability to arrive at abstract phonological codes, which is also known as the *normalization* process (Ladefoged & Broadbent, 1957; Nygaard & Pisoni, 1998). In other words, abstractionist accounts have focused more on the processing of linguistic information, and the role of indexical information was largely ignored. Episodic accounts (Goldinger, 1998), on the other hand, challenge abstractionist views by supporting the notion that indexical properties are retained and always encoded alongside linguistic information.

In the literature, there have been substantial findings in support of both abstractionist and episodic accounts on the sound, lexical, and sentence level (see also Chapter 2), but the results of this dissertation were not as unequivocal as some previous studies. Generally speaking, results from Chapter 4 support abstractionist accounts since no evidence for a talker effect on phonetic-to-lexical mapping was obtained. However, when using the NRV framework for post-hoc interpretation a talker effect was found, which can be loosely explained by episodic models. As mentioned above, episodic models postulate that indexical information is stored in the memory of the listener, which can influence processing (Goldinger, 1998). As such, previous experience with child speech possibly made vowel information of the child voice an unreliable indicator for the lexical mapping process. This is in line with research on foreign-accented speech, showing that previous experience has been found to tolerate deviations in pronunciations (e.g., Eisner et al., 2013; Trude et al., 2013; Zwitserlood, 1989). By contrast, Experiment 2 of Chapter 4 showed no impact of talker information on the mapping of phonetic-to-lexical representations, thus possibly supporting abstractionist accounts such that indexical information is discarded early in the process. The mixed findings illustrate that our results cannot be strictly explained by either abstractionist or episodic accounts. In fact, there is emerging evidence showing a coexistence of both abstract lexical representations and

talker-specific information. For example, behavioral studies showed facilitation when words were produced by familiar talkers, thus indicating that both indexical and linguistic information is encoded and stored together in representations of spoken words in memory (Goldinger et al., 1991; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Palmeri et al., 1993; Schacter & Church, 1992).

To this end, a comprehensive model of speech comprehension should be able to account for both types of information: those highlighting the role of abstract information in speech comprehension and those underlying the importance of indexical information. Hybrid theories have been developed to answer the question of how abstract and indexical information is stored and when indexical information plays a role during spoken-language comprehension (e.g., Luce & Lyons, 1998). We will discuss our results from the point of view of the time-course hypothesis (Luce, McLennan, & Charles-Luce, 2003). Luce and Lyons (1998) propose that details of a talker's voice are not available as early as abstract representations during processing, because in their repetition priming experiment participants responded faster to words in an old/new word judgment task than in a lexical decision task when produced by a familiar talker. This pattern was successfully reproduced in subsequent studies (e.g., C. Y. Lee & Zhang, 2015; Luce et al., 2003). For instance, Luce et al. (2003) demonstrated an indexical effect for longer and lower-frequency words in a repetition priming experiment. Hence, McLennan and Luce (2005) put forward the idea that if processing conditions are ideal, identification is fast and effortless. Consequently, abstract forms will dominate speech processing and indexical information is moved to the background. However, this effect is not solely limited to repetition priming. C. Y. Lee and Zhang (2015), for example, found an effect of talker variability in both repetition priming and semantic priming, both of which involved a lexical decision task. However, the magnitude of the effect

was smaller for semantic priming than for repetition priming, suggesting that talker information is less evident in processing word meaning compared to processing word form. This effect may be attributed to the fact that word meaning presumably involves a deeper level of processing than word form, suggesting that processing talker information occurs primarily at a relatively shallow level (Goldinger, 1996).

Following this interpretation, it is conceivable that participants in Chapter 4 did not have enough time to process indexical information. Importantly, the talker effect was only obtained when we conducted an exploratory analysis with the focus on the NRV framework. When attempting to reproduce the findings in Experiment 2, a few modifications had to be done. The major change was that some word pairs had to be newly paired, which resulted in an exclusion of sixteen word pairs. We suspected that these adjustments came at the expense of participants having less time and opportunity to process talker information. Although the time-course hypothesis gives an interesting explanation for our findings, the present results still cannot provide a definite answer to the question about when indexical information becomes relevant in the comprehension process as the time course was not the focus of our investigation. At the same time, Chapters 5.1 to 5.3 showed no influence of talker information in higher cognitive processes like memory encoding, which again shows support for the time-course hypothesis, but since the experimental setup differed significantly from the studies described above, the interpretation has to be approached with caution. They predominantly addressed it from the audio-only perspective (e.g., C. Y. Lee & Zhang, 2015; Luce, 1986; Luce & Lyons, 1998; McLennan & Luce, 2005) whereas Chapters 5.1 to 5.3 approached the role of indexical information from the audio-visual perspective.

The addition of visual cues to auditory cues is known to augment speech understanding. Hence, being able to see the talker's lip and jaw movements presents

a distinct advantage for participants, presenting a substantial improvement in speech recognition (Jesse & Janse, 2012; Massaro, 1987). The addition of visual information to the auditory input probably enhanced the comprehension process, thereby potentially diminishing the talker effect in the audio-visual scenario. However, this argument does not hold very well because, as stated above, participants reported having no difficulties understanding child speech in any of the studies. In fact, processing auditory-visual information also comes with its disadvantages as it is more demanding to combine sources from two modalities, which requires additional cognitive resources (Pichora-Fuller et al., 2016). Hence, it seems plausible that the lack of a talker effect was due to the task's design, since memory encoding likely resulted in a weak encoding of indexical cues. As such, participants' attention was distracted and not focused on surface information such as acoustic variability of the stimuli (e.g., Kittredge, Davis, & Blumstein, 2006). Thus, Chapters 5.1 to 5.3 lend support to McLennan and Luce (2005) stating that processing talker information requires time. However, we would not want to go so far as to say that indexicality does not matter in later stages of processing, because findings of Chapter 6 showed different patterns of talker effects.

The diverging patterns suggested that talker information was not discarded from further processing. Rather, talker information was preserved in memory and was accessed, possibly triggered by certain contextual situations. For example, we found different results depending on the pairing of the underlying factors like gender, age, and language background. Somewhat provoking yet intriguing was the result that male adults were judged as more believable when paired with female children; however, we did not find this pattern when pairing the same children with female adults. When comparing child speech with non-native female adult speech, non-native adults were not rated as less credible than native children. This

is somewhat comforting, because the findings showed that prejudice did not form native adults' evaluations. That said, depending on the social context, talker information opens the door to explore and understand social differences that are conveyed through indexical cues.

Taken together, this discussion set talker information into the perspective of current theoretical models of spoken-language comprehension. We started our discussion with the observation that the scientific endeavor aimed at understanding how listeners understand spoken language has still not fully understood how indexical information influences speech comprehension. Emerging evidence challenges traditional views and argues for talker information being an integral part in speech communication. Hence, without any modifications, traditional models cannot account for the present findings. Therefore, hybrid theories proposed a coexistence between linguistic and indexical information during speech comprehension.

Although this dissertation provided mixed talker effects, our findings add to the growing body of literature that challenges long-standing theoretical paradigms that postulate exclusive processing of linguistic information. These challenges may lead to the next generation of models of spoken language comprehension that account for both message and messenger.

Directions for future research

We are still at the beginning stages of the investigation on the influence of child speech, as one form of variation, on spoken-language comprehension, thus leaving many questions open for future research studies.

One open question, for example, concerns the influence of speech by children who are younger than seven years of age. In this thesis, the child speech stimuli were produced by children who were between the ages of seven and eleven years old. A serious challenge for studies using child speech is to find the immaculate balance between the desired speech variation and the ability of child participants to cooperate during recording.

Thus, utilizing this particular age range for our child speech stimuli can be attributed to two main reasons: First, children at this age can understand recording instructions without any problems and have enough stamina to make it through a full recording session. Second, they can produce the items as intended without any obvious hesitations and disfluencies. These criteria came with the advantage of recording child talkers in a highly controlled recording environment. In this thesis, we focused on child speech which introduced some variations in pronunciation but which were still intelligible to adults' ears but less intelligible than adult speech. We predicted that such variation may evoke additional cognitive resources and mental effort in the comprehension process, similar to foreign-accented speech (Van Engen & Peelle, 2014). Although our mixed findings demonstrated hints of an influence of child speech on spoken-language comprehension, the findings never showed a clear disadvantage for child talkers. Possibly this pattern is co-determined by the fact that our child talkers were relatively advanced in their language production skills which made their pronunciations perhaps too similar to native adult speech, even

though their pronunciation introduced some variation in the speech input and their voices were clearly child voices.

This makes it all the more interesting to investigate speech produced by younger child talkers which extends the scope of the present thesis and thereby paints a broader picture of the role that child speech variation may have in spoken-language comprehension. Generally, younger children have less control over their motor skills and are therefore less accurate in their pronunciation, thus introducing more variation to the speech input. Speech from younger children deviates more strongly from adult norms as opposed to the speech stimuli used in this thesis. Thus, it leaves open the question of what effects may occur. Stronger deviation can reduce intelligibility (Bent & Bradlow, 2003; Munro & Derwing, 1995) and consequently increase listening effort which in turn can obstruct comprehension (Van Engen & Peelle, 2014) and memory encoding (Romero-Rivas, Thorley, Skelton, & Costa, 2019). With this in mind, it seems possible that different result patterns might occur. Potentially, effects may vary according to speech variation strength. Importantly, stronger variation does not necessarily equal stronger effects. Sometimes, weaker effects may emerge for stronger variation as was shown in research on foreign-accented speech. For instance, free recall was significantly affected when the talker had a strong foreign accent compared to a mild foreign accent or a native accent (Romero-Rivas et al., 2019). This is in parallel with findings focusing on word recognition. Witteman et al. (2013), for example, tested how varying strengths of foreign accents and listeners' experience with the accent can influence word recognition. Witteman et al. (2013) observed weaker effects for stronger variation compared to mild or weak accents. Specifically, Dutch listeners correctly interpreted Dutch words when they were produced with a weak or mild German accent, but they had considerably more difficulties recognizing

strongly-accented words, thus showing weaker effects. These effects were modulated by prior experience, illustrating that adaptation to foreign accents depends on accent strength and on listeners' experience.

This brings us to another possible avenue for future research, and that is whether experience with child speech can enhance comprehension. Hence, including listeners' experience as a variable might further advance this area. Research on foreign-accented speech has consistently shown that listeners can readily handle variation in speech by foreign-accented talkers despite it being more challenging initially (e.g., Bradlow & Bent, 2008; C. Clarke & Garrett, 2004; Witteman et al., 2013). Again, Witteman et al. (2013) showed that adaptation to foreign accents was particularly modulated by listeners' experience with the accent. For instance, Dutch listeners with ample prior experience had less problems interpreting strongly German accented words as opposed to listeners with little experience. Following this line of logic, we propose that interpreting speech of young children might be significantly harder for listeners with less or no exposure to child speech compared to listeners who have extensive experience with child talkers (e.g., caregivers, preschool teachers, and parents). This would help us better understand whether familiarity with child speech can have a positive effect on child speech comprehension.

Taken together, child speech variation is an understudied field and much research needs to be done to expand our understanding of the influence of child speech variation on spoken-language comprehension. Future research needs to bear in mind that additional challenges need to be taken into account when utilizing speech samples from younger child talkers. As previously stated, finding the appropriate balance between the level of speech variation and cooperation from child talkers, especially with younger children, seems to be a challenging task. It is seemingly easier to select foreign-accented talkers for stronger speech variation

rather than younger children, as a controlled recording environment cannot easily be ensured when working with children. Future research might therefore need to take into account adjusting recording instructions and possibly adapting experimental methodology, such as investigating the perception of spontaneous child speech for younger talkers. This might mirror a naturalistic listening situation which much resembles real-life speech perception conditions, thus shedding more light onto the phenomenon of child speech variation.

Bibliography

Abercrombie, D. (1967). *Elements of general phonetics*. University of Chicago Press.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

Anderson, S., Downs, S. D., Faucette, K., Griffin, J., King, T., & Woolstenhulme, S. (2007). How accents affect perception of intelligence, physical attractiveness, and trustworthiness of middle-easter-, latin-american-, british- and standard-american-english-accented speakers. *Intuition: The BYU Undergraduate Journal in Psychology*, *3*(1), 5–11.

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. (2019). Gorilla in our midst: An online behavioral experiment builder. *bioRxiv*. `https://doi.org/10.1101/438242`

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, *37*(5), 715–727. `https://doi.org/10.1037/0022-3514.37.5.715`

Arkes, H. R., Hackett, C., & Boehm, L. (1989). The generality of the relation between familiarity and judged validity. *Journal of behavioral decision making*, *2*(2), 81–94.

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British journal of psychology*, *92*(2), 339–55.

Aydelott, J., & Bates, E. (2004). Effects of acoustic distortion and semantic context on lexical access. *Language and Cognitive Processes*, *19*(1), 29–56. https://doi.org/10.1080/01690960344000099

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using r*. Cambridge: Cambridge University Press.

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

Bacon, F. T. (1979). Credibility of repeated statements. memory for trivia. *Journal of Experimental Psychology: Human learning and memory*, *5*(3), 241–252.

Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2020). *Memory* (3rd ed.). Routledge.

Bala, N., Ramakrishnan, K., Lindsay, R., & Lee, K. (2005). Judicial assessment of the credibility of child witnesses. *Alberta Law Review*, *42*(4), 995–1017. https://doi.org/10.29173/alr1270

Barr, D. J., & Keysar, B. (2006). Perspective taking and the coordination of meaning in language use. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (Second Edition ed., pp. 901–938). Academic Press. https://doi.org/https://doi.org/10.1016/B978-012369374-7/50024-9

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. https://doi.org/10.48550/arXiv.1506.04967

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using eigen and s4: R package. *Journal of Statistical Software*, *67*(1), 1–48. Retrieved from http://CRAN.R-project.org/package=lme4 https://doi.org/10.18637/jss.v067.i01

Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. *Psychological research*, *74*(1), 110-20. https://doi.org/10.1007/s00426-008-0185-z

Baus, C., McAleer, P., Marcoux, K., Belin, P., & Costa, A. (2019). Forming social impressions from voices in native and foreign languages. *Scientific Reports*, *9*(1), 414. https://doi.org/10.1038/s41598-018-36518-6

Begg, I., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, *121*(1), 446–458. https://doi.org/10.1037/0096-3445.121.4.446

Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, *102*(4), 711–725. https://doi.org/10.1111/j.2044-8295.2011.02041.x

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acousticsl Society of America*, *114*(3), 1600-1610. https://doi.org/10.1121/1.1603234

Bernstein, L., Auer, E., & Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lipreading. *Speech Communication*, *44*, 5-18. https://doi.org/10.1016/j.specom.2004.10.011

Best, C. (1995). A direct realist view of crosslanugage speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium: New York Press.

Black, J. B., & Bern, H. (1981). Causal coherence and memory for events in narratives. *Journal of Verbal Learning and Verbal Behavior*, *20*(3), 267–275.

Boduch-Grabka, K., & Lev-Ari, S. (2021). Exposing individuals to foreign accent increases their trust in what nonnative speakers say. *Cognitive Science*, *45*(11), e13064. https://doi.org/10.1111/cogs.13064

Boersma, P., Weenink, D. (2018). *Praat: doing phonetics by computer.*

Bordon, G. J., & Gay, T. (1979). Temporal aspecrts of articulatory movements for /s/-stop clusters. *Phonetica*, *36*(1), 21–31.

Bosker, H. R., Quené, H., Sanders, T., & De Jong, N. H. (2014). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language*, *75*, 104–116. https://doi.org/10.1016/j.jml.2014.05.004

Bottalico, P., Murgia, S., Puglisi, G. E., Astolfi, A., & Kirk, K. I. (2020). Effects of masks on speech intelligibility in auralized classrooms. *The Journal of the Acoustical Society of America*, *148*(5), 2878–2884. https://doi.org/10.1016/j.jml.2014.05.004

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Brashier, N. M., & Marsh, E. J. (2020). Judging truth. *Annual review of psychology*, *71*, 499–515. https://doi.org/10.1146/annurev-psych-010419-050807

Brashier, N. M., Umanath, S., Cabeza, R., & Marsh, E. J. (2017). Competing cues: Older adults rely on knowledge in the face of fluency. *Psychology and aging*, *32*(4), 331–337. https://doi.org/10.1037/pag0000156

Bricker, P. D., & Pruzansky, S. (1966). Effects of stimulus content and duration on talker identification. *The Journal of the Acoustical Society of America*, *40*(6), 1441-1449. https://doi.org/10.1121/1.1910246

Briñol, P., Petty, R. E., & Tormala, Z. L. (2006). The malleable meaning of subjective ease. *Psychological Science*, *17*(3), 200–206. https://doi.org/10.1111/j.1467-9280.2006.01686.x

Brown, A. S., & Nix, L. A. (1996). Turning lies into truths: Referential validation of falsehoods. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1088–1100. https://doi.org/10.1037/0278-7393.22.5.1088

Brown, V. A., Van Engen, K. J., & Peelle, J. E. (2021). Face mask type affects audiovisual speech intelligibility and subjective listening effort in young and older adults. *Cognitive Research: Principles and implications*, *6*(1), 49. https://doi.org/10.1186/s41235-021-00314-0

Burgoon, J. K., T., B., & Pfau, M. (1990). Nonverbal behaviors, persuasion, and credibility. *Human communication research*, *17*, 140–169. https://doi.org/10.1111/j.1468-2958.1990.tb00229.x

Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philosophical transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1001-1010. https://doi.org/10.1098/rstb.2007.2155

Campbell-Kibler, K. (2010). Sociolinguistics and perception. *Language Linguistics Compass*, *4*(6), 377–389. https://doi.org/10.1111/j.1749-818X.2010.00201.x

Casasanto, L. S. (2008). Does social information influence sentence processing? *Proceedings of the Annual Meeting of the Cognitive Science Society*, *30*(30), 799–804.

Chaiken, S. (1979). Communicator physical attractiveness and persuasion. *Journal of Personality and social Psychology*, *37*(8), 1387. https://doi.org/10.1037/0022-3514.37.8.1387

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

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

Clarke, C., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented english. *The Journal of the Acoustical Society of America*, *116*(6), 3647–58. https://doi.org/10.1121/1.1815131

Clarke, C. M. (2002). Perceptual adjustment to foreign-accented english with short term exposure. In *Proceedings of the 7th international conference on spoken language processing.*

Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour*, *60*(6), 773–780. https://doi.org/10.1006/anbe.2000.1523

Cook, C., Heath, F., Thompson, R. L., & Thompson, B. (2001). Score reliability in web- or internet-based surveys: Unnumbered graphic rating scales versus likert-type scales. *Educational and Psychological Measurement*, *61*(4), 697–706. https://doi.org/10.1177/00131640121971356

Cooper, A., Fecher, N., & Johnson, E. K. (2020). Identifying children's voices. *The Journal of the Acoustical Society of America*, *148*(1), 324–333. https://doi.org/10.1121/10.0001576

Corey, R. M., Jones, U., & Singer, A. C. (2020). Acoustic effects of medical, cloth, and transparent face masks on speech signals. *The Journal of the Acoustical Society of America*, *148*(4), 2371-–2375. https://doi.org/10.1121/10.0002279

Creel, S. C., & Bregman, M. R. (2011). How talker identity relates to language processing. *Language and Linguistics Compass*, *5*(5), 190–204. https://doi.org/10.1111/j.1749-818X.2011.00276.x

Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., & Floccia, C. (2012). Linguistic processing of accented speech across the lifespan. *Frontiers in Psychology*, *3*.

Crystal, D., & Crystal, B. (2014). *You say potato: A book about accents*. Macmillan.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, *34*(2), 269–284. https://doi.org/10.1016/j.wocn.2005.06.002

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of english phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, *116*(6), 3668–3678. https://doi.org/10.1121/1.1810292

Dahan, D., & Magnuson, J. (2006). Spoken word recognition. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 249–283). Academic Press. https://doi.org/10.1016/B978-012369374-7/50009-2

Davis, C., & Kim, J. (2004). Audio–visual interactions with intact clearly audible speech. *The Quarterly Journal of Experimental Psychology Section A*, *57*(6), 1103–1121. https://doi.org/10.1080/02724980343000701

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, *14*(2), 238–257. https://doi.org/10.1177/1088868309352251

DeGroot, T., Aime, F., Johnson, S. G., & Kluemper, D. (2011). Does talking the talk help walking the walk? an examination of the effect of vocal attractiveness in leader effectiveness. *The Leadership Quarterly*, *22*(4), 680–689. https://doi.org/10.1016/j.leaqua.2011.05.008

De Groot, T., & Gooty, J. (2009). Can nonverbal cues be used to make meaningful personality attributions in employment interviews? *Journal of Business and Psychology*, *24*(2). https://doi.org/10.1007/s10869-009-9098-0

DeGroot, T., & Kluemper, D. (2007). Evidence of predictive and incremental validity of personality factors, vocal attractiveness and the situational interview. *International Journal of Selection and Assessment*, *15*(1), 30–39. https://doi.org/10.1111/j.1468-2389.2007.00365.x

DeGroot, T., & Motowidlo, S. (1999). Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *Journal of Applied Psychology*, *84*(6), 986–993. https://doi.org/10.1037/0021-9010.84.6.986

De la Vaux, D. W., S. K. Massaro. (2004). Audiovisual speech gating: Examining information and information processing. *Cognitive Processing*, *5*, 106–112. https://doi.org/10.1007/s10339-004-0014-2

De Meo, A. (2012). How credibly is a non-native speaker? prosody and surroundings. In S. A. Busà Maria Grazia (Ed.), *Methodological perspectives on second language prosody* (pp. 3–9). Papers from ML2P 2012.

De Rue, N. P. W. D., Snijders, T. M., & Fikkert, P. (2021). Contrast and conflict in dutch vowels. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.629648

Dewaele, J.-M., & McCloskey, J. (2015). Attitudes towards foreign accents among adult multilingual language users. *Journal of Multilingual and Multicultural Development*, *36*(3), 221–238. https://doi.org/doi.org/10.1080/01434632.2014.909445

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*(3), 285–290.

Dixon, J., Mahoney, B., & Cocks, R. (2002). Accents of guilt? effects of regional accent, race, and crime type on attributions of guilt. *Journal of Language and Social Psychology*, *21*(2), 162–168. https://doi.org/10.1177/02627X02021002004

Douglas, R. M., & Hemilä, H. (2005). Vitaminc c for preventing and treating the common cold. *PLoS*, *2*(6), e168–e217. https://doi.org/10.1371/journal.pmed.0020168

Drager, K. (2010). Sociophonetic variation in speech perception. *Language and Linguistics Compass*, *4*(7), 473–480. https://doi.org/10.1111/j.1749-818X.2010.00210.x

Drijvers, L., & Özyürek, A. (2017). Visual context enhanced: The joint contribution of iconic gestures and visible speech to degraded speech comprehension. *Journal of Speech, Language, and Hearing Research*, *60*(1), 212–222. https://doi.org/10.1044/2016_JSLHR-H-16-0101

Drijvers, L., & Özyürek, A. (2020). Non-native listeners benefit less from gestures and visible speech than native listeners during degraded speech comprehension. *Language and Speech*, *63*(2), 209–220. https://doi.org/10.1177/0023830919831311

Duller, D. B., LePoire, B. A., Aune, R. K., & Eloy, S. V. (1992). Social perceptions as mediators of the effect of speech rate similarity on compliance. *Human Communication Research*, *19*(2), 286–311. https://doi.org/10.1111/j.1468-2958.1992.tb00303.x

Edwards, J. (1999). Refining our understanding of language attitudes. *Journal of Language and Social Psychology*, *18*(1), 101-110. https://doi.org/10.1177/0261927X99018001007

Eisner, F., Melinger, A., & Weber, A. (2013). Constraints on the transfer of perceptual learning in accented speech. *Frontiers in Psychology*, *4*, 148. https://doi.org/10.3389/fpsyg.2013.00148

Erdener, V. D., & Burnham, D. K. (2005). The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning*, *55*(2), 191–228. https://doi.org/10.1111/J.0023-8333.2005.00303.X

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 540–551. https://doi.org/10.1016/S0022-5371(81)90165-1

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*(1), 162 - 173.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*(2), 345-360. https://doi.org/10.1016/j.wocn.2007.11.002

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*(5), 993–1002. https://doi.org/10.1037/xge0000098

Fazio, L. K., Rand, D. G., & Pennycook, G. (2019). Repetition increases perceived truth equally for plausible and implausible statements. *Psychonomic Bulletin, and Review*, *26*, 1705–1710. https://doi.org/10.3758/s13423-019-01651-4

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulation of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour*, *69*(3), 561–568. https://doi.org/10.1016/j.anbehav.2004.06.012

Ferguson, M. J., & Zayas, V. (2009). Automatic evaluation. *Current Directions in Psychological Science*, *18*(6), 362–366. https://doi.org/10.1111/j.1467-8721.2009.01668.x

Fisher, C. G. (1968). Confusions among visually perceived consonants. *Journal of Speech and Hearing Research*, *11*(4), 796–804.

Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, *4*(7), 258–267. https://doi.org/10.1016/S1364-6613(00)01494-7

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–273). Baltimore, York Press.

Floccia, C., Butler, J., Goslin, J., & Ellis, L. (2009). Regional and foreign accent processing in english: Can listeners adapt? *Journal of psycholinguistic research*, *38*, 379–412. https://doi.org/10.1007/s10936-008-9097-8

Frances, C., Costa, A., & Baus, C. (2018). On the effects of regional accents on memory and credibility. *Acta Psychologica*, *186*, 63-70. https://doi.org/https://doi.org/10.1016/j.actpsy.2018.04.003

Fraser, S., Gagné, J. P., Alepins, M., & Dubois, P. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: The effects of providing visual speech cues. *Journal of Speech, Language, and Hearing Research*, *53*(1), 18–33. https://doi.org/10.1044/1092-4388(2009/08-0140)

Freitas, A. L., Azizian, A., Travers, S., & Berry, S. A. (2005). The evaluative connotation of processing fluency: Inherently positive or moderated by motivational context? *Journal of Experimental Social Psychology*, *41*(6), 636–644. https://doi.org/https://doi.org/10.1016/j.jesp.2004.10.006

Friedrich, C. K., Felder, V., Lahiri, A., & Eulitz, C. (2013). Activation of words with phonological overlap. *Frontiers in Psychology*, *4*, 556. https://doi.org/10.3389/fpsyg.2013.00556

Friedrich, C. K., Lahiri, A., & Eulitz, C. (2008). Neurophysiological evidence for underspecified lexical representations: asymmetries with word initial variations. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1545–1559. https://doi.org/10.1037/a0012481

Frumkin, L. A., & Stone, A. (2020). Not all eyewitnesses are equal: Accent status, race and age interact to influence evaluations of testimony. *Journal of Ethnicity in Criminal Justice*, *18*(2), 123-145. https://doi.org/10.1080/15377938.2020.1727806

Fuertes, J., Gottdiener, W., Martin, H., Gilbert, T., & Giles, H. (2012). A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology*, *42*(1), 120–133. https://doi.org/10.1002/ejsp.862

Fuertes, J. N., Potere, J. C., & Ramirez, K. Y. (2002). Effects of speech accents on interpersonal evaluations: Implications for counseling practice and research. *Cultural diversity and ethnic minority psychology*, *8*(4), 346–56. https://doi.org/10.1037/1099-9809.8.4.347

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin, and Review*, *13*, 361-77. https://doi.org/10.3758/BF03193857

Gaskell, G., & Marslen-Wilson, W. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language*, *44*(3), 325–349. https://doi.org/10.1006/jmla.2000.2741

Gilbert, R. C., Chandrasekaran, B., & Smiljanic, R. (2014). Recognition memory in noise for speech of varying intelligibility. *The Journal of the Acoustical Society of America*, *135*, 389—399. https://doi.org/10.1121/1.4838975

Giles, H. (1970). Evaluative reactions to accents. *Educational Review*, *22*(3), 211–227. https://doi.org/10.1080/0013191700220301

Giles, H., & Trudgill, P. (1983). Sociolinguistics and linguistic value judgements. *On Dialect. Oxford: Basil Blackwell*.

Gluszek, A., & Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review*, *14*(2), 214-237. https://doi.org/10.1177/1088868309359288

Goldin, A., & Weinstein, B. E. (2020). How do medical masks degrade speech perception? *Hearing Review*, *27*(5), 8-9.

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of experimental psychology: Learning, memory, and cognition*, *22*(5), 1166. https://doi.org/10.1037/0278-7393.22.5.1166

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279. https://doi.org/10.1037/0033-295X.105.2.251

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(1), 152. https://doi.org/10.1037/0278-7393.17.1.152

Goslin, J., Duffy, H., & Floccia, C. (2012). An erp investigation of regional and foreign accent processing. *Brain Language*, *122*(2), 92–102. https://doi.org/10.1016/j.bandl.2012.04.017

Gosselin, P. A., & Gagné, J. P. (2011). Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. *International Journal of Audiology*, *50*(11), 786–792. https://doi.org/10.3109/14992027.2011.599870

Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America*, *108*(3), 1197–1208. https://doi.org/10.1121/1.1288668

Grant, K. W., Walden, B. e., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *The Journal of the Acoustical Society of America*, *103*(5), 2677–2690. https://doi.org/10.1121/1.422788

Greifeneder, R., Alt, A., Bottenberg, K., Seele, T., Zelt, S., & Wagener, D. (2010). On writing legibly: Processing fluency systematically biases evaluations of handwritten material. *Social Psychological and Personality Science*, *1*(3), 230–237. https://doi.org/10.1177/1948550610368434

Grohe, A.-K., & Weber, A. (2018). Memory advantage for produced words and familiar native accents. *Journal of Cognitive Psychology*, *30*(5-6), 570-587. https://doi.org/10.1080/20445911.2018.1499659

Halle, M. (2003). Speculations about the representations of words in memory. In *From memory to speech and back: Papers on phonetics and phonology 1954 - 2002* (pp. 122–136). Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110871258.122

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive science*, *28*(1), 105–115. https://doi.org/10.1207/s15516709cog2801_5

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*(1), 43–61. https://doi.org/10.1016/S0749-596X(03)00022-6

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: the effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887. https://doi.org/10.1162/jocn_a_00103

Hanulíková, A., & Weber, A. (2012). Sink positive: Linguistic experience with th substitutions influences nonnative word recognition. *Attention, Perception, and Psychophysics*, *74*(3), 613–629. https://doi.org/10.3758/s13414-011-0259-7

Hanzlíková, D., & Skarnitzl, R. (2017). Credibility of native and non-native speakers of english revisited: Do non-native listeners feel the same? *Research in Language*, *15*(3), 285–298. https://doi.org/10.1515/rela-2017-0016

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, *16*(1), 107–112. https://doi.org/10.1016/S0022-5371(77)80012-1

Hawkins, S. A., & Hoch, S. J. (1992). Low-involvement learning: Memory without evaluation. *Journal of Consumer Research,*, *19*(2), 212–225.

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, *48*(4). https://doi.org/10.1515/ling.2010.027

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, *23*(3), 351–379. https://doi.org/10.1515/TLR.2006.014

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of phonetics*, *34*(4), 458–484. https://doi.org/10.1016/j.wocn.2005.10.001

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., & Chunge, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America*, *119*(3), 1740–1751. https://doi.org/10.1121/1.2166611

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by japanese learners of english. *Speech Communication*, *47*(3), 360–378. https://doi.org/10.1016/j.specom.2005.04.007

Hillenbrand, J. M., & Clark, M. (2009). The role of f0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, and Psychophysics*, *71*(5), 1150–1166. https://doi.org/10.3758/APP.71.5.1150

Hillenbrand, J. M., Getty, L., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of american english vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099–3111. https://doi.org/10.1121/1.409456

Hintzman, D. L. (1986). "schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*(4), 411–428. https://doi.org/10.1037/0033-295X.93.4.411

Hirschberg, J. B., & Rosenberg, A. (2005). *Acoustic/prosodic and lexical correlates of charismatic speech.*

Hughes, S. M., & Miller, N. E. (2016). What sounds beautiful looks beautiful stereotype: The matching of attractiveness of voices and faces. *Journal of Social and Personal Relationships*, *33*(7), 984–996. https://doi.org/10.1177/0265407515612445

Jaeger, T. F. (2008). Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Jalbert, M., Newman, E., & Schwarz, N. (2020). Only half of what i'll tell you is true: Expecting to encounter falsehoods reduces illusory truth. *Journal of Applied Research in Memory and Cognition*, *9*(4), 602–613. https://doi.org/10.1016/j.jarmac.2020.08.010

Janse, E., & Adank, P. (2012). Predicting foreign-accent adaptation in older adults. *Quarterly Journal of Experimental Psychology*, *65*(8), 1563-1585. https://doi.org/10.1080/17470218.2012.658822

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. https://doi.org/10.1080/01690965.2011.620335

Jesse, A., & Massaro, D. W. (2010). The temporal distribution of information in audiovisual spoken-word identification. *Attention, Perception, and Psychophysics*, *72*(1), 209–225. https://doi.org/10.3758/APP.72.1.209

Jesse, A., Vrignaud, N., Cohen, M. M., & Massaro, D. W. (2000/2001). The processing of information from multiple sources in simultaneous interpreting. *Interpreting*, *5*, 95-115.

Jesse, A., Vrignaud, N., Cohen, M. M., & Massaro, D. W. (2001-2002). The processing of information from multiple sources in simultaneous interpreting. *Interpreting*, *5*(2), 95–115. https://doi.org/10.1075/intp.5.2.04jes

Jiang, X., & Pell, M. D. (2017). The sound of confidence and doubt. *Speech Communication*, *88*, 106–126.

Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. *The Journal of Acoustical Society of America*, *94*(2), 701–14. https://doi.org/10.1121/1.406887

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., & Vukovic, J. (2010). A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour*, *79*(1), 57–62. https://doi.org/10.1016/j.anbehav.2009.10.003

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149. https://doi.org/10.1037/0033-295X.99.1.122

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.

Kaiser, A. R., Kirk, K. I., Lachs, L., & Pisoni, D. B. (2003). Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *Journal of Speech Language, and Hearing Research*, *46*(2), 390–404. https://doi.org/10.1044/1092-4388(2003/032)

Kawase, S., Hannah, B., & Wang, Y. (2014). The influence of visual speech information on the intelligibility of english consonants produced by non-native speakers. *The Journal of the Acoustical Society of America*, *136*(3), 1352–1362. https://doi.org/10.1121/1.4892770

Keerstock, S., & Smiljanic, R. (2018). Effects of intelligibility on within- and cross-modal sentence recognition memory for native and non-native listeners. *The Journal of the Acoustical Society of America*, *144*(5), 2871–2881. https://doi.org/10.1121/1.5078589

Keerstock, S., & Smiljanic, R. (2019). Clear speech improves listeners' recall. *The Journal of the Acoustical Society of America*, *146*(6), 4604-4610. https://doi.org/10.1121/1.5141372

Keidser, G., Best, V., Freeston, K., & Boyce, A. (2015). Cognitive spare capacity: evaluation data and its association with comprehension of dynamic conversations. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00597

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of speech and hearing research*, *19*(3), 421–47. https://doi.org/10.1044/jshr.1903.421

Kinzler, K. D., & DeJesus, J. M. (2013). Northern= smart and southern= nice: The development of accent attitudes in the united states. *Quarterly Journal of Experimental Psychology*, *66*(6), 1146–1158. https://doi.org/10.1080/17470218.2012.731695

Kittredge, A., Davis, L., & Blumstein, S. E. (2006). Effects of nonlinguistic auditory variations on lexical processing in broca's aphasics. *Brain and Language*, *97*(1), 25–40. https://doi.org/10.1016/j.bandl.2005.07.012

Klatt, D. H. (1979). Speech perception: A model of acousti-phonetic analysis and lexical access. *Journal of Phonetics*, *7*(3), 279–312.

Klofstadt, C. A., Anderson, R. C., & S., P. (2012). Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings. Biological sciences / The Royal Society*, *279*, 2698–704. https://doi.org/10.1098/rspb.2012.0311

Koops, C., Gentry, E., & Pantos, A. (2008). The effect of perceived speaker age on the perception of pin and pen vowels in houston, texas. *University of Pennsylvania Working Papers in Linguistics*, *14*(2), 93–101.

Kramer, E. (1963). Judgment of personal characteristics and emotions from nonverbal properties of speech. *Psychological Bulleting*, *60*(4), 408–420. https://doi.org/10.1037/h0044890

Kreiman, J., Gerrat, B. R., K., P., & Berke, G. S. (1992). Individual differences in voice quality perception. *Journal of speech and hearing research*, *35*(3), 512–20. https://doi.org/10.1044/jshr.3503.512

Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Wiley-Blackwell Publication. https://doi.org/10.1002/9781444395068

Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., Jr., Cute, S. L., Humes, L. E., Dubno, J. R., & Ecker, M. A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology*, *50*(1), 23–34.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the acoustical society of America*, *29*(1), 98–104. https://doi.org/10.1121/1.1908694

Landis, T., Buttet, J., Assal, G., & Graves, R. (1982). Dissociation of ear preference in monaural word and voice recognition. *Neuropsychologia*, *20*(4), 501–504. https://doi.org/10.1016/0028-3932(82)90049-5

Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology*, *21*(4), R143–R145. https://doi.org/https://doi.org/10.1016/j.cub.2010.12.033

Lee, C. Y., & Zhang, Y. (2015). Processing speaker variability in repetition and semantic/associative priming. *Journal of Psycholinguistic Research*, *44*(44), 237–250. https://doi.org/10.1007/s10936-014-9307-5

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*(3), 1455–1468. https://doi.org/10.1121/1.426686

Lenth, R. (2018). *emmeans: Estimated marginal means, aka least-squares means. r package version 1.2.4.*

Lev-Ari, S. (2015). Comprehending non-native speakers: Theory and evidence for adjustment in manner of processing. *Frontiers in psychology*, *5*, 1546. https://doi.org/10.3389/fpsyg.2014.01546

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? the influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096. https://doi.org/10.1016/j.jesp.2010.05.025

Levi, S., & Pisoni, D. B. (2007). Indexical and linguistic channels in speech perception: Some effects of voiceovers on advertising outcomes. In T. M. Lowrey (Ed.), *Psycholinguistic phenomena in marketing communications* (pp. 203–219). Lawrence Erlbaum Associates.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36. https://doi.org/10.1016/0010-0277(85)90021-6

Lick, D. J., & Johnson, K. L. (2015). The interpersonal consequences of processing ease: Fluency as a metacognitive foundation for prejudice. *Current Directions in Psychological Science*, *24*(2), 143–148. https://doi.org/10.1177/0963721414558116

Lindemann, S. (2003). Koreans, chinese or indians? attitudes and ideologies about non-native english speakers in the united states. *Journal of Sociolinguistics*, *7*(3), 348–364. https://doi.org/10.1111/1467-9481.00228

Llamas, C., Harrison, P., Donnelly, D., & Watt, D. (2009). Effects of different types of face coverings on speech acoustics and intelligibility. *York Papers in Linguistics Series 2*(9), 80–104.

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception and Psychophysics*, *39*(3), 155–158. https://doi.org/10.3758/BF03212485

Luce, P. A., & Lyons, E. A. (1998). Specificity of memory representations for spoken words. *Memory and Cognition*, *26*(4), 708–715. https://doi.org/10.3758/BF03211391

Luce, P. A., McLennan, C. T., & Charles-Luce, J. (2003). Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming. In J. Bowers & C. Marsolek (Eds.), *Rethinking implicit memory* (pp. 197–214). Oxford: Oxford University Press.

Lüttke, S. C. (2018). *What you see is what you hear. visual influences on auditory speech perception* (phdthesis). Radboud University.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perseption and Psychophysics*, *24*(3), 253–257. https://doi.org/10.3758/bf03206096

Mack, S., & Munson, B. (2012). The influence of /s/ quality on ratings of men's sexual orientation: Explicit and implicit measures of the 'gay lisp' stereotype. *Journal of Phonetics*, *40*(1), 198–212. https://doi.org/10.1016/j.wocn.2011.10.002

Maclagan, M. A., & Gordon, E. (2000). The near/square merger in new zealand english. *Asia Pacific Journal of Speech, Language and Hearing*, *5*(3), 201-207. https://doi.org/10.1179/136132800805576951

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, *21*(2), 131–141. https://doi.org/10.3109/03005368709077786

Magee, M., Lewis, C., Noffs, G., Reece, H., Chan, J. C. S., Zaga, C. J., ... Vogel, A. P. (2020). Effects of face masks on acoustic analysis and speech perception: Implications for peri-pandemic protocols. *The Journal of the Acoustical Society of America.*, *148*(6), 3562. https://doi.org/10.1121/10.0002873

Marinis, T. (2018). Cross-modal priming in bilingual sentence processing. *Bilingualism: Language and Cognition*, *21*(3), 456–461. https://doi.org/10.1017/S1366728917000761

Marsh, E. J., & Umanath, S. (2014). Knowledge neglect: Failures to notice contraditions with stored knowledge. In D. N. Rapp & J. L. G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 161–181). Boston Review.

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, *101*(4), 653–675. https://doi.org/10.1037/0033-295X.101.4.653

Masapollo, M., Polka, L., Molnar, M., & Ménard, L. (2017). Directional asymmetries reveal a universal bias in adult vowel perception. *The Journal of the Acoustical Society of America*, *141*(4), 2857. https://doi.org/10.1121/1.4981006

Masapollo, M., Polka, L., & Ménard, L. (2015). Asymmetries in vowel perception: Effects of formant convergence and category "goodness". *Journal of the Acoustical Society of America*, *137*(4), 2385. https://doi.org/10.1121/1.4920678

Masapollo, M., Polka, L., Ménard, L., Franklin, L., Tiede, M., & Morgan, J. (2018). Asymmetries in unimodal visual vowel perception: The roles of oral-facial kinematics, orientation, and configuration. *Journal of experimental psychology. Human perception and performance*, *44*(7), 1103–1118. https://doi.org/10.1037/xhp0000518

Massaro, D. W. (1987). Categorical partition: A fuzzy-logical model of categorization behavior. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 254–283). Cambridge University Press.

Massaro, D. W. (1989). Testing between the trace model ad the fuzzy logical model of speech perceptio. *Cogitive Psychology*, *21*(3), 398–421. https://doi.org/10.1016/0010-0285(89)90014-5

Massaro, D. W. (1998). *Perceivig talkig faces: From speech perceptio to a behavioral principle*. The MIT Press.

Massaro, D. W., & Chen, T. H. (2008). The motor theory of speech perception revisted. *Psychonomic Bulletin, and Review*, *15*(2), 453–462. https://doi.org/10.3758/pbr.15.2.453

Massaro, D. W., & Cohen, M. M. (1993). Perceiving asynchronous bimodal speech in consonant-vowel and vowel syllables. *Speech Communication*, *13*(1), 127–134. https://doi.org/10.1016/0167-6393(93)90064-R

Massaro, D. W., & Jesse, A. (2007). Audiovisual speech perception and word recognition. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 19–35). Oxford: Oxford University Press.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7-8), 953–978. https://doi.org/10.1080/01690965.2012.705006

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*(3), 543–562. https://doi.org/10.1080/03640210802035357

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'hello'? personality impressions from brief novel voices. *PLoS One*, *9*, e90779. https://doi.org/10.1371/journal.pone.0090779

McClelland, J. L., & Elman, J. L. (1986). The trace model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *Quarterly Journal of Experimental Psychology*, *58*(1). https://doi.org/10.1080/02724980443000151

McGlone, M. S., & Tofighbakhsh, J. (2000). Birds of a feather flock conjointly (?): Rhyme as reason in aphorisms. *Psychological Science*, *11*(5), 424–428. https://doi.org/10.1111/1467-9280.00282

McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5), 745–748. https://doi.org/10.1038/264746a0

McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Learning, memory, and cognition*, *31*(2), 306–321. https://doi.org/10.1037/0278-7393.31.2.306

McQueen, J. M. (2005). Speech perception. In K. Lamberts & R. Goldstone (Eds.), *Handbook of cognition* (pp. 255–275). London: Sage Publications.

McQueen, J. M., Dahan, D., & Cutler, A. (2003). Continuity and gradedness in speech processing. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 39–78). Berlin, Germany: Mouton de Gruyter.

McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimantal Psychology: Learning, Memory, Cognition*, *20*(3), 621. https://doi.org/10.1037//0278-7393.20.3.621

Mendel, L. L., Gardino, J. A., & Atcherson, S. R. (2008). Speech understanding using surgical masks: A problem in health care? *Journal of the American Academy of Audiology*, *19*(9), 686–95. https://doi.org/10.3766/jaaa.19.9.4

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*(2), 201–213. https://doi.org/10.1016/S0749-596X(03)00028-7

Miller, G. R., & Hewgill, M. A. (1964). The effect of variations in nonfluency on audience ratings of source credibility. *Quarterly Journal of Speech*, *50*(1), 36-44. https://doi.org/10.1080/00335636409382644

Miller, N., Maruyama, G., Beaber, R. J., & Valone, K. (1976). Speed of speech and persuasion. *Journal of Personality and Social Psychology*, *34*(4), 615–624. https://doi.org/10.1037/0022-3514.34.4.615

Mirman, D. (2017). *Growth curve analysis and visualization using r*. London: Taylor Francis.

Mishra, S., Lunner, T., Stenfelt, S., Rönnberg, J., & Rudner, M. (2013a). Seeing the talker's face supports executive processing of speech in steady state noise. *Frontiers in Systems Neuroscience*, *7*(96). https://doi.org/10.3389/fnsys.2013.00096

Mishra, S., Lunner, T., Stenfelt, S., Rönnberg, J., & Rudner, M. (2013b). Visual information can hinder working memory processing of speech. *Journal of Speech, Language, and Hearing Research*, *56*(3), 1120–1132. https://doi.org/10.1044/1092-4388(2012/12-0033)

Mitchell, a., Gottfried, J., Barthel, M., & Sumida, N. (2018). *Distinguishing between factual and opinion statements in the news.* Retrieved from https://www.pewresearch.org/journalism/2018/06/18/distinguishing-between-factual-and-opinion-statements-in-the-news/

Moult, M. (2011). Cued recall. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of clinical neuropsychology* (pp. 751–752). New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-79948-3_1116

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics*, *47*(4), 379–390. https://doi.org/10.3758/BF03210878

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*(3), 289–306. https://doi.org/10.1177/002383099503800305

Munro, M. J., Derwing, T. M., & Satō, K. (2006). Salient accents, covert attitudes: Consciousness-raising for pre-service second language teachers..

Murry, T., & Singh, S. (1980). Multidimensional analysis of male and female voices. *The Journal of the Acoustical Society of America*, *68*(5), 1294-1300. https://doi.org/10.1121/1.385122

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research*, *71*, 4–12. https://doi.org/10.1007/s00426-005-0031-5

Newman, E. J., & Schwarz, N. (2018). Good sound, good research: How audio quality influences perceptions of the research and researcher. *Science Communication*, *40*(2), 246–257. https://doi.org/10.1177/1075547018759345

Niebuhr, O., Brem, A., Novák-Tót, E., & Voße, J. (2016). *Charisma in business speeches – a contrastive acoustic-prosodic analysis of steve jobs and mark zuckerberg.* MIT, Boston, USA.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of language and social psychology*, *18*(1), 62–85. https://doi.org/10.1177/0261927X99018001005

Nisbett, S. L., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*(4), 250–256. https://doi.org/10.1037/0022-3514.35.4.250

Nissen, S. L., Dromey, C., & Wheeler, C. (2007). First and second language tongue movements in Spanish and Korean bilingual speakers. *Phonetica*, *64*(4), 201–216. https://doi.org/10.1159/000121373

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189–234. https://doi.org/10.1016/0010-0277(94)90043-4

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics*, *60*(3), 355–376. https://doi.org/10.3758/bf03206860

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42–46. https://doi.org/10.1111/j.1467-9280.1994.tb00612.x

Oldenburger Satztest: Handbuch und Hintergrundwissen. (2000). Oldenburg, Germany: Hörtech GmbH.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237–241. https://doi.org/10.1016/j.tics.2008.02.014

Owens, E., & Blazek, B. (1985). Visemes observed by hearing-impaired and normal-hearing adult viewers. *Journal of Speech and Hearing Research*, *28*(3), 381–393. https://doi.org/10.1044/jshr.2803.381

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Learning, Memory*, *19*(2), 309–328. https://doi.org/10.1037/0278-7393.19.2.309

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, *39*(2), 204–214. https://doi.org/10.1097/AUD.0000000000000494

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*, 175–184. https://doi.org/10.1121/1.1906875

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, *19*, 123–205. https://doi.org/10.1016/S0065-2601(08)60214-2

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., ... Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (fuel). *Ear and Hearing*, *37*, 5S–27S. https://doi.org/10.1097/AUD.0000000000000312

Podlipsky, V., Simackova, S., & Petráž, D. (2016). Is there an interlanguage speech credibility benefit? *Topics in Linguistics*, *17*(1). https://doi.org/10.1515/topling-2016-0003

Polka, L., & Bohn, O.-S. (2011). Natural referent vowel (nrv) framework: An emerging view of early phonetic development. *Journal of phonetics*, *39*(4), 467–478. https://doi.org/10.1016/j.wocn.2010.08.007

Polka, L., Molnar, M., Zhao, T. C., & Masapollo, M. (2021). Neurophysiological correlates of asymmetries in vowel perception: An english-french cross-linguistic event-related potential study. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.607148

Prieur, J., Barbu, S., Blois-Heulin, C., & Lemasson, A. (2020). The origins of gestures and language: history, current advances and proposed theories. *Biological Reviews*, *95*(3), 531-554. https://doi.org/https://doi.org/10.1111/brv.12576

Ptacek, P. H., & Sander, E. K. (1966). Age recognition from voice. *Journal of speech and hearing research*, *9*(2), 273–277. https://doi.org/10.1044/jshr.0902.273

Pörschmann, C., Lübeck, T., & Arend, J. M. (2020). Impact of face masks on voice radiation. *The Journal of the Acoustical Society of America*, *148*(6), 3663-3670. https://doi.org/10.1121/10.0002853

R Core Team. (2018). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for statistical computing. Retrieved from https://www.r-project.org (versions 3.5.0 to 4.0.5)

Rabbitt, P. (1968). Channel-capacity, intelligibility and immediate memory. *The Quarterly Journal of Experimental Psychology*, *20*(3), 241–248. https://doi.org/10.1080/14640746808400158

Rabbitt, P. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with iq. *Acta Oto-Laryngologica*, *111*(sup476), 167–176. https://doi.org/10.3109/00016489109127274

Randazzo, M., Koenig, L. L., & Priefer, R. (2020). The effect of face masks on the intelligibility of unpredictable sentences. *Proceedings of Meetings on Acoustics*, *42*(1), 032001. https://doi.org/10.1121/2.0001374

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, *8*(3), 338 –342. https://doi.org/10.1006/ccog.1999.0386

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, *8*(4), 364-382. https://doi.org/10.1207/s15327957pspr0804_3

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 97–113). Erlbaum, Hillsdale, NJ.

Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of experimental psychology. Human perception and performance*, *23*(3), 651–66. https://doi.org/10.1037//0096-1523.23.3.651

Repp, B. H., & Crowder, R. G. (1990). Stimulus order effects in vowel discrimination. *The Journal of the Acoustical Society of America*, *88*(5), 2080–2090. https://doi.org/10.1121/1.400105

Rezlescu, C., Penton, T., Walsh, V., Tsujimura, H., Scott, S. K., & Banissy, M. J. (2015). Dominant voices and attractive faces: The contribution of visual and auditory information to integrated person impressions. *Journal of Nonverbal Behavior*, *39*(4), 355–370. https://doi.org/10.1007/s10919-015-0214-8

Riedinger, M., Nagels, A., Werth, A., & Scharinger, M. (2021). Asymmetries in accessing vowel representations are driven by phonological and acoustic properties: Neural and behavioral evidence from natural german minimal pairs. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.612345

Romero-Rivas, C., Thorley, C., Skelton, K., & Costa, A. (2019). Foreign accents reduce false recognition rates in the drm paradigm. *Journal of Cognitive Psychology*, *31*(5-6), 507–521. https://doi.org/10.1080/20445911.2019.1634576

Rommers, J., & Federmeier, K. D. (2018). Predictability's aftermath: Downstream consequences of word predictability as revealed by repetition effects. *Cortex*, *101*, 16–30. https://doi.org/doi.org/10.1016/j.cortex.2017.12.018

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 environment. *Cerebral Cortex*, *17*(5), 1147–1153. https://doi.org/10.1093/cercor/bhl024

Roster, C. A., Lucianetti, L., & Albaum, G. S. (2015). Exploring slider vs. categorical response formats in web-based surveys. *Journal of Research Practice*, *11*(1), 1–16.

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., ... Rudner, M. (2013). The ease of language understanding (elu) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*, 31. https://doi.org/10.3389/fnsys.2013.00031

Schacter, D. L., & Church, B. A. (1992). Auditory priming: implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 915. https://doi.org/10.1037/0278-7393.18.5.915

Schroeder, L. D., Sjoquist, D. L., & Stephan, P. E. (2017). *Understanding regression analysis: An introductory guide* (Second Edition ed.; B. Entwisle, Ed.). Sage Publications, Inc. https://doi.org/10.4135/9781506361628

Schwartz, J., Abry, C., Boe, L.-J., Ménard, L., & Vallée, N. (2005). Asymmetries in vowel perception, in asymmetries in vowel perception, in the the context of the dispersion-focalisation theory. *Speech Communication*, *45*(4), 425–434. https://doi.org/10.1016/j.specom.2004.12.001

Schwarz, J., Li, K. K., Sim, J. H., Zhang, Y., Buchanan-Worster, E., Post, B., ... McDougall, K. (2022). Semantic cues modulate children's and adults' processing of audio-visual face mask speech. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.879156

Schwarz, N. (2004). Meta-cognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, *14*(4), 332–348. https://doi.org/10.1207/s15327663jcp1404_2

Schwarz, N. (2015). Metacognition. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *Apa handbook of personality and social psychology, volume 1: Attitudes and social cognition.* (pp. 203–229). American Psychological Association. https://doi.org/10.1037/14341-006

Schwarz, N., & Jalbert, M. (2020). When (fake) news feels true: Intuitions of truth and the acceptance and correction of misinformation. In *The psychology of fake news: Accepting, sharing, and correcting misinformation* (chap. When (fake) news feels true: Intuitions of truth and the acceptance and correction of misinformation). London, UK: Routledge.

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in chinese subjects. *Perception and Psychophysics*, *59*(1), 73–80. https://doi.org/10.3758/BF03206849

Sekiyama, K., & Tohkura, Y. I. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, *21*(4), 427–444. https://doi.org/10.1016/S0095-4470(19)30229-3

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., ... Gurven, M. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B: Biological Sciences*, *277*, 3509–3518. https://doi.org/10.1098/rspb.2010.0769

Service, E., Simola, M., Metsänheimo, O., & Maury, S. (2010). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology*, *14*(3), 383–408. https://doi.org/10.1080/09541440143000140

Smiljanic, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass*, *3*(1). https://doi.org/10.1111/j.1749-818X.2008.00112.x

Smiljanic, R., Keerstock, S., Meerman, K., & Ransom, S. M. (2021). Face masks and speaking style affect audio-visual word recognition and memory of native and non-native speech. *The Journal of Acoustical Society of America*, *149*(6), 4013–4023. https://doi.org/10.1121/10.0005191

Smith, B., Sugarman, M., & Long, S. (1983). Experimental manipulation of speaking rate for studying temporal variability in children's speech. *The Journal of the Acoustical Society of America*, *74*(3), 744-–749. https://doi.org/10.1121/1.389860

Smith, B. L., Brown, B. L., Strong, W. J., & Rencher, A. C. (1975). Effects of speech rate on personality perception. *Language and Speech*, *18*(2), 145–52. https://doi.org/10.1177/002383097501800203

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(Suppl1)*, 62S–68S. https://doi.org/10.1097/AUD.0000000000000322

Song, H., & Schwarz, N. (2009). If it's difficult to pronounce, it must be risky: Fluency, familiarity, and risk perception. *Psychological Science*, *20*(2), 135-138. https://doi.org/10.1111/j.1467-9280.2009.02267.x

Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. *Cognition*, *92*(3), B13–B23. https://doi.org/https://doi.org/10.1016/j.cognition.2003.10.005

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, *45*(3), 412–432. https://doi.org/10.1006/jmla.2000.2783

Souza, A. L., & Markman, A. B. (2013). Foreign accent does not influence cognitive judgments. In *Cogsci* (Vol. 35).

Steinlen, A. K. (2005). *The influence of consonants on native and non-native vowel production: A cross-linguistic study*. Tübingen: Gunter Narr.

Stevens, A. A. (2004). Dissociating the cortical basis of memory for voices, words and tones. *Cognitive Brain Research*, *18*(2), 162–171.

Stocker, L. (2017). The impact of foreign accent on credibility: An analysis of cognitive statement ratings in a swiss context. *Journal of Psycholinguistic Research*, *46*, 617–628. https://doi.org/10.1007/s10936-016-9455-x

Stoel-Gammon, C., & Menn, L. (2005). Phonological development: Learning sounds and sound patterns. In J. B. Gleason & N. B. Ratner (Eds.), *The development of language* (pp. 52–85). Pearson.

Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology*, *18*(1), 86–100. https://doi.org/10.1177/0261927X99018001006

Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., & Nishi, K. (2007). Acoustic variability within and across german, french, and american english vowels: phonetic context effects. *The Journal of the Acoustical Society of America*, *122*(2), 1111–1129. https://doi.org/10.1121/1.2749716

Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research*, *11*(2), 139–169. https://doi.org/10.1111/J.1468-2958.1984.TB00043.X

Street, R. L., Jr., & Brady, R. M. (1982). Speech rate acceptance ranges as a function of evaluative domain, listener speech rate, and communication context. *Communication Monographs*, *49*(4), 290–308.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*(2), 212. https://doi.org/10.1121/1.1907309

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, *335*(1273), 71-78. https://doi.org/10.1098/rstb.1992.0009

Surawski, M. K., & Ossoff, E. P. (2006). The effects of physical and vocal attractiveness on impression formation of politicians. *Current Psychology*, *25*, 15-27. https://doi.org/10.1007/S12144-006-1013-5

Swinney, D. A. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*(6), 645–659. https://doi.org/https://doi.org/10.1016/S0022-5371(79)90355-4

Tanenhaus, M. K., & Trueswell, J. C. (2005). Eye movements as a tool for bridging the language-as-product and language-as-action traditions. *Approaches to studying world-situated language use: Bridging the language-as-product and language-action traditions*, 3–37.

Tesink, C. M. J. Y., Petersson, K. M., Van Berkum, J. J. A., Van den Brink, D., Buitelaar, J. K., & Hagoort, P. (2008). Unification of speaker and meaning in language comprehension: An fmri study. *Journal of Cognitive Neuroscience*, *21*(11), 2085–2099. https://doi.org/10.1162/jocn.2008.21161

Tigue, C. C., Borak, D. J., O'Connor, J. J. M., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior*, *33*(3), 210 - 216. https://doi.org/https://doi.org/10.1016/j.evolhumbehav.2011.09.004

Tingley, B. M., & Allen, G. D. (1975). Development of speech timing control in children. *Child Development*, *46*(1), 186–194. https://doi.org/10.2307/1128847

Trude, A. M., Tremblay, A., & Brown-Schmidt, S. (2013). Limitations on adaptation to foreign accents. *Journal of memory and language. - Amsterdam ([u.a.]*, *69*(3), 349–367. https://doi.org/10.1016/j.jml.2013.05.002

Truong, T. L., Beck, S. D., & Weber, A. (2021). The impact of face masks on the recall of spoken sentences. *The Journal of the Acoustical Society of America*, *149*(1), 142-144. https://doi.org/10.1121/10.0002951

Truong, T. L., & Weber, A. (2021). Intelligibility and recall of sentences spoken by adult and child talkers wearing face masks. *The Journal of the Acoustical Society of America*, *150*(3), 1674-1681. https://doi.org/10.1121/10.0006098

Tsantani, M. S., Belin, P., Paterson, H. M., & McAleer, P. (2016). Low vocal pitch preference drives first impressions irrespective of context in male voices but not in female voices. *Perception*, *45*(8), 946-963. https://doi.org/10.1177/0301006616643675

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 381–391. https://doi.org/10.1016/S0022-5371(66)80048-8

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. https://doi.org/10.1037/pag0000094

Unkelbach, C. (2006). The learned interpretation of cognitive fluency. *Psychological Science*, *17*(4), 339–345. https://doi.org/10.1111/j.1467-9280.2006.01708.x

Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of experimental psychology. Learning, memory, and cognition*, *33*(1), 219-30. https://doi.org/10.1037/0278-7393.33.1.219

Unkelbach, C., & Greifeneder, R. (2018). Experiential fluency and declarative advice jointly inform judgments of truth. *Journal of Experimental Social Psychology*, *79*, 78–86. https://doi.org/10.1016/j.jesp.2018.06.010

Unkelbach, C., & Stahl, C. (2009). A multinomial modeling approach to dissociate different components of the truth effect. *Consciousness and Cognition*, *18*(1), 22–38. https://doi.org/10.1016/j.concog.2008.09.006

Van Berkum, J. J. A., Van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of cognitive neuroscience*, *20*(4), 580–91. https://doi.org/10.1162/jocn.2008.20054

Van Den Brink, D., Van Berkum, J. J. A., Bastiaansen, M. C. M., Tesink, C. M. J. Y., Kos, M., Buitelaar, J. K., & Hagoort, P. (2010). Empathy matters: Erp evidence for inter-individual differences in social language processing. *Social Cognitive and Affective Neuroscience*, *7*(2), 173–183. https://doi.org/10.1093/scan/nsq094

Van den Noort, M. W. M. L., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist*, *11*(4), 289–296. https://doi.org/10.1027/1016-9040.11.4.289

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, *40*(3), 191–210. https://doi.org/10.1017/S0261444807004338

Van der Zande, P. (2013). *Hearing and seeing speech: Perceptual adjustments in auditory-visual speech processing* (phdt hesis). Radboud University Nijmegen.

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, *8*, 577. https://doi.org/10.3389/fnhum.2014.00577

Van Lancker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia: A dissociation between familiar and unfamiliar voices. *Cortex*, *24*(2), 195–209. https://doi.org/10.1016/S0010-9452(88)80029-7

Van Son, N., Huiskamp, T. M. I., Bosman, A. J., & Smoorenburg, G. F. (1994). Viseme classifications of dutch consonants and vowels. *The Journal of the Acoustical Society of America*, *96*(3), 1341–1355. https://doi.org/10.1121/1.411324

Vejnovic, D., Milin, P., & Zdravković, S. (2010). Effects of proficiency and age of language acquisition on working memory performance in bilinguals. *Psihologija*, *43*(3), 219–232. https://doi.org/10.2298/PSI1003219V

Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1510–45. https://doi.org/10.1044/1092-4388(2007/104)

Wade, T., Jongman, A., & Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica*, *64*(2-3), 122–144. https://doi.org/10.1159/000107913

Wagner, A. E., Toffanin, P., & Başkent, D. (2016). The timing and effort of lexical access in natural and degraded speech. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00398

Wang, Y., Behne, D. M., & Jiang, H. (2008). Linguistic experience and audio-visual perception of non-native fricatives. *The Journal of the Acoustical Society of America*, *124*(3), 1716–1726. https://doi.org/10.1121/1.2956483

Warren, P., Hay, J., & Thomas, B. (2007). Chapter ?? the loci of sound change effects in recognition and perception. *Laboratory Phonology 9*, *9*.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0

Weber, A., & Scharenborg, O. (2012). Models of spoken-word recognition. *WIREs Cognitive Science*, *3*(3), 387–401. https://doi.org/10.1002/wcs.1178

Weick, M., & Guinote, A. (2008). When subjective experiences matter: power increases reliance on the ease of retrieval. *Journal of Personality and Social Psychology*, *94*(6), 956–970. https://doi.org/10.1037/0022-3514.94.6.956

Wells. (1977). Peter ladefoged, a course in phonetics. *Journal of Linguistics*, *13*(2), 329–334. https://doi.org/10.1017/S002222670000551X

Wetzel, C. G., Wilson, T. D., & Kort, J. (1981). The halo effect revisited: Forewarned is not forearmed. *Journal of Experimental Social Psychology*, *17*(4), 427–439. https://doi.org/https://doi.org/10.1016/0022-1031(81)90049-4

Wetzel, M., Zufferey, S., & Gygax, P. (2021). Do non-native and unfamiliar accents sound less credible? an examination of the processing fluency hypothesis. *Journal of articles in support of the null hypothesis*, *17*(2).

Whittlesea, B. W. A., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language*, *29*(6), 716 - 732. https://doi.org/https://doi.org/10.1016/0749-596X(90)90045-2

Wiese, R. (2000). *The phonology of german*. Oxford: Oxford University Press.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Winter, B. (2019). *Statistics for linguists: An introduction using r*. Routledge. https://doi.org/10.4324/9781315165547

Winters, S. J., Levi, S. V., & Pisoni, D. B. (2008). Identification and discrimination of bilingual talkers across languages. *The Journal of the Acoustical Society of America*, *123*(6), 4524–4538. https://doi.org/10.1121/1.2913046

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, and Psychophysics*, *75*(3), 537–556. https://doi.org/10.3758/s13414-012-0404-y

Wittemann, M. J., Weber, A., & McQueen, J. M. (2014). Tolerance for inconsistency in foreign-accented speech. *Psychonomic Bulletin, and Review*, *21*, 512–519. https://doi.org/10.3758/s13423-013-0519-8

Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, and Psychophysics*, *79*, 2064–2072. https://doi.org/10.3758/s13414-017-1361-2

Xie, Z., Yi, H. G., & Chandrasekaran, B. (2014). Nonnative audiovisual speech perception in noise: Dissociable effects of the speaker and listener. *PloS one*, *9*, e114439. https://doi.org/10.1371/journal.pone.0114439

Xu, Y., Lee, A., Wu, W.-L., Liu, X., & Birkholz, P. (2013). Human vocal attractiveness as signaled by body size projection. *PLoS One*, *8*(4), e62397. https://doi.org/10.1371/journal.pone.0062397

Yi, H.-G., Phelps, J. E. B., Smiljanic, R., & Chandrasekaran, B. (2013). Reduced efficiency of audiovisual integration for nonnative speech. *The Journal of the Acoustical Society of America*, *134*(5), EL387–EL393. https://doi.org/10.1121/1.4822320

Zhao, T. C., Masapollo, M., Polka, L., Ménard, L., & Kuhl, P. K. (2019). Effects of formant proximity and stimulus prototypicality on the neural discrimination of vowels: Evidence from the auditory frequency-following response. *Brain and Language*, *194*, 77–83. https://doi.org/10.1016/j.bandl.2019.05.002

Zuckerman, M., & Driver, R. E. (1989). What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior*, *13*(2), 67–82. https://doi.org/10.1007/BF00990791

Zuckerman, M., & Miyake, K. (1993). The attractive voice: What makes it so? *Journal of Nonverbal Behavior*, *17*(2), 119-135.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, *32*(1), 25–64. https://doi.org/10.1016/0010-0277(89)90013-9

Appendix A

Prime-target word pairs used in Chapter 4.

Word pairs			Continued	Word pairs	
Item	**Prime/Target**	**Prime/Target**	**Item**	**Prime/Target**	**Prime/Target**
1	A-no-rak	A-na-nas	31	Mi-ne-ral	Mi-na-rett
2	Ex-a-men	Ex-o-tik	32	E-le-ment	E-lo-quenz
3	Ka-ra-ffe	Ka-ro-sse	33	Ga-lee-re	Ga-llo-ne
4	Ka-ta-log	Ka-tho-lik	34	Kon-se-quenz	Kon-so-nant
5	Ma-tra-tze	Ma-tro-se	35	Kor-re-lat	Kor-ro-sion
6	Me-la-nom	Me-lo-die	36	Me-li-sse	Me-lo-ne
7	Me-ta-pher	Me-tho-de	37	Di-a-gramm	Di-op-trie
8	Ok-ta-ve	Ok-to-ber	38	Ko-man-do	Ko-mmo-de
9	Os-ma-ne	Os-mo-se	39	Ka-len-der	Ka-li-ber
10	Pa-ra-de	Pa-ro-le	40	Ka-ser-ne	Ka-si-no
11	Pa-ra-sit	Pa-ro-die	41	La-ter-ne	La-ti-no
12	Pi-a-nist	Pi-o-nier	42	La-ven-del	La-wi-ne
13	Pis-ta-zie	Pis-to-le	43	Ak-ri-bie	Ak-ro-bat
14	Py-ra-mi-de	Py-ro-ma-ne	44	An-gi-na	An-go-ra
15	An-te-nne	An-ti-ke	45	Ho-ri-zont	Ho-ros-kop
16	I-de-al	I-di-om	46	Ka-nin-chen	Ka-no-ne
17	No-ve-lle	No-vi-ze	47	Ka-pi-tel	Ka-pu-ze
18	Pas-te-te	Pas-ti-lle	48	Mi-nis-ter	Mi-nu-te
19	Ko-li-bri	Cho-le-ra	49	Fa-cet-te	Fa-ssa-de
20	Al-bi-no	Al-ba-ner	50	Ga-le-rie	Ga-la-xie
21	Ka-bi-ne	Ka-ba-le	51	Ka-the-ter	Ka-thar-sis
22	Ka-bi-nett	Ka-ba-rett	52	In-fek-tion	In-for-mant
23	Li-bi-do	Li-ba-non	53	In-se-rat	In-sol-venz
24	Re-li-gion	Re-la-tion	54	Ar-gu-ment	Ar-go-naut
25	Ter-ri-ne	Ter-ra-sse	55	Mo-nu-ment	Mo-no-pol
26	A-ppe-tit	A-pa-thie	56	Fu-ro-re	Fu-run-kel
27	Fi-ne-sse	Fi-na-le			
28	Kom-men-tar	Kom-man-dant			
29	Kom-pe-tenz	Kom-pa-nie			
30	Li-te-rat	Li-ta-nei			
		Continued			

Appendix B

Sentence items used in Chapter 5.1, 5.2, and 5.3

Item	Cue	Verb	Verb+1	Adjective	Noun
1	Der Junge	bekommt	zwei	alte	Autos
2	Das Mädchen	bewundert	drei	große	Bäder
3	Das Kind	probiert	vier	grüne	Blumen
4	Der Bruder	begutachtet	fünf	kleine	Dosen
5	Die Schwester	kauft	sieben	nasse	Messer
6	Der Lehrer	malt	acht	rote	Ringe
7	Die Frau	lobt	neun	schöne	Schuhe
8	Der Mann	bestellt	elf	schwere	Sessel
9	Der Onkel	sieht	zwölf	teure	Steine
10	Die Tante	erkennt	achtzehn	weiße	Tassen
11	Die Mutter	findet	selten	neue	Beeren
12	Der Vater	bemerkt	besonders	bunte	Vorhänge
13	Der Chef	betrachtet	erneut	edle	Uhren
14	Die Studentin	entdeckt	zwanzig	verschiedene	Röcke
15	Die Dame	sucht	ebenfalls	graue	Katzen
16	Die Erbin	klaut	hundert	antike	Stühle
17	Der Freund	streichelt	vierzehn	schnelle	Hühner
18	Der Bauer	mag	viele	rosa	Lampen
19	Der Neffe	folgt	einem	kleinen	Hund
20	Die Lehrerin	repariert	zwei	silberne	Ketten
21	Der Nachbar	liebt	alle	frechen	Fische
22	Der Verkäufer	beobachtet	dreißig	schwarze	Dackel
23	Die Chefin	braucht	sehr	schicke	Schränke
24	Der Mitarbeiter	plant	häufig	tolle	Feste

The table continues on the next page

Item	Cue	Verb	Verb+1	Adjective	Noun
25	Die Großmutter	zeichnet	gerne	lila	Kühe
26	Der Verwandte	trägt	bald	dünne	Mützen
27	Die Kollegin	hat	trotzdem	mehrere	Häuser
28	Die Schülerin	trinkt	wieder	saure	Milch
29	Der Student	testet	öfters	scharfe	Suppen
30	Der Enkel	ordnet	vierzig	dicke	Bretter
31	Der Herr	verkauft	niemals	blinde	Mäuse
32	Der Großvater	übt	hoffentlich	wilde	Tänze
33	Der Kellner	nervt	nur	starke	Frauen
34	Die Köchin	hilft	montags	armen	Kindern
35	Das Schwein	riecht	immer	staubige	Kissen
36	Der Pfleger	behält	zehn	helle	Mäntel
37	Der Sohn	lernt	stets	fade	Texte
38	Die Ärztin	erbt	endlich	stumpfe	Scheren
39	Die Freundin	druckt	siebzehn	moderne	Kunstwerke
40	Der Bäcker	füttert	vermutlich	junge	Schafe
41	Der Vermieter	bemalt	eine	krumme	Flasche
42	Die Nichte	kocht	meistens	heiße	Tomaten
43	Das Kleinkind	kennt	kaum	lustige	Witze
44	Der Pfarrer	pflegt	gern	laute	Vögel
45	Die Sängerin	zeigt	manchmal	offene	Boote
46	Die Künstlerin	wäscht	doch	lange	Hosen
47	Der Maler	übersieht	kein	braunes	Haus
48	Die Nachbarin	erzieht	unzählige	mutige	Enten

Appendix C

Trivia statements used in Chapter 6.

Item	Sentences	Condition	Truth
1	Wenn eine Essiggurke an Strom angeschlossen wird, leuchtet sie im Dunkeln.	experimental	True
2	Manche Arten von Schnecken haben mehr als 20.000 Zähne.	experimental	True
3	Kängurus können nicht rücktwärts springen.	experimental	True
4	Der Elefant ist das einzige Säugetier, dass nicht springen kann.	experimental	True
5	In Europa gibt es keine einzige Wüste.	experimental	True
6	Giraffen sind die einzigen Tiere, die mit Hörnern geboren werden.	experimental	True
7	Das Feuerzeug wurde vor dem Streichholz erfunden.	experimental	True
8	Flöhe können um das hundertfache ihrer eigenen Körperlänge in die Höhe springen.	experimental	True
9	Die Sonne wird jede Stunde eineinhalb Meter kleiner.	experimental	True
10	Tiger haben nicht nur ein gestreiftes Fell, sondern auch gestreifte Haut.	experimental	True
11	Die meisten Eisbären sind Linkshänder.	experimental	True
12	Eisbären schwimmen hunderte Kilometer ohne Pause.	experimental	True
13	Der Dosenöffner wurde erst Jahrzente nach der Konservendose erfunden.	experimental	True
14	Ein Flusspferd kann schneller rennen als ein Mensch.	experimental	True
15	Eine Giraffe kann länger ohne Wasser leben als ein Kamel.	experimental	True
16	Das Gehirn des Vogel Strauß ist kleiner als sein Auge.	experimental	True
17	Austern können ihr Geschlecht wechseln.	experimental	True
18	Ameisen schlafen nicht.	experimental	True
19	Obwohl das Eisbärenfell weiß aussieht, ist es eigentlich farblos.	experimental	True
20	Ein Albatros kann schlafen während er fliegt.	experimental	True
21	Die Sonne hat fast 100 Prozent der Masse des gesamten Sonnensystems.	experimental	True
22	Kamele haben drei Augenlider, um sich vor Sand zu schützen.	experimental	True
23	Alle Schwäne in England gehören der Königin.	experimental	True
24	Eine Mücke hat zwei Zähne.	experimental	False
25	Deutschland ist das erste Land, in dem es Briefmarken gab.	experimental	False
26	Der Koala ist das einzige Tier, dass nie krank wird.	experimental	False
27	Rom war die erste Stadt mit einer Polizeibehörde.	experimental	False
28	Ein Adler hat ungefähr 20.000 Federn.	experimental	False
29	Männliche Blauhaie haben doppelt so dicke Haut wie weibliche.	experimental	False
30	Ein neugeborener Eisbär kann ein Jahr lang nichts sehen und hören.	experimental	False

Table continues on the next page

Item	Sentences	Condition	Truth
31	Der Blutegel hat 32 Herzen.	experimental	False
32	Jerusalem ist die älteste Stadt der Welt.	experimental	False
33	Haie greifen Frauen zehn Mal öfter an als Männer.	experimental	False
34	Falken sind die einzigen Vögel, die die Farbe blau sehen können.	experimental	False
35	Frauen blinzeln 10 Mal öfter als Männer.	experimental	False
36	Bananen wachsen auf Bäumen.	experimental	False
37	Hunde schwitzen mit der Spucke.	experimental	False
38	Krokodile können nicht ohne Nahrung mehrere Tage überleben.	experimental	False
39	Der Jupiter dreht sich andersrum als die anderen Planten im Sonnensystem.	experimental	False
40	Die erste öffentliche Bibliothek war in Wien.	experimental	False
41	Nur junge Eisbären halten einen Winterschlaf.	experimental	False
42	Irland hat nach Amerika die meisten Brauereien.	experimental	False
43	Eine Schnecke kann Jahrzehnte lang schlafen.	experimental	False
44	Nur 15 Prozent vom Wasser auf der Erdoberfläche kann man trinken.	experimental	False
45	Regenwürme haben fünf Gehirne.	experimental	False
46	Delfine gehören zu den Fischen.	filler	False
47	Haie können auch rückwärts schwimmen.	filler	False
48	Ostereier sammelt man an Pfingsten.	filler	False
49	Brokkoli ist ungesund.	filler	False
50	Heuschnupfen gibt es nur im Frühling.	filler	False
51	Am Muttertag bekommt der Papa ein Geschenk.	filler	False
52	Zwerge sind mindestens 2 Meter groß.	filler	False
53	Manche Krokodile essen andere Krokodile.	filler	False
54	Weihnachten ist im Winter.	filler	True
55	Wenn man eine Katze streichelt, dann schnurrt sie manchmal.	filler	True
56	Papageien plappern gerne Menschen nach.	filler	True
57	Das Fell eines Kaninchens ist kuschelig weich.	filler	True
58	Das Jahr hat vier Jahreszeiten.	filler	True
59	Eine Woche hat sieben Tage.	filler	True
60	Fische atmen unter Wasser.	filler	True

List of Tables

List of Figures