Individual Differences in Speech Production

And Maximum Speech Performance

Chen Shen

沈晨

Funding body
This research was funded by EU's H2020 research and innovation programme under the MSCA GA 675324 (ENRICH European Training Network, early stage researcher fellowship for Chen Shen) (www.enrich-etn.eu).

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

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

ISBN
978-94-92910-35-6

Cover
Chen Shen & Promotie In Zicht

Design/lay-out
Promotie In Zicht | www.promotie-inzicht.nl

Print
Ipskamp Printing

Individual Differences in Speech Production

And Maximum Speech Performance

Proefschrift

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

woensdag 8 juni 2022
om 12.30 uur precies

door

Chen Shen
geboren op 8 maart 1990
te Lanzhou (China)

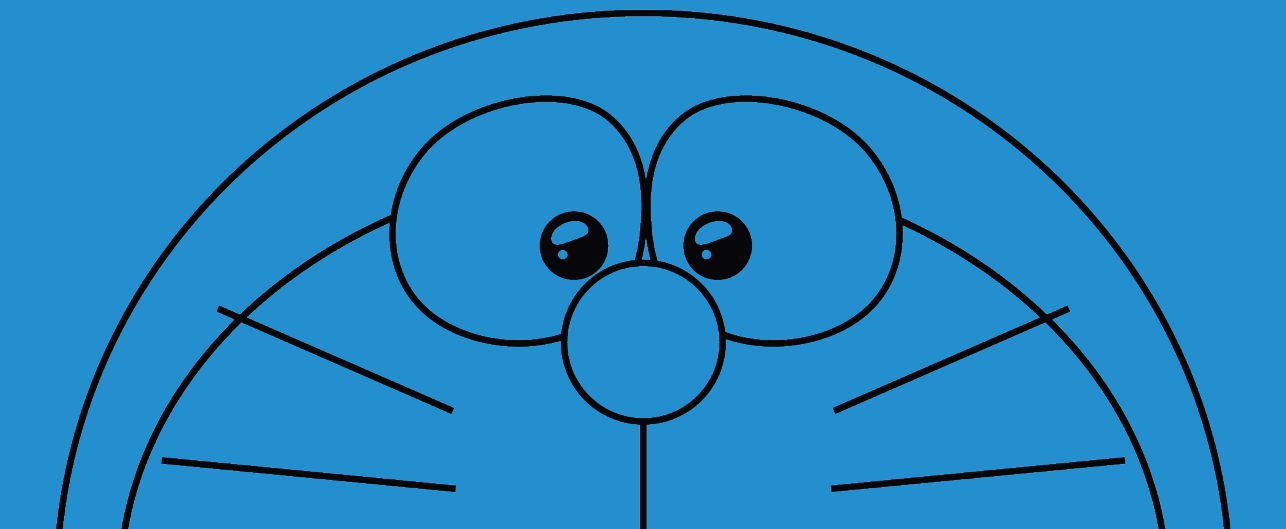
Promotor
Prof. dr. Mirjam Ernestus

Copromotor
Dr. Esther Janse

Manuscriptcommissie
Prof. dr. Amalia Arvaniti
Prof. dr. Ben Maassen (Rijksuniversiteit Groningen)
Prof. dr. Hugo Quené (Universiteit Utrecht)
Dr. Hans Rutger Bosker (MPI)
Dr. Marina Laganaro (Université de Genève, Zwitserland)

Van het concert des levens krijgt niemand een program.

Contents



1

General Introduction

In the world of around seven-and-a-half-billion people, no two speakers are alike. The process of speech production begins with an intended message and ends with articulated audible speech. Apart from differences in what words or phrases speakers use, speakers each have a unique combination of different features in their speech: tempo, accent, pitch, voice quality, and speech clarity are among the most salient features.

Apart from general differences between speakers, speakers also vary their speech depending on the situation or environment they communicate in. As our everyday speech communication rarely happens in a noise-free environment, speakers often need to adapt their ways of talking to counter ambient noise (e.g., in noisy restaurants or pubs) to make sure their messages are properly understood by listeners. Lombard speech, for instance, is commonly referred to as the type of speech produced by speakers to compensate for loud background noise during speech communication (Junqua, 1993; Lombard, 1911; Van Summers et al., 1988). In order to improve speech intelligibility in the presence of ambient noise and/or when the interlocutor suffers from (slight) hearing loss, speakers typically are capable of enriching their speech (Ferguson & Morgan, 2018). Additionally, speakers are even able to produce various types of listener-oriented clear speech depending on their interlocutors (for a review, see Cooke et al., 2014). Despite this general adaptive ability, we may experience that some speakers are more intelligible, particularly in noisy settings, than others (e.g., Bradlow, et al., 1996; Ferguson, 2004, 2012; Hazan & Markham, 2004). How do speakers convey their messages such that they meet their communicative intentions? Moreover, if speakers change their speech to counter ambient noise, how do they estimate and weigh the needs of their interlocutor(s) and their own? Do speakers mainly speak up to hear themselves better or to facilitate their interlocutor's understanding? What is the vocal and articulatory effort speakers have to put into speaking more loudly and clearly? Apart from controlling and adapting their speech to meet communicative needs in adverse communicative settings, speakers also need to control their speech output to prevent speech errors. Even though adult speakers already have years of experience speaking, they may still stumble over sentences such as 'she sells sea-shells by the seashore', where constant alternation is needed between the 's' and the 'sh' sound at word onsets. What kind of control abilities are required to fluently and successfully produce such 'tongue-twisting' sentences?

To address these questions, this thesis investigates different aspects of speech variability. One important source of between speaker variability may be associated with their cognitive abilities. **Chapters 2** and **3** address the link between speakers' cognitive control and maximum speech performance (indexing speakers' speech motor control). As ageing is often associated with cognitive decline, **Chapter 4**

focuses on age differences in the time speakers need to prepare and initiate speech. As mentioned above, speakers also differ in the way they modify their speech in noisy conditions. In **Chapter 5**, speakers' speech enrichment is investigated from two angles: the consistency with which they maintain their speech enrichment; and the potential link between their speech enrichment behaviour and indices of their speech motor control.

In the remainder of this chapter, I first explain which processes in speech production are considered as 'late stages' of speech production. Then I steer the focus to individual differences in late stages of speech production to look at how individuals differ in their speech production when they are asked to flexibly adapt their speech according to instructions (e.g., to speak as fast as possible or to speak clearly). Finally, I outline the research questions in this thesis in more detail.

1.1. Late stages of speech production

According to Levelt's classical model of speech production (Levelt, 1989; Levelt et al., 1999), the process of generating speech from thoughts can be categorised into three main stages, namely, conceptualisation, formulation, and articulation (see Figure 1). Generally speaking, speakers need to firstly set their communicative intention and the target concepts, and then retrieve words, generate the morpho-syntactic structure as well as the phonological shape of utterances. After the completion of these conceptualisation and formulation steps, speakers proceed to the 'late stages' of speech production to plan, programme and execute articulatory programmes to articulate. Speech production has often been studied through measuring speakers' reaction times in tasks like picture naming. In a meta-analysis of word production (including picture-naming) studies by Indefrey and Levelt (2004), the late stages of speech production (i.e., syllabification and phonetic planning till initiation of articulation) were estimated to take up almost half of the time (around 270 ms) needed for the entirety of the production process (around 600 ms). The complex process of articulation involves a collection of motor programmes that control over 100 muscles (e.g., the articulators, vocal folds, and the abdomen). The terms articulatory control and speech motor control are often used interchangeably to refer to the systems that regulate the production of speech, including the planning and preparation of movements and the execution of movement plans that result in muscle contractions (Kent, 2000).

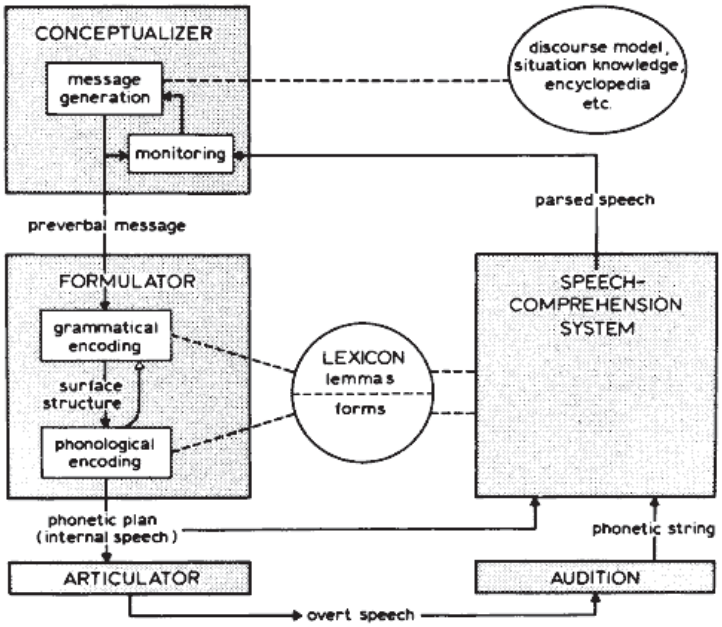


Figure 1.1
An illustration of the blueprint for the speaker: the processing components involved in the generation of speech (taken from Levelt, 1989, Chapter 1)

After phonological encoding, which is the latest step in the formulation phase, the process of speaking still requires neuromotor mechanisms to implement the correct sequencing of motor programmes, action execution, and response monitoring (Tremblay et al., 2019). Monitoring of one's own speech in the Levelt 'blueprint for the speaker', depicted in Figure 1, is accomplished in two ways. Monitoring here specifically functions to intercept speech errors either *before* they have been realised (through monitoring of inner or internal speech) or *after* they have been made in overt speech (monitoring of speech through audition), with both monitoring routes feeding into one's own speech comprehension system (i.e., the perceptual loop theory). Once errors have been detected in inner or overt speech, repairs can be initiated (Dell, 1986; Levelt, 1989). Newer models of verbal monitoring have proposed that monitoring is accomplished via *production-internal* mechanisms, as laid out in the conflict monitoring account (Nozari et al., 2011) and follow-up models (e.g., Gauvin & Hartsuiker, 2020). These conflict monitoring accounts propose that the production system monitors whether there is 'conflict' between highly-active competing representations. Correct productions would then be cases in which there is only a

single highly-active target representation, whereas erroneous productions would be cases with multiple representations with high activation levels. Conflict information from lexical-semantic or phonological selection levels is passed on to a domain-general executive control system for potential error detection.

The 'late stages' of speech production have been described in more detail in the Hierarchical State Feedback control model (Hickok, 2012) and in its precursor, the DIVA (Directions Into Velocities of Articulators) model, which is a computationally implemented neural-network model of speech motor control (Guenther et al., 2006; Guenther, 2016; Guenther & Vladusich, 2012; Tourville & Guenther, 2011). As in the model of speech production by Levelt and colleagues (Levelt, 1989; Levelt et al., 1999), the DIVA model also assigns an important role to sensory monitoring systems in speech production such that speakers check whether their speech comes out as planned. More specifically, the model consists of a feedback control system (including auditory control and somatosensory control sub-systems) and a feedforward control system. The feedforward control system enables the stored motor programmes to be executed and sets up predictions (or forward models) on what the speaker is about to feel and hear (upon realising the speech). Upon speech realisation, the feedback control system then compares the observed somatosensory and auditory input to expected sensory targets. The DIVA model assigns a crucial role to auditory and somatosensory feedback during first language acquisition when infants need to learn the relationship between motor actions and their auditory and somatosensory consequences (Guenther & Vladusich, 2012; Tourville & Guenther, 2011). In adult speech, the role of auditory feedback may have become less important than during language acquisition because speakers can now rely on a stable feedforward system. Despite a stable feedforward system, adult speakers are still thought to integrate their auditory feedback into their sensorimotor control through monitoring of their speech output. If mismatches between the expected and the actual motor commands are detected, the feedback control system may generate compensatory/corrective motor commands. Consequently, these compensatory/corrective motor commands can be used to update the feedforward control system (Guenther & Vladusich, 2012; Tourville & Guenther, 2011).

Whereas the DIVA model stems from a research tradition focusing on motor control, most psycholinguistic studies following up on the Levelt blueprint for the speaker have focused on higher-level linguistic processes *preceding* the motor acts required for articulation (Hickok, 2012). The involvement of control mechanisms in the so-called 'early stages' of speech production has been studied extensively over the past years (e.g., Costa et al., 2006; Piai & Roelofs, 2013; Shao et al., 2012; Sikora et al., 2016). The involvement of control mechanisms in the 'late stages' of speech production,

on the other hand, has received less research attention. Given that the articulation stages take up about half of the time needed to produce a spoken word (Indefrey & Levelt, 2004), this research focus seems rather unbalanced. Furthermore, while there is ample evidence that 'early' stages like lemma selection are linked to cognitive control (Piai & Roelofs, 2013; e.g., Shao et al., 2012; Sikora et al., 2016), fewer studies so far have investigated the link between articulatory and cognitive control, and their results are also mixed.

Studies from different fields have found different results concerning whether the late stages of speech production are 'automatic' or not. Some psycholinguistic studies have argued that stages of speech production following lexical selection, such as phonetic encoding, motor programming, and articulatory execution, do not require processing resources (Ferreira & Pashler, 2002; Garrod & Pickering, 2007). However, recent studies looking into late stages of speech production have provided evidence for the involvement of executive control abilities such as sustained attention in the (articulation) processes following phonological encoding (e.g., Jongman et al., 2015). Studies that looked into articulatory control have also suggested a potential relationship between executive control and articulatory control abilities (Dromey & Benson, 2003; Nijland et al., 2015). One study compared the occurrence and detection of speech errors made by people who stutter and fluent controls (Brocklehurst & Corley, 2011). Participants in their study were asked to rapidly produce four-word sequences like 'rag lap lash rap' in different speaking conditions. Speech errors could involve word onsets (the typical substitutions induced by tongue twister phrases that were assumed to arise during phonological encoding), or word-order errors, which would arise during late stages where the planned speech has to be articulated in the correct order. Results showed that, across speaking conditions and groups, those with better working memory performance (as quantified by digit-span performance), produced fewer word-onset and word-order errors (Brocklehurst & Corley, 2011). According to the authors, these results attested to the general association between working memory and phonological encoding (see also Acheson & MacDonald, 2009a, 2009b). All in all, the results from the majority of these studies seem to argue against automaticity of the late stages of articulation (i.e., phonological encoding and later processes).

Across speech studies that have linked speech to cognitive control, different terms and tasks have been used to refer to cognitive or 'executive' control. Executive control is known as a set of general-purpose control mechanisms that regulate our thoughts and actions (Gilbert & Burgess, 2008; Logan, 1985). According to Miyake and colleagues (2000), executive control consists of three main components: inhibitory control (the ability to suppress activation of unwanted information in order to resolve

conflict), cognitive switching (the ability to rapidly switch back and forth between mental sets or operations), and updating of working memory (the ability of maintaining or actively refreshing the contents of working memory while processing incoming information). This thesis adopts this framework of executive control by Miyake and colleagues and its three main components to investigate the relation between speech production and executive control.

If articulatory control relates to general executive or cognitive control, adult ageing may also affect the control of speech movement through age-related changes in cognitive abilities (Glisky, 2007; Salami et al., 2012). Evidence on the link between executive control and action control comes from the field of (hand) movement control. Age differences in the planning and execution of complex rhythm production (in the form of hand tapping) have been argued to be modulated by executive control (Krampe et al., 2005). Using a computerised motor sequencing task (i.e., Push-Turn-Taptap task), Niermeyer and colleagues (2017) found that the learning aspect of motor sequencing was uniquely associated with executive control for older adults, especially for complex sequences. Additionally, in a study on step-initiation (requiring leg control), age differences in motor control were found to be modulated by inhibition requirements (Sparto et al., 2014), illustrating that age differences may be enlarged in more cognitively demanding conditions.

Returning to the study of speech, a few recent studies on speech motor performance have revealed that speech production in cognitively healthy older adults may be affected by age-related declines in the planning and execution of speech movements and speech motor performance (Tremblay, et al., 2018; Tremblay et al., 2017). Specifically, using a pseudoword production task, Tremblay and colleagues (2018) studied the effects of ageing on motor aspects of speech production. They found that age affected the speed and stability with which the speech was realised (acoustically). Additionally, using the Stroop interference paradigm, MacPherson (2019) showed age differences in the size of the Stroop effect on speakers' speech motor performance. Compared to younger adults, older adults displayed more articulatory variability and longer movement duration especially in the more cognitively demanding (incongruent) condition. MacPherson (2019) therefore suggested that older adults' speech motor performance may have been affected by age-related decline in cognitive and motoric functions. However, from these studies, it is unclear which aspect(s) of speech motor planning may be specifically susceptible to age-related decline. It is open to question whether age specifically affects speech planning, or initiation, or execution of speech. This thesis will attempt to decompose the late stages of speech production in order to better understand the relationship between cognitive decline, age effects, and late stages of speech production.

1.2. Speech enrichment in noisy environments

As noted earlier, speakers are capable of adjusting their speech according to various environmental and communicative needs. This flexibility is backed up by the hyper- and hypo-articulation (H&H) theory, which proposes that speakers can vary their speech output along a continuum of hyper-speech and hypo-speech to strike a balance between speaking as clearly as possible for the sake of the listener (hyper-articulated speech), while spending as little effort as possible (hypo-articulated speech) (Lindblom, 1990). When communication takes place in noisy environments, speakers would generally produce the so-called 'Lombard speech', which often displays characteristics that reflect speakers' increased vocal effort of 'speaking up' (e.g., Van Summers et al. 1988). Typically, Lombard speech exhibits acoustic features such as reduced articulation rate, raised fundamental frequency (F_0), expanded F_0 range, and enhanced 1-3 kHz frequency emphasis (Bradlow et al., 1996; Cooke & Lu, 2010; Garnier & Henrich, 2014; e.g., Junqua, 1993; Lu & Cooke, 2009; Tuomainen & Hazan, 2016). A number of perception studies have demonstrated the existence of a Lombard intelligibility gain (Pittman & Wiley 2001; Lu & Cooke 2008; Cooke & Lecumberri 2012). This intelligibility gain implies that Lombard speech is more intelligible than 'plain' speech if it is presented in the noisy background it was supposed to counter (e.g., Lu & Cooke, 2008). Moreover, this gain is perceived by both native- and non-native listeners, illustrating that this speaking style is generally more robust against noise degradation (e.g., Cooke & Lecumberri, 2012). However, there have been debates on the nature or cause of this Lombard type of 'enriched speech'. Some researchers have argued that Lombard speech is mainly produced as some sort of involuntary reflex in response to reduced auditory feedback caused by loud noise (Lombard, 1911; Pick et al., 1989), such that the reflex would be instantaneous and driven by speakers' need to hear their own speech for monitoring purposes (Huettig & Hartsuiker, 2010). Others have argued that Lombard speech is listener-driven, and is motivated by the (perceived) need to compensate for reduced intelligibility caused by noise for the listeners (Garnier et al., 2008; Garnier et al., 2010; Lane & Tranel, 1971). It is likely that both speaker- and listener-driven mechanisms contribute to the changes made by speakers in response to noisy environments, especially when there is a clear communicative intent (Hazan & Baker, 2011; Villegas et al., 2021; Zollinger & Brumm, 2011).

There are more speaking styles that can be thought of as 'enriched' speech. Instructed clear speech (see Lam & Tjaden, 2013), a type of speech that speakers produce to overcome various difficult communicative needs (including noise), has been reported to exhibit overlapping acoustic features with Lombard speech, such as slower articulation rate, enhanced pitch modulation, and enhanced vowel articulation (for a

review, see Uchanski, 2008). This clear speaking style has been considered as a conscious choice by the speaker for the benefit of the listener. Indeed, upon presentation of sentences spoken in either a conversational style or an instructed clear style (as if you address someone with a hearing loss), both normal-hearing younger adults and older adults with hearing loss rated the clear speech to be clearer than conversational speech (Ferguson & Morgan, 2018). The clear-speech features require hyper-articulation or extra articulatory effort from a speaker (Hazan & Baker, 2011; Krause & Braida, 2004; Maniwa et al., 2009), and speakers have been shown to tailor their speech to listeners' needs according to different adverse listening conditions (e.g., Hazan et al., 2012) and linguistic backgrounds (e.g., Lee & Baese-Berk, 2020).

Most literature on various types of enriched speech focuses on the (across-the-board) presence or absence of certain acoustic features. More detailed aspects of speakers' enriched Lombard speech production, such as the maintenance of it, still remain poorly understood (but see Ferguson, 2012 and Ferguson and Morgan, 2018 for effects of experience talking to people with hearing loss). Are speakers able to immediately switch to a clearer speaking style, and do they maintain it for a longer time? One recent study by Lee and Baese-Berk (2020) addressed these questions through studying speakers' use and maintenance of clear speech in an interactive speech task (i.e., Diapix, where speakers have to describe their version of a pictured scene to their interlocutor who has a slightly different version of the picture). Although not explicitly instructed to speak clearly, native-English speakers in their study were found to have produced more intelligible speech when talking to non-native (rather than native) English listeners. Moreover, their speech was found to be more intelligible in the early instead of the late portions of the conversation, and that speakers tended to 'reset' to clear speech whenever they started their description of a new picture. Lee and Baese-Berk therefore concluded that the initiation of clear speech at topic boundaries could be listener-oriented to clearly set the stage while introducing a new topic. At the same time, once the conversation topic has been established, speakers may gradually spend less articulatory effort when the interlocutor needs less clarification, thereby following the H&H theory (Lindblom, 1990).

Lee and Baese-Berk (2020) showed that speakers may generally be able to adapt their speech in a dynamic manner depending on whom they are talking to (i.e., their interlocutor's language status being native or non-native) and depending on how the conversation proceeds. However, individual speakers have been shown to exhibit speaker-specific and language-independent traits that make them differ in their baseline speech intelligibility (e.g., Bradlow et al., 1996). Individual speakers also differ in the extent to which they enrich their speech in noisy communicative settings

(e.g., Ferguson, 2012), and most likely, also in their maintenance of this clear speaking style. It is unclear whether specific articulatory control abilities might predict which speakers are better able to enrich their speech in noise, or to better maintain their clear speaking style. Therefore, an investigation into the inter- and intra-speaker differences in speech intelligibility in quiet and noisy settings can speak to the nature of these changes in clear speaking style.

1.3. Outline and research questions

The research in this thesis concerns late (i.e., phonological encoding and articulation) stages of speech production. There are two main goals in this thesis. The first goal is to investigate the link between executive control and speech motor control. The second goal is to investigate inter- and intra-speaker variability in the way speakers adapt their speaking style when they are asked to speak clearly while hearing loud noise. For the first goal, how speech motor control relates to executive control (**Chapters 2** and **3** with **Chapter 2** focusing on measuring speech motor control mainly), and which processes of speech planning (i.e., motor programming and response initiation) are affected by ageing (**Chapter 4**) were investigated. In **Chapter 5**, the second goal was investigated through examining the variability of enriched speech production. Specifically, speakers' speech enrichment capabilities in terms of consistency and enrichment success were explored in relation to their individual speech motor control ability.

Chapter 2 sets out to investigate the link between two tasks that have been used to probe speakers' maximum speech performance. The one task has been used in clinical settings to study age differences or differences between patient groups and controls (i.e., the Diadochokinesis or in short, DDK task) (Bernthal et al., 2009; Duffy, 2013). In this DDK task, speakers are asked to rapidly repeat one or more pseudowords. Speakers' DDK task performance has been argued to be a stable index of their oral motor skill (Duffy, 2013; Fletcher, 1972; Kent et al., 1987). The other task, the tongue twister task, has been used to elicit speech errors in psycholinguistic studies (Goldrick & Blumstein, 2006; Wilshire, 1999). As tongue twister phrases are typically constructed with repeating and alternating phonemes, they may elicit phoneme selection errors. Even though these two tasks both tap articulatory control, to our knowledge, task performance has never been compared. Establishing the relationship between speakers' performance on the two tasks was done in preparation for the further analysis (in **Chapter 3**) of the link between articulatory control as assessed by these two tasks on the one hand, and executive control on the other.

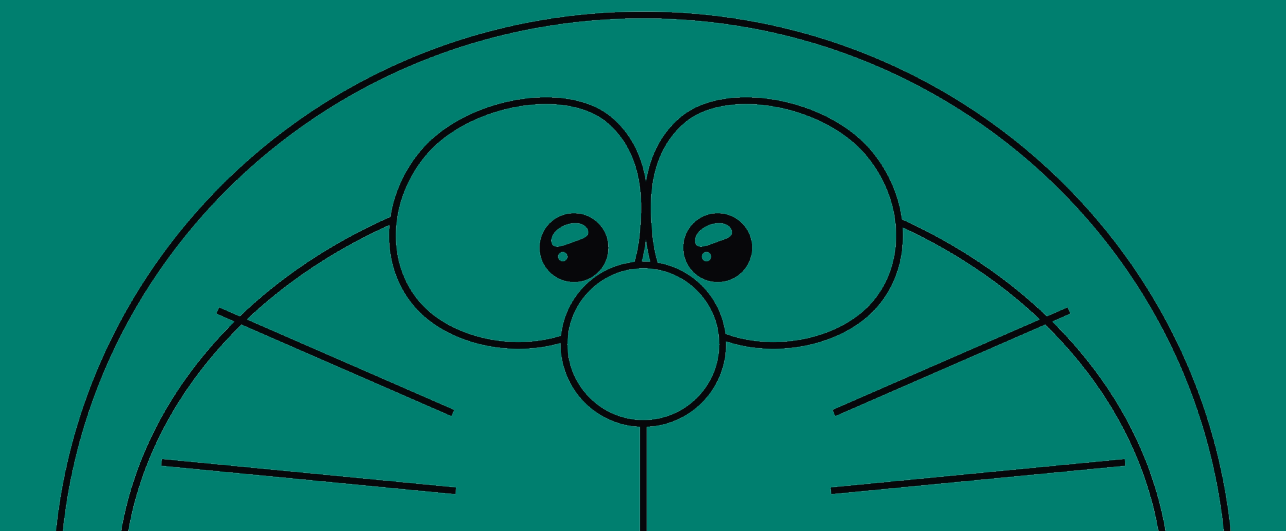
Chapter 3 addresses the research question '**does articulatory control (quantified as maximum speech performance) relate to executive control abilities?**'. This chapter follows up on the two tasks introduced in **Chapter 2**, and it relates speakers' maximum speech performance to measures of executive control. Executive control is operationalised as three main components: inhibitory control, cognitive switching, and updating of working memory. In the context of speech production, executive control may be involved in the following ways: updating ability may be needed for speakers to manage communicative goals while monitoring their production processes. Additionally, when producing a tongue twister phrase, speakers may need to inhibit activated but incorrect phonemes, and to constantly switch between similar phonemes or similar motor programmes. Maximum speech performance on the two aforementioned speech tasks may relate to executive control because both speech tasks require speakers to reach the limits of their motor speech system and to control their articulators. An individual-differences approach is used to investigate whether speakers' executive control abilities (of inhibition, switching, and updating of working memory) predict their ability to rapidly and successfully alternate between similar syllables during speech production (at maximum performance levels).

Ageing is another factor that contributes to between speaker variability. **Chapter 4** addresses the research question '**does adult ageing affect speech motor planning and/or speech initiation?**'. In **Chapter 4**, an age group comparison is carried out to test for potential age-related decline in late stages of speech production. The focus is on processes of speech preparation that could be gleaned from vocal onset RTs (the time before speech realisation is acoustically measurable). Using a speeded speech production task with which the processes of speech motor planning and initiation could be distinguished, which (if any) aspect(s) of speech motor planning is susceptible to age-related slowing is examined.

Another aspect of speech variability lies in speech enrichment. **Chapter 5** addresses the research question '**how consistent are speakers in their speech enrichment strategies over the course of a sentence list?**'. Using an individual-differences approach, inter-speaker variability in speech enrichment modifications that speakers apply moving from baseline (habitual or 'plain') speech to instructed clear-Lombard speech was examined first. In this instructed clear-Lombard speech, participants are instructed to read out sentence lists clearly while hearing loud noise played over headphones. The consistency of speakers' habitual speech over time, as indexed by several acoustic features, as well as the consistency of speakers' speech enrichment modifications over a certain period of time were then examined. More specifically, I investigated whether speaker characteristics, such as articulatory control ability, relate to (the consistency of) speakers' speech enrichment modifications and their

speech intelligibility (as measured by an automatically derived acoustic metric that serves as a proxy of speech intelligibility in noise).

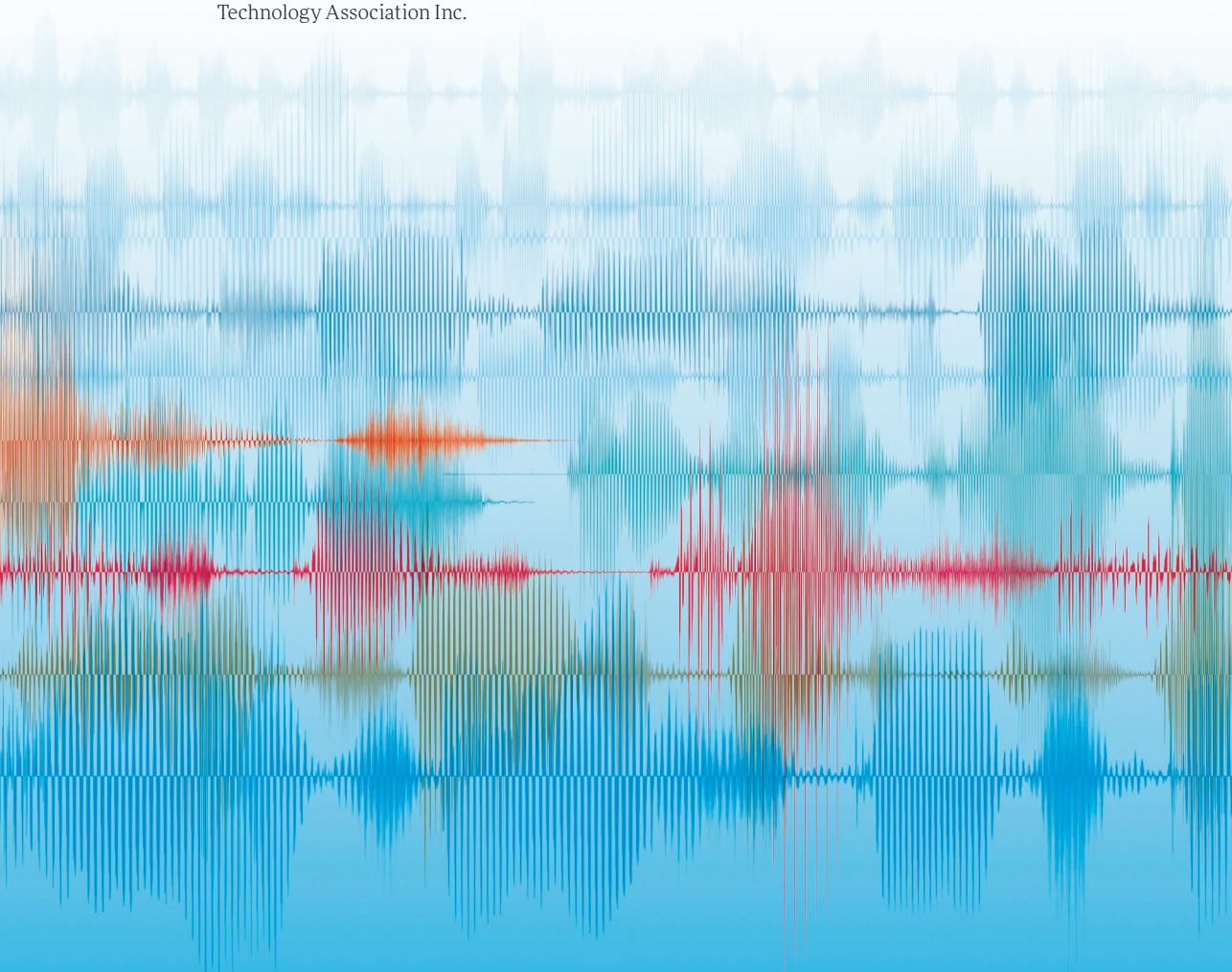
Lastly, **Chapter 6** summarises the results of the experimental **Chapters 2-5** exploring age and individual differences in late stages of speech production. Additionally, it discusses the results from the previous empirical chapters in light of frameworks of speech production and the hyper- and hypo-articulation theory. Furthermore, it addresses the limitations of the present studies and provides follow-up directions for future studies. Last but not least, **Chapter 6** also describes two corpora developed in this thesis based mainly on data collected for **Chapters 2-5**.



2

Articulatory Control in Speech Production

This chapter is based on the following:
Shen C. & Janse E. (2019). Articulatory control in speech production. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 2533–2537). Canberra, Australia: Australasian Speech Science; Technology Association Inc.



Abstract

Articulatory control can be quantified in various ways. Clinical studies frequently use maximum performance measures (e.g., diadochokinesis or DDK) to elicit speakers' maximum rate of repeating syllable sequences. Psycholinguistic studies, on the other hand, often use tongue twister phrases to elicit speech errors in healthy populations. Although both tasks require speakers to rapidly alternate between similar syllables, no direct comparison has been made to investigate the expected overlap between speakers' performance in these two tasks. We collected speech data from 78 healthy young adults, testing their maximum performance on syllable repetitions and tongue twister sentences, and their habitual reading rate. Our results show that individual maximum speech rate in tongue twister sentences was predicted by maximum DDK rate, illustrating that both tasks contain elements of articulatory control. Speakers' habitual sentence reading rate was, however, not correlated to their maximum rate, highlighting a dissociation between maximum and actual performance in speech rate.

2.1. Introduction

Clinical evaluation of articulatory control often uses repetitive syllable sequences for assessing speech motor capacity in persons with speech disorders (e.g., dysarthria, Duffy, 2013). In this so-called diadochokinesis (henceforth DDK) task, speakers are asked to repeat the same syllable as fast and as accurately as possible (e.g., 'papapapa...') or to alternate between syllables (e.g., to produce 'pataka' repeatedly). The latter task thus asks speakers for their maximum performance (in terms of rate and accuracy) in quickly alternating between syllables that only differ in place of articulation of the onset consonant: labial-alveolar-velar. Likewise, production of tongue twister sentences also requires speakers to alternate between similar onset phonemes, between similar onset clusters, or between singleton onset phonemes and onset clusters. Evidently, production of a meaningful sentence such as a tongue twister sentence entails more linguistic processing than repeating nonsensical syllable sequences. First, a tongue twister sentence requires reading or memorising of a longer fragment than a DDK stimulus. The longer fragment naturally has more variegated alternation between similar syllable onsets than a DDK stimulus. Second, sentence production entails grammatical and semantic processes that are absent in sequence repetition.

Although both DDK and tongue twister tasks contain elements of articulatory control, to our knowledge, no study so far has investigated the relationship between speakers' maximum performance on these two tasks. This is most likely due to the former (DDK) task being typically used in a clinical setting (Duffy, 2013; Fletcher, 1972; Wang et al., 2004), and the latter task being mainly used in psycholinguistic studies (Acheson & Hagoort, 2014; Goldrick & Blumstein, 2006; McMillan & Corley, 2010) as a means to elicit speech errors from healthy speakers. To quantify articulatory control from different angles, we examined variations within and associations between maximum performance in the two speech tasks in a healthy adult population.

Reference rates for healthy control speakers already exist for DDK in multiple languages, including Dutch (Knuijt et al., 2017). Additionally, several studies have investigated rate differences between DDK performance on repetitions of non-words versus real words in native speakers of multiple languages (Ben-David & Icht, 2017; Icht & Ben-David, 2015). As speakers have access to stored motor programmes for real words, but not for non-words, maximum performance can be expected to be better for word than non-word repetition. Indeed, school-aged children as well as healthy older adults achieved faster repetition rates in producing real word relative to non-word stimuli in DDK tasks (Ben-David & Icht, 2017; Icht & Ben-David, 2015).

In addition, several clinical studies have addressed the question of whether patients' DDK performance is actually representative of their 'normal' speech behaviour, operationalised as their habitual speech rate in sentence reading. Some have stressed the discrepancy between patients' maximum performance on DDK stimuli and their sentence reading rate (Ziegler, 2002), thereby questioning the utility of DDK as a clinical measure. Others have observed that habitual rate in healthy adults is associated with their maximum articulation rate, but note that they have used the very same reading materials for eliciting both habitual and maximum rate (Tsao & Weismer, 1997).

In this study, we aimed to quantify articulatory control using two maximum performance speech tasks (a DDK and a tongue twister task) that require fast and accurate alternation between similar syllables. Through the novel combination of these two speech tasks, we aimed to achieve the following three objectives. First, through the maximum performance speech rate and accuracy measures, we investigated the variability in a sample of young healthy adult speakers on stimuli that differ in the level of linguistic content (ranging from non-words to real words to tongue twister sentences). Second, we examined whether speakers' tongue twister performance is related to their maximum articulatory (DDK) performance, as measured with words and non-words, to explore the underlying articulatory control mechanisms these measures may reflect. Third, we investigated whether speakers' maximum rate measures (in DDK and tongue twister tasks) are associated with their habitual sentence reading rate.

2.2. Methods

2.2.1. Participants

In total, 78 participants (age: M = 23 years, SD = 3; 61 females) completed the speech tasks in the Centre for Language Studies lab at Radboud University Nijmegen. They were reimbursed for their time through course credits or gift vouchers. Participants were all native speakers of Dutch, with no speech, hearing, or reading disabilities, nor past diagnosis of speech pathology or brain injury. Normal or corrected-to-normal vision was also required. All 78 participants gave informed consent for their audio recordings to be analysed.

2.2.2. Description and analysis of the speech tasks

Two speech tasks were used to elicit participants' maximum performance (rate and accuracy) as indices of their articulatory control. An additional sentence reading task was used to gather data for participants' habitual speech rate. Stimuli of all three

tasks were presented using PowerPoint slides on a 24" full HD monitor placed on a table in front of the participant. Recordings were made using a Sennheiser ME 64 cardioid capsule microphone through a pre-amplifier (Audi Ton) onto a steady-state 2 wave/mp3 recorder Roland R-05 in a sound-attenuating recording booth. The first author monitored participants' task progress and controlled the changing of stimulus slides outside the recording booth on the stimulus computer (Dell Precision T3600).

DDK task description and analysis

Clinical DDK task normally contains repetitions of mono- and tri-syllabic nonsense words such as 'pa' and 'pataka'. Given the focus of this study on alternating articulatory movements, we only selected the commonly used tri-syllabic non-word 'pataka' /pataka/, and added the reversed syllable-order variant 'katapa' /katapa/. In addition, two common real Dutch words that were closest to the nonsense words 'pataka' and 'katapa' were added: 'pakketten' /pɑˈkɛtə(n)/ (packages) and 'kapotte' /kaˈpɔtə/ (broken). Whereas no stress pattern was available for the non-words, both real words had lexical stress on the second syllable. The mono- and di-syllabic nonsense stimuli ('pa', 'ta', 'ka', 'pata', 'taka') were presented as practice trials. All of the nonsense words used here were phonotactically legal in Dutch.

During the task, each DDK stimulus was presented in the centre of a full-screen PowerPoint slide. To elicit repetitive production of the stimulus, multiple (nonsense) words were presented next to each other, for instance 'patakapatakapataka...'. Participants were instructed to repeatedly produce the presented stimulus as accurately and as fast as possible. A pre-recorded example was played prior to the practices to familiarise the participants with the task. A brief line of text reminding them about the accuracy and speed of repetition was constantly on-display at the top of each slide. A 2-second pause (preparation time) followed by a 75-millisecond beep-tone was used to mark the start of articulation. Each stimulus was to be repeated for around 10 seconds. Mean DDK task duration was three minutes.

Participants' maximum performance in terms of articulation rate and accuracy was analysed acoustically in *Praat* (Boersma & Weenink, 2017). Most participants were already making some errors in a 3-second time window, but errors generally increased in longer time windows. We therefore opted for a relatively long time-window (7s) to capture accuracy and rate in a reliable and representative way.

Individual DDK articulation rate (syllables/sec) was calculated by multiplying the total number of correct-and-full (non)words produced by each participant in a 7-second time window (or as close to 7-second as possible for the repetition counts to be an integer) by three (syllables), and dividing this number of total syllables by the

actual production time (total-duration minus error-duration, in-breaths, and pauses longer than 200 ms between repetitions).

Individual DDK accuracy (fraction) was calculated as number of correct-and-full repetitions divided by number of all repetitions in the same 7-second time window. A repetition was only counted as correct if it did not contain any form of error or pauses longer than 200 ms within the sequence.

Tongue twister task description and analysis

Following Wilshire's tongue twister paradigm (Wilshire, 1999), we selected four tongue twister sentences that contain a combination of repetition and alternation of word-initial consonants (e.g., **p**oes **k**otst **p**ostzak, and **fr**its **v**indt **v**is **fr**ietjes). Below are the four Dutch tongue twister sentences that were used as test stimuli with their literal English translations in parentheses:

De **p**oes **k**otst in de **p**ostzak (*The cat puked in the mail bag*)

- **Fr**its **v**indt **v**is**fr**ietjes **vr**eselijk **vi**es (*Frits finds fish-fries terribly gross*)
- Ik **b**ak een **pl**ak **b**ak**bl**oedworst (*I fry a slice of blood-sausage*)
- **P**apa **p**akt de **bl**auwe **pl**atte **b**ak**p**an (*Daddy grabs the blue flat frying pan*)

Prior to the task stimuli, two additional tongue twister sentences were presented as practice stimuli:

- **Sl**imme **Sj**aantje **sl**oeg de **sl**ome **sl**ager (*Smart Sjaantje hit the slow butcher*)
- **B**akker **B**as **b**akt de **b**olle **br**oodjes **br**uin (*Baker Bas bakes the round buns brown*)

Participants were instructed to repeat the tongue twister sentences minimally five times as accurately and as fast as possible. As in the DDK task, tongue twister stimuli were each presented in the centre of a full-screen PowerPoint slide with a reminder of the accuracy and speed of repetition. A picture related to the meaning of each tongue twister sentence (e.g., a blue frying pan) was shown on the same slide, and disappeared after about two seconds (preparation time). Participants were instructed to start repeating the tongue twister as soon as the picture disappeared. Mean tongue twister task duration was four minutes.

Maximum performance (rate and accuracy) was analysed acoustically in *Praat* (Boersma & Weenink, 2017). Individual tongue twister rate (syllables/sec) was calculated by averaging the articulation rate of the correct repetitions of the four tongue twister sentences. Rate of each correct stimulus was measured by dividing the number of syllables in a tongue twister sentence by the time used for that repetition.

Similar to accuracy measures in the DDK task, individual tongue twister accuracy (fraction) for the first five repetitions per sentence was calculated by number of correct and fluent repetitions divided by five. A repetition was counted as fluent if it did not contain any form of error or pause longer than 200 ms in the tongue twister sentence.

Sentence reading task description and analysis

In addition to the two maximum performance speech tasks, participants also performed a sentence reading task. The reading task contained 48 meaningful Dutch sentences that are between 12 and 16 syllables in length (e.g., De grote kat heeft de vaas per ongeluk gebroken '*The big cat has accidentally broken the vase*'). Participants were instructed to read the sentences fluently in a natural way. Habitual articulation (HA) rate (syllables/sec) of each speaker was averaged across all 48 sentences.

2.3. Results

2.3.1. Quantifying variability in speech performance

Table 2.1 Speech task performance.

	Rate (syll./sec)			Accuracy (fraction)		
	Mean	*SD*	*CV%*	*Mean*	*SD*	*CV%*
DDK (real word)	6.33	0.71	11.2	0.94	0.06	6.2
DDK (non-word)	5.91	0.93	15.8	0.89	0.10	10.7
Tongue Twister	4.22	0.49	11.5	0.59	0.16	26.8
Habitual Articulation	5.62	0.61	10.9	na	na	na

CV corresponds to coefficient of variation as a variability index ((SD/mean)*100%)

Table 2.1 presents the descriptive statistics of the two maximum performance speech tasks and the sentence reading task. Rate and accuracy measures averaged over task stimuli were entered as dependent variables in two models for rate and accuracy respectively. Task (DDK real word, DDK non-word, and tongue twister) was entered as the fixed effect of interest, with participant as random effect (Baayen et al., 2008). Results from linear mixed-effects analysis, using the lme4 package (Bates et al., 2015), showed that real word DDK performance is significantly better than non-word DDK performance for both rate (t = 5.45, p < .001) and accuracy (t = 2.71, p < .01). Maximum

performance in the tongue twister task is significantly worse than in DDK non-word repetition (t = -25.17, p < .001 and t = -18.24, p < .001 for rate and accuracy respectively). This indicates that the difficulty level of repetitively producing tongue twister sentences is relatively high for healthy young adult speakers, possibly also due to the fact that the tongue twister sentences contain syllables of varying complexity (e.g., some have consonant clusters) and voicing alternation in consonants. Furthermore, the more difficult the speech task, the higher the variability in accuracy between speakers, as evident from the coefficient of variation values (cf. Table 2.1).

2.3.2. Correlations between maximum performance measures and between maximum and habitual rate

Our second question was whether individual's maximum performance in tongue twister and DDK tasks are associated. Figure 1 below shows the between-task correlations for maximum rate.

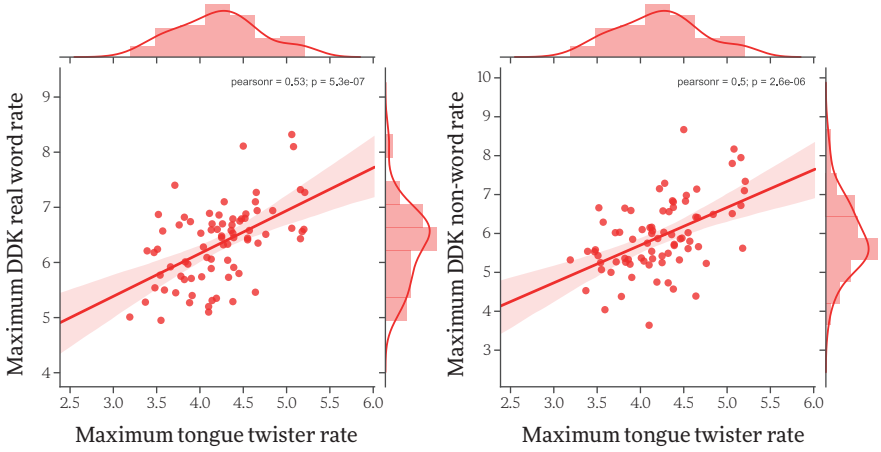


Figure 2.1

Correlations between maximum rate measures in tongue twister and DDK (real word to the left and non-word to the right) tasks

Rates in the two maximum performance speech tasks correlated significantly (r = .53*** for tongue twister and DDK real word rate, r = .50*** for tongue twister and DDK non-word rate). Accuracy of tongue twister production was not correlated with DDK accuracy: neither for DDK word stimuli (r = .13), nor for DDK non-word stimuli (r = .16). This lack of association between accuracy levels may be due to limited variability in DDK accuracy (cf. Table 2.1).

Our last question was whether speakers' maximum rate measures are associated with their habitual sentence reading rate. None of the correlations between rate performance measured in the two maximum performance speech tasks on the one hand and habitual articulation rate on the other reached significance (all r values < .14), suggesting that speech rates which speakers can maximally obtain alternating between similar syllables are not clearly reflected in their habitual sentence reading.

2.4. Discussion

In this study, we investigated articulatory control in a young adult speaker sample through examining their maximum performance (rate and accuracy) in two speech tasks as indices of articulatory control. More specifically, we used a repetitive syllable-sequence production (DDK) task, which is often used in clinical settings, and a tongue twister task, which is typically used as an experimental means to elicit speech errors in non-clinical populations.

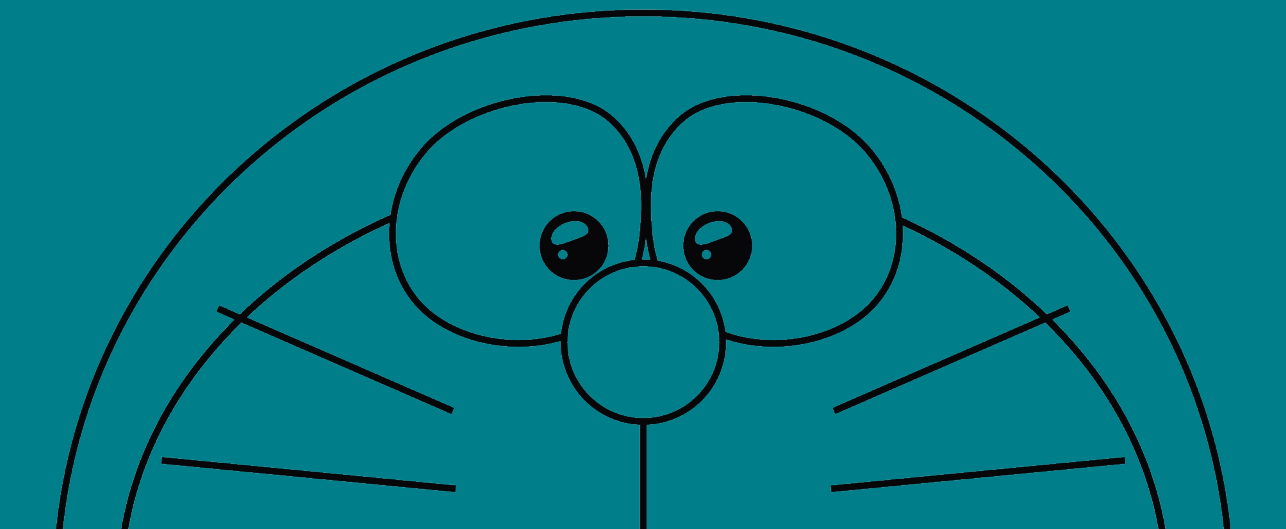
The descriptive statistics show that maximum rate in DDK non-word production and maximum accuracy in tongue twister production were highly variable, even in our homogeneous young and non-clinical speaker group. This variability illustrates that speakers differ considerably in their articulatory control ability.

Our observation of faster DDK performance on real words than nonsense sequences is in line with findings for other languages with school-aged children and older adults (Ben-David & Icht, 2017; Icht & Ben-David, 2015). This may suggest that speakers were better able to rapidly move their articulators in the correct manner when they are more familiar with the required motor programmes. Alternative explanations, however, cannot be ruled out. For instance, confounded with lexicality, words in Dutch have lexical stress patterns (and hence involve unstressed syllables that are reduced acoustically) that are lacking in meaningless sequences like 'pataka' or 'katapa'. Additionally, the word sequences also contained short vowels whereas the non-words only consisted of long vowels, which might have contributed to the rate differences observed between real and nonsense words too.

Our second aim was to examine whether speakers' tongue twister performance is related to their maximum articulatory (DDK) performance, given that both tasks require rapid alternation between similar syllables. Maximum speech rates but not accuracy measures in the two speech tasks were correlated. The rate correlation suggests that both tasks contain elements of speakers' ability to plan and execute similar articulatory programmes, despite differences between tasks in terms of

difficulty level, length of the speech stimuli, and the amount of linguistic processing involved. This then provides evidence for both tasks tapping articulatory control.

Our third and last aim was to assess whether speakers' habitual articulation rate, as measured with a sentence reading task, is associated with their speech rate on (either of) the two maximum performance measures. In line with patient data (Ziegler, 2002) and with rate measures of speakers' semi-spontaneous speech (De Jong & Mora, 2017), maximum rates obtained with neither DDK nor tongue twister production were predictive of speakers' habitual articulation rate. These results highlight a dissociation between maximum and actual performance in speech rate, likely due to differences in task demands.

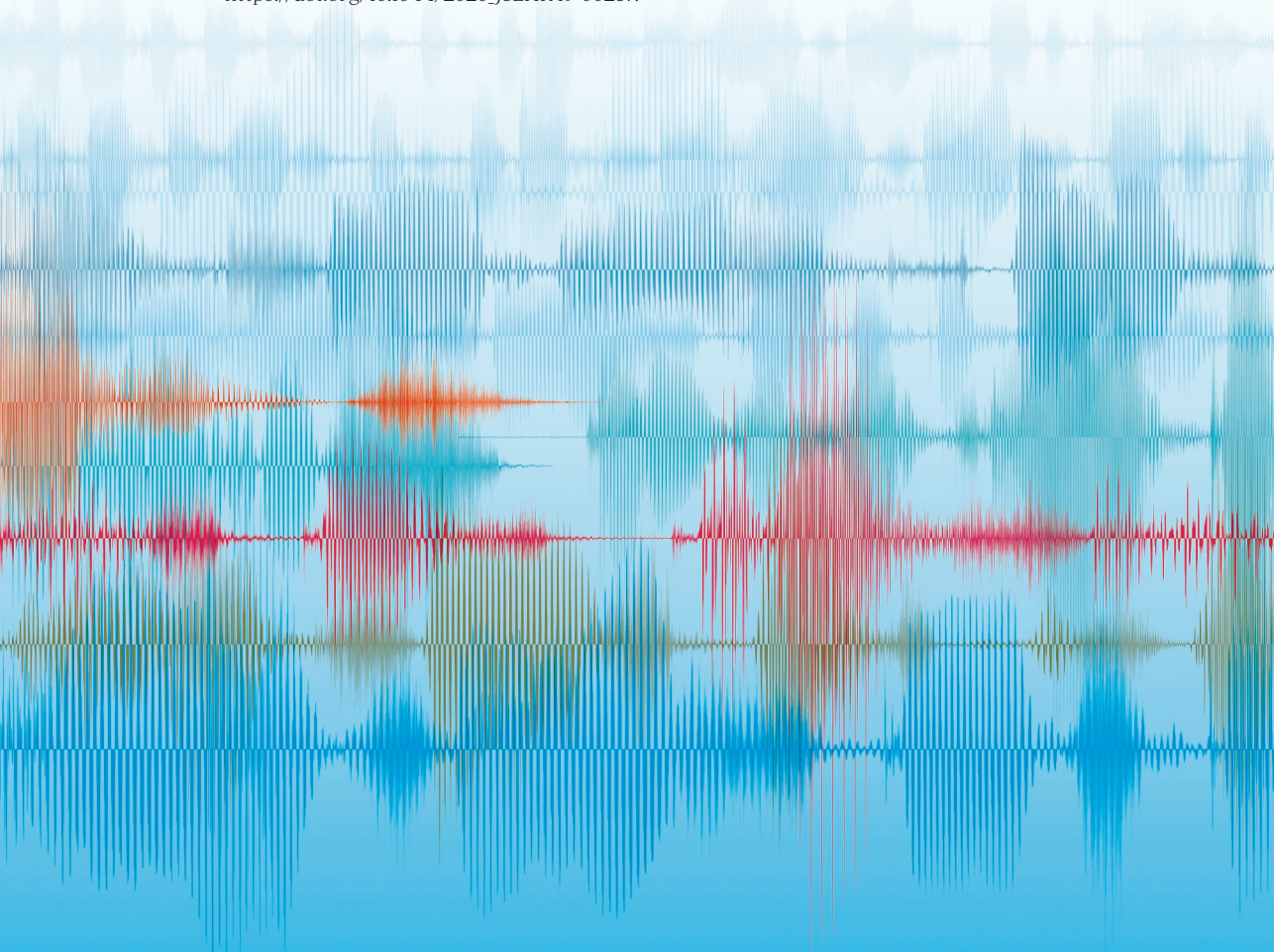


3

Maximum Speech Performance and Executive Control in Young Adult Speakers

This chapter is based on the following:

Shen C., & Janse E. (2020). Maximum speech performance and executive control in young adult speakers. *Journal of Speech, Language, and Hearing Research, 63*(11), 3611-3627. https://doi.org/10.1044/2020_JSLHR-19-00257.



Abstract

This study investigated whether maximum speech performance, more specifically, the ability to rapidly alternate between similar syllables during speech production, is associated with executive control abilities in a nonclinical young adult population.

Seventy-eight young adult participants completed two speech tasks, both operationalised as maximum performance tasks, to index their articulatory control: a diadochokinetic (DDK) task with non-word and real-word syllable sequences and a tongue-twister task. Additionally, participants completed three cognitive tasks, each covering one element of executive control (a Flanker interference task to index inhibitory control, a letter-number switching task to index cognitive switching, and an operation span task to index updating of working memory). Linear mixed-effects models were fitted to investigate how well maximum speech performance measures can be predicted by elements of executive control.

Participants' cognitive switching ability was associated with their accuracy in both the DDK and tongue-twister speech tasks. Additionally, non-word DDK accuracy was more strongly associated with executive control than real word DDK accuracy (which has to be interpreted with caution). None of the executive control abilities related to the maximum rates at which participants performed the two speech tasks.

These results underscore the association between maximum speech performance and executive control (cognitive switching in particular).

3.1. Introduction

Adult speakers have years of experience speaking, yet they often stumble over sentences such as 'she sells sea-shells by the seashore', where constant alternation between /s/ and /ʃ/ at word onsets is needed. What kind of control abilities is required from speakers to successfully produce the alternations in such 'tongue-twisting' sentences? Recent clinical studies have suggested that articulatory control abilities may relate to executive control abilities (e.g., Dromey & Benson, 2003; Nijland et al., 2015). Some psycholinguistic studies, on the other hand, have argued that stages of speech production following lexical selection, such as articulation (covering phonetic encoding, motor programming, and articulatory execution), do not require processing resources (e.g., Ferreira & Pashler, 2002). Our study will take an individual differences approach to investigate whether executive control abilities predict the ability to rapidly alternate between similar syllables during speech production. We investigated individual differences in maximum speech performance in a nonclinical population of young adult speakers. The choice of this population enabled us to include a relatively large group of participants, as individual differences research should preferably be carried out with large samples. Moreover, even in relatively homogeneous student populations, language performance has been demonstrated to be variable enough to show relationships between cognitive control and lexical access (e.g., Piai & Roelofs, 2013).

3.1.1. Studies on articulatory control

The terms articulatory control and speech motor control are often used interchangeably to refer to the "systems and strategies that regulate the production of speech, including the planning and preparation of movements and the execution of movement plans to result in muscle contractions and structural displacements" (Kent, 2000, p. 391). Articulatory control in clinical settings is often quantified by various maximum performance speech tasks in the assessment of motor speech disorders (Kent et al., 1987). Among those maximum performance speech tasks, rapid repetition rate or the diadochokinetic (DDK) rate task has been one of the most commonly used tasks. It is relatively simple to conduct and administer, and speakers' performance on this task has been claimed to be a stable index of oral motor skills (Bernthal et al., 2009; Duffy, 2013; Fletcher, 1972; Kent et al., 1987).

In a DDK task, participants are typically asked to accurately and rapidly repeat the same nonsense syllables (e.g., pa, pa, pa) or to alternate between different nonsense syllables (e.g., pa, ta, ka) (Duffy, 2013; Fletcher, 1972; Yang et al., 2011). Previous research has investigated differences in maximum performance for repetition of nonsense sequences compared to repetition of real words. As speakers have more experience speaking real words and hence have access to stored motor programmes for words,

but not for non-words, they can be expected to reach faster rates for real word repetition than non-word repetition. Indeed, for Hebrew, school-age children (Icht & Ben-David, 2015) and healthy older adults (Ben-David & Icht, 2017) achieved faster repetition rates in producing the real (familiar) Hebrew word *bodeket* relative to the trisyllabic non-word 'pataka' (note, however, that lexical status is confounded with voicing of plosives here, which may also influence rate differences in the two stimulus types). Additionally, for languages with lexical stress, real words, but not non-words, have fixed stress patterns that may lead to reduction of unstressed syllables. This would also lead to potentially faster rates for real word than non-word repetitions.

This effect of lexical status also brings up the much-debated question of how representative DDK maximum performance based on non-word repetitions is for patients' speech performance (Maas, 2017). Ziegler and colleagues (cf. also Staiger et al., 2017) have argued that motor requirements for 'nonspeech' (i.e., DDK) behaviour may differ from those for natural speech. Ziegler (2002), for instance, found that patient groups who had comparable sentence production rates differed significantly in their (non-word) DDK rates. Moreover, whereas one pathology (apraxia of speech) might affect sentence production more than DDK, another pathology (cerebellar dysarthria) would affect DDK performance more than sentence production. Possibly, speech tasks differ in their involvement of executive control. Performance on tasks involving articulation of less familiar (nonsensical) sequences may be more variable and more vulnerable to differences (within and between speakers) than production of familiar phrases or words. Repetition of unfamiliar (such as DDK) sequences may therefore be expected to involve more executive control than repetition of familiar sequences.

Kent (2004) illustrated that speech, as a motor behaviour, is influenced by cognition and that speaking should be viewed as a "cognitive-motor accomplishment" (Kent, 2004, p. 3). Thus, in order to successfully complete the stages of speech production, a certain amount of executive control may be required from speakers. Executive control (or executive functions) is known as a set of general-purpose control mechanisms that regulate our thoughts and actions (Gilbert & Burgess, 2008; Logan, 1985). Executive control is proposed to have three main underlying components, namely, inhibitory control, cognitive switching, and updating of working memory (Miyake et al., 2000). More specifically, inhibitory control is the ability to suppress activation of unwanted information in order to resolve conflict. Cognitive switching is defined as the ability to rapidly switch back and forth between mental sets or operations. Lastly, updating of working memory refers to the ability of maintaining or actively refreshing the contents of working memory while processing incoming information (Miyake et al., 2000).

In a review article, Kent (2000) suggested motor speech disorders should be investigated in relation to (phonological and) cognitive systems. The question of whether articulatory control may be related to executive control has been investigated in different clinical populations. In children with childhood apraxia of speech (CAS), a relationship between memory abilities and speech production has been observed (Nijland et al., 2015). Significant correlations were found between scores on two cognitive factors (extracted from a set of complex sensorimotor and sequential memory tasks) and speech scores (based on maximum repetition rates of non-speech stimuli such as the trisyllabic, 'pataka'-type, maximum repetition task) of children with CAS (Nijland et al., 2015). Similar associations between cognitive and speech performance were found in a study testing adults with dyslexia and adults with a probable history of CAS (Peter et al., 2018). Peter et al. (2018) used a battery of speech tasks (non-word repetition, multisyllabic real-word repetition, and non-word decoding), testing for patients' sensory encoding, memory, retrieval, and motor planning/programming abilities. Their results showed that the two disordered groups performed significantly worse on all three speech tasks compared to adults from the control group, again suggesting links between sensory encoding, (short-term) memory, and speech motor programming (Peter et al., 2018).

Perhaps more direct evidence for a relationship between cognition and speech motor performance has been found among nonclinical populations in studies where cognitive load was manipulated experimentally. For instance, using kinematic measures of lip movement, Dromey and Benson (2003) found healthy young adults' speech production to be more variable in a sentence repetition task when repetition was paired with cognitive or linguistic distractors (i.e., a higher cognitive load), relative to simple repetition. Similar results were obtained in follow-up studies, for instance, Bailey and Dromey (2015) on effects of dual tasking on speech motor performance of younger, middle-age, and older adults and MacPherson (2019) on increased cognitive load effects, as induced by Stroop interference, on speech motor performance in healthy younger and older adult speakers.

Results of these studies on clinical and nonclinical populations, therefore, suggest a relationship between cognitive and articulatory control. Nevertheless, several psycholinguistic studies, to be reviewed below, have argued that the involvement of executive control in 'late' stages of speech production, such as phonological encoding and articulation, is minimal.

3.1.2. Psycholinguistic studies

There is now ample evidence for a relationship between executive control and formulation and lemma selection stages (or the 'early' stages) of speech production, such as in language control in bilinguals (e.g., Costa et al., 2006; Rodriguez-Fornells et al., 2006), in noun-phrase production (e.g., Sikora et al., 2016), and in word-level lemma selection (e.g., Piai & Roelofs, 2013; Shao et al., 2012). The relationship between executive control and 'late' stages of speech production, such as phonological encoding and articulation, however, remains less straightforward. Garrod and Pickering (2007) argued that the processes of syllable or phoneme selection and articulation are largely automatic comparing to, for instance, the process of lexical selection. Their claim was supported by experimental evidence by Ferreira and Pashler (2002), who used a dual-task paradigm to test participants' performance on a picture-naming and a concurrent manual tone discrimination task. Ferreira and Pashler manipulated the availability of processing resources for lemma selection, phonological word form selection, and phoneme selection by introducing a secondary task. They found that both lemma retrieval and morphological encoding delayed the latencies of the secondary tone discrimination task, while phonological encoding did not show such interference. Their argumentation, based on these results, was that phoneme selection did not require central processing resources (Ferreira & Pashler, 2002).

Roelofs (2008) followed up on these results and examined dual-task interference using a slightly different paradigm than Ferreira and Pashler (2002). Roelofs found that phonological encoding on picture naming did spill over to performance on an unrelated manual task. This result thus suggests that some form of executive control may be required for phonological encoding as well. Additionally, in a more recent experimental study also using a dual-task paradigm, Jongman et al. (2015) tested whether sustained attention (related to executive control) is consistently needed throughout the different stages of speech production. Their evidence suggests that individual differences in sustained attention were mainly related to the processes of phonetic encoding and initiation of articulation (Jongman et al., 2015).

Evidence for selection and resisting interference from competitors at the level of phonological and phonetic encoding comes from studies using the tongue-twister paradigm (Wilshire, 1999). For instance, using tongue-twister–like utterances, McMillan and Corley (2010) manipulated the phonemic similarity of onset consonants and compared speakers' production of word sets with and without phonemic competition (e.g., *kef def def kef* vs. *kef kef kef kef*). Their results showed that articulation of onset phonemes in tongue-twister–like word sets is influenced by competing phonemes, such that even if speakers do not produce full-blown errors, their productions are less target-like and more variable in the context of competing

phonemes (*kef def def kef*) than when produced in a context without competing phonemes (*kef kef kef kef*). Furthermore, their results also showed that the more similar the competing phoneme to the target phoneme (/t/ being more of a competitor for /k/ than is /d/), the greater the effect of articulatory interference (McMillan & Corley, 2010). These results suggest higher level executive control may be needed during the 'late' stages of tongue-twister production to resist interference in order to select the correct target phoneme. In summary, despite the mixed findings listed above, some psycholinguistic studies are in line with the speech kinematics evidence by Bailey and Dromey (2015) and Dromey and Benson (2003), that executive control may be involved during the 'late' stages of speech production (phonological encoding and articulation), thereby challenging the claims that the 'late' stages of speech production are largely automatic (Ferreira & Pashler, 2002; Garrod & Pickering, 2007).

3.1.3. This study

We now return to our initial question of what control abilities are required for speakers to successfully produce 'tongue-twisting' sentences that contain constant alternations between similar syllables, whereby we focus on the three elements of executive control (inhibition, shifting, and updating of working memory) in the Miyake model (Miyake & Friedman, 2012). Note that there are multiple models of cognitive abilities, working memory, or attentional abilities (e.g., Baddeley & Della Salla, 1996; Posner & Peterson, 1990) and that different models have distinguished different elements. For this study, we chose to investigate the link between speech performance and the three executive control elements defined in the Miyake model.

The interference induced by phoneme similarity in the McMillan and Corley (2010) study suggests that phonologically similar phonemes are jointly activated due to shared features and that similar phonemes compete for selection. As resolving competition at lexical selection has been linked to executive control (Piai & Roelofs, 2013), we hypothesise that higher level executive control may be needed during phonological encoding or the 'late' stages of tongue-twister production to resist interference and to select the correct target phoneme. More specifically, in order to accurately and fluently produce tongue-twister phrases or sentences, inhibitory control may be involved in the suppression of coactivated but incorrect competing phonemes and/or phoneme clusters. Additionally, speakers producing tongue-twister phrases typically need to switch between two or more similar competing onset phonemes, between similar onset clusters, or between singleton onset phonemes and onset clusters. As such, we hypothesise that production of alternating sequences, such as tongue twisters, may require cognitive switching. Furthermore, in line with evidence that speech performance is associated with sequential memory functioning

in children with CAS (Nijland et al., 2015), we investigate whether updating ability relates to tongue-twister performance as speakers need to constantly update the planning and programming of the required speech movements during production.

Similar to the tongue-twister paradigm, the maximum performance speech task that we discussed earlier, the DDK task, also contains several elements that may require executive control. For instance, in order to repetitively produce the DDK sequence 'pataka', the amount of shared phonetic features in the syllable-initial consonants may require speakers to suppress the coactivated but incorrect phoneme (cf. McMillan & Corley, 2010). Additionally, fast alternation between the similar syllable-onset phonemes requires that speakers constantly switch between them. Lastly, the involvement of updating ability could be reflected in having to constantly update the planning and programming of familiar or unfamiliar sequences during speech production.

Note that our tongue-twister and DDK speech tasks are maximum performance tasks in which maximum speed is stressed. Therefore, we also investigate whether maximum performance on the two speech tasks (i.e., accuracy and rate) relates to the general ability of information-processing speed.

Clearly, the two maximum performance speech tasks of tongue twisters and DDK have typically been used in separate research fields for different purposes. The tongue-twister paradigm has mainly been used in psycholinguistic studies as a means to elicit speech errors or blends, while the DDK task has typically been used in a clinical setting as an index of speech motor control. According to Levelt's model of speech production (Levelt, 1989; Levelt et al., 1999), tongue-twister errors and blends may occur at the level of phonological selection and/or at the level of phonetic encoding. DDK performance has been suggested to index speech motor ability, and hence, DDK performance concerns an even later stage than the phonological encoding stage involved in tongue twisters. However, despite their differences, both tasks may capture elements of speakers' articulatory control. In a recent study in which we administered both tasks as maximum performance tasks, we found a significant correlation between maximum performance on tongue-twister and DDK repetition (Shen & Janse, 2019). This finding suggests that these two tasks tap into a task-independent articulatory control component.

The current study was thus set up to investigate the potential link(s) between maximum speech performance and executive control abilities. More specifically, we examined whether cognitive measures of inhibitory control ability, cognitive switching ability, working memory capacity, and baseline processing speed predict

articulatory control as measured by DDK and tongue-twister (rate and accuracy) performance in a healthy young adult population. Finding out whether the late stages of speech production (phonological and phonetic encoding and execution) relate to cognitive control is important for (psycholinguistic or speech-motor) theories on speech production. Knowing about possible relationships between a clinical speech measure like DDK and executive control is also important for clinical practice, as it may have implications for DDK administration with patient populations suffering from cognitive impairment or comorbidities.

3.2. Methods

3.2.1. Participants

A total number of 78 participants (age: *M* = 23 years, *SD* = 3; 61 women) were recruited online through the Radboud Research Participation System (note that all of them were enrolled in bachelor's or master's programmes or had already graduated). Participants were all native Dutch speakers with normal or corrected-to-normal vision and had no reported history of speech, hearing, or reading disabilities nor past diagnosis of speech pathology or brain injury. Our study protocol was evaluated and approved by the Ethics Assessment Committee Humanities at Radboud University. Participants had all given informed consent for their data to be analysed anonymously, and they either received course credits or gift vouchers as compensation for their time.

3.2.2. General procedure

Participants were tested individually in the Centre for Language Studies Lab at Radboud University. They completed a battery of five tasks during the experimental session; three of which were cognitive tasks (a flanker interference task, a letter-number switching task, and an operation span task), and two were maximum performance speech tasks (a DDK task and a tongue-twister task). The whole session lasted for 60–75 min. During the experimental session, participants first completed the three cognitive tasks and then performed the two speech tasks. For the three cognitive tasks, Presentation software (Version 18.0, Neurobehavioral Systems, Inc.) was used to present the visual stimuli and to record participants' responses. For the two speech tasks, PowerPoint slides were used to present speech stimuli. One audio recording was made per participant using a Sennheiser ME 64 cardioid capsule microphone on an adjustable table stand. The speech was recorded through a preamplifier (Audi Ton) onto a steady-state 2 wave/mp3 recorder (Roland R-05). All tasks were completed in a sound-attenuating recording booth. All visual stimuli from the cognitive and speech tasks were presented on a Ben Q XL 2420T 24-in. full HD

monitor placed on a table in front of the participant. Participants were encouraged to sit comfortably to have a good view of the computer screen.

The experimenter monitored participants' performance in both the cognitive and speech tasks from outside the recording booth during practice trials. Whenever participants were confused or misunderstood the task requirements during the practice phase, the experimenter would verbally communicate with the participant and restart the practice to make sure all participants had sufficient understanding of the task(s). The progress of the cognitive tasks and the presentation of stimulus slides for the speech tasks were controlled by the experimenter on the stimulus computer (Dell Precision T3600).

3.2.3. Cognitive tasks

The three cognitive tasks used in this study were each meant to tap into one aspect of executive control: a flanker task was used to index inhibitory control, a letter-number task was used to index switching ability, and an operation span task was used to index working memory capacity. The three tasks are described in more detail below.

Flanker task

Task description

The flanker task, developed by Eriksen and Eriksen (1974), measures inhibition of dominant (flanking) stimuli. During the task, participants were presented with a sequence of five symbols, and they were asked to pay attention to the direction in which the middle symbol (an arrowhead '<' or '>') was pointing. They had to respond to the target (middle) stimulus by pressing a response button with their left thumb or index finger when the stimulus was pointing left ('<') or with their right thumb or index finger when it was pointing right ('>'). The two target response buttons on the six-button button box were labelled with '<' on the left-hand side and '>' on the right to clarify the association between the target stimulus and the response buttons.

The target stimulus appeared in three conditions, namely, the congruent condition (target stimulus pointing in the same direction as the flanker stimuli, '<<<<<' or '>>>>>'), the incongruent condition (target stimulus pointing in the opposite direction of the flanker stimuli, '<<><<' or '>><>>'), and the neutral condition (target stimulus embedded in the middle of neutral stimuli, '––<––' or '––>––'). In total, 72 trials were presented with an equally distributed number of repetitions across the three conditions (24 trials per condition, of which 12 targets were pointing left and 12 were pointing right). The order of the 72 test trials was randomised for each participant.

On-screen instructions in Dutch were given at the beginning of the task, followed by 12 practice trials to familiarise participants with the task. On each trial, a fixation cross was presented for 750 ms, followed by a target stimulus for 500 ms. A 1,000-ms blank screen was presented immediately after the target stimulus for participants to respond (timing choices were piloted with a small sample of different younger adults to verify that the task was doable yet challenging). Participants were encouraged to respond as quickly and as accurately as possible, and any response exceeding the response duration was logged as a 'miss'. After 12 practice trials, a wait screen was presented, asking whether the participant had understood the task correctly and was ready to begin. Once a ready signal was received from the participant, the experimenter proceeded the task on the main computer outside the recording booth.

Analysis

Participants' response times (RTs) and response accuracy were measured. We only calculated individual RT means for those participants who had actually paid attention to the stimulus on screen, as evident from accuracy levels well above chance. Data of seven participants had to be excluded because they failed to meet our minimum accuracy requirement, that is, having an accuracy level of at least 2/3 correct responses overall (i.e., minimally 48 correct out of 72) and 2/3 correct responses in each individual condition (i.e., 16 correct out of the 24 trials per condition). Overall accuracy of the remaining 72 participants ranged between 86% and 100%. RT data of the remaining 72 participants (correct trials only) were analysed using RStudio (Version 1.1.463), the R packages languageR (Version 1.4.1; Baayen, 2013), and lme4 (Version 1.1-19; Bates et al., 2015). Data points that were more than 3 *SD*s of the individual's overall mean were removed (47 data points in total or < 1%). Mean RT was 362 ms (*SD* = 73) for the congruent condition and 451 ms (*SD* = 73) for the incongruent condition. To examine whether there is a potential trade-off between speed and accuracy on this task, we correlated individual accuracy and overall RT. Speed and accuracy were not correlated ($r = .18$, $p > 0.1$).

RTs (from valid responses only) were log-transformed (to make the distribution more normal) and entered as a numerical dependent variable into a linear mixed-effects model. Condition (congruent, incongruent, or neutral, with the congruent condition mapped on the intercept) of the flanker trials was entered as the fixed effect of interest, with direction (pointing direction of the target arrow) and trial being included as fixed control predictors. By including the latter two variables in the statistical model, we can account for variance that is otherwise left unexplained (participants generally speeding up over trials and participants being generally faster on arrows pointing to the right than pointing to the left). Participant was included as random effect (Baayen et al., 2008), with condition being a random

by-participant slope to capture individual variability among participants in the size of the condition effect. Across participants, RTs were longer going from the congruent to the incongruent condition (reflecting the general condition effect). The by-participant slopes reflected the modelled individual adjustment to this general slowing effect. To make the interpretation of these slopes more straightforward, we reversed the individual slopes (negative values made positive and vice versa). In this way, participants with an originally negative value of this by-participant condition adjustment (i.e., those who were less slowed, relative to the averaged condition effect, changing from congruent to incongruent flanker trials) now got a positive value, indicating better inhibitory control. Conversely, participants who originally had a positive value, indicating that they were slowed more than average, now got a negative value, indicating worse inhibitory control.

Letter-number task

Task description

The task-switching paradigm, first introduced by Jersild (1927) and then popularised by Rogers and Monsell (1995), has mainly been used to measure the 'switching cost' incurred during switching back and forth between different trials or sets of trials. During this letter-number task, participants were presented with letter-number combinations (e.g., C8). They were instructed to pay attention to the quality of the number being even or odd (2, 4, 6, and 8 for even; 3, 5, 7, and 9 for odd) or to the case of the letter being upper or lower (a, d, f, and h for lower case; B, C, E, and G for upper case) in the letter-number combinations. The task consisted of three blocks.

During the entire task, the experiment-monitor screen was divided into four equal quadrants by a graphic cross. In Block 1, letter-number combinations only showed up in the top two quadrants, with stimulus location changing following a left-to-right manner from trial to trial. Participants were asked to only pay attention to the number in the letter-number combination and judge whether the number was even or odd by pressing the buttons labelled with the Dutch word 'Even (*even*)' or 'Oneven (*odd*)' on the button box. In Block 2, only the bottom two quadrants of the computer screen were used. Stimulus location also followed a left-to-right manner from trial to trial. Participants were instructed to only pay attention to the letter in the letter-number combination and judge whether the letter case was capital or small by pressing the buttons labelled with 'Hoofd (*capital*)' or 'Klein (*small*)'. Note that only two buttons on the button box were used for this task with top halves of the buttons labelled with 'Even' and 'Oneven' and lower halves with 'Hoofd' and 'Klein'. This was to ensure stimulus-response mapping: left index finger/thumb for even and capital stimuli and right index finger/thumb for odd or small stimuli (Rogers & Monsell, 1995).

The first two blocks were single-task blocks in which participants had to either pay attention to the number being even or odd (Block 1) or to the letter being in lower or upper case (Block 2). The third block was a mixed-task block, in which the position of the letter-number combination on the screen (i.e., the quadrant the combination appeared in) determined what aspect of the letter-number combinations participants had to pay attention to. In total, there were 192 trials; Blocks 1 and 2 both consisted of 48 trials, and Block 3 consisted of 96 trials.

In Block 3, the mixed-task block, the whole screen was used for the presentation of letter-number combinations. Stimulus location changed following a clockwise manner from trial to trial (starting in the upper left quadrant, then upper right, followed by lower right, then lower left). Participants were required to judge the number of the letter-number combination as being odd or even if the letter-number stimuli were presented in the upper left and right quadrants and to judge the letter of the letter-number combination as being upper or lower case if the stimuli were presented in the lower quadrants. Each letter-number stimulus was presented until the participant pressed one of the response buttons, up to a maximum of 5,000 ms. The third block thus consisted of no-switch trials where participants had to pay attention to the aspect they also paid attention to on the previous trial (i.e., the no-switch trials appearing in the upper right quadrant and the lower left quadrant) and switch trials where participants needed to switch from responding to the one dimension to the other dimension (i.e., the switch trials appearing in the lower right quadrant and the upper left quadrant). Blocks 1 and 2 were practice blocks, while Block 3 was the experiment block of interest.

Participants were instructed to respond as quickly and as accurately as possible, and any response exceeding maximum trial duration was logged as a 'miss'. Instructions in Dutch were displayed on screen prior to each block of trials. Upon reading the instructions of each block, participants were asked whether they had any questions understanding the task. Once everything was clear, the experimenter proceeded the task on the main computer outside the recording booth. After each block, participants were presented with visual on-screen feedback on their accuracy score for that block. This block-based feedback enabled the experimenter to evaluate whether participants had sufficient understanding of the task requirements during the first two (practice) blocks before they moved on to the third (test) block.

Analysis

Participants' RTs and response accuracy in the third (mixed-task) block were measured. One participant's data were excluded due to technical failure. Data of all remaining 77 participants met the minimum accuracy requirement, that is, each

participant having at least two-thirds of correct responses in the third block (64 correct out of 96 trials). Accuracy rates in the third block ranged between 88% and 100%. Data of these 77 participants were analysed using RStudio (in the same way as described above for the flanker data analysis). Similar to the flanker task, 106 outliers (104 data points were more than 3 *SD*s of the individual's mean RTs in Block 3, and two data points were lower than the 200 ms threshold) of the RT data were removed (< 1.5%). Mean RT was 794 ms (*SD* = 454) for the no-switch trials and 1,353 ms (*SD* = 621) for the switch trials. We correlated individuals' response speed and accuracy on this task to test for potential trade-offs. Individual RT and accuracy were not significantly correlated (r = -.11, p > 0.1).

Similar to the flanker task, RTs (from correct responses only) were log-transformed (to make the distribution more normal) and entered as a numerical dependent variable into a linear mixed-effects model. Condition (the target trial being a switch or no-switch trial, with the no-switch condition being mapped on the intercept) of the letter-number trials was entered as the fixed effect of interest, with trial being included as a fixed control predictor. Participant was included as random effect, with condition as a random by-participant slope to capture individual variability among participants in the size of the condition effect. The general condition effect showed that participants' RTs generally increased going from a no-switch to a switch trial. The by-participant slopes reflected the modelled individual adjustment to this general switching effect, such that the lower the value, the less they were slowed, changing from no-switch to switch letter-number trials (relative to the averaged condition effect), indicating a smaller switching cost. Similar to the analysis of flanker responses above, we also reversed the individual slopes here (negative values made positive and vice versa). Thus, those with original lower values for this individual condition adjustment (i.e., those with lower negative values) were slowed less than average, changing from no-switch to switch trials, now got a positive value, indicating better switching ability.

Operation span task

Task description

The operation span task (Turner & Engle, 1989), as one of the complex span tasks, is taken to assess the capacity to efficiently update working memory. The task requires participants to store and regularly update memory representations while performing another cognitively demanding task. For example, in the original version of the task, participants have to solve simple mathematical problems while memorising word lists of varying lengths. The adapted version of the operation span task used in this study (from Shao et al., 2012) required participants to judge the accuracy of simple mathematical problems while remembering randomly ordered letter lists of varying

length. The main reasons for using letters rather than words, as used in Shao et al. (2012), are twofold. First, we intended to increase the difficulty of the task by replacing meaningful words with meaningless, randomly sequenced letters, such that the letter lists did not resemble any familiar Dutch or English acronyms. Second, we aimed to test 'purer' executive control by avoiding interference from language ability as much as possible.

For the task, 65 mathematical operations each followed by one letter (letters were selected from the alphabet) were used as trials. These 65 trials were divided over 17 lists, ranging from two to six trials per list. Two lists of two and three trials, respectively, were used as practices lists. Detailed instructions were given on screen before the practice lists. During the task, a fixation cross was presented for 800 ms at the start of each trial. After a blank screen of 100 ms, a mathematical operation followed by a letter was presented in the centre of the screen, for example, (4 × 2) - 3 = 2 D. Participants were instructed to read both the operation and the letter out loud in the order presented and then press one of the buttons labeled 'Ja (yes)' or 'Nee (no)' on the button box to judge whether or not the operation was correct while trying to remember the letter. At the end of each list of trials, a recall cue: 'Nu graag typen! (Type now please!)' was presented. Upon presentation of this cue, participants were asked to recall all the letters seen since the beginning of the list and to type them in the same order as they had been presented using a keyboard. They were also encouraged to mark the position of any missing letters using '.' if they could not recall the letters themselves. The experimenter monitored participants' performance during the practice trials. Participants were reminded to read the mathematical operation and the letter following it out loud if they forgot to do so.

Analysis

Participants' response accuracy for the mathematical operations and their scores for the letter sequence recall were measured. Results from two participants were excluded because of poor performance on the math problems (less than 85% correct, following Unsworth et al., 2005). Updating span (recall score) was calculated as the sum of the letters that were recalled correctly in the correct position (Unsworth et al.,2005). The higher the recall score, the better the working memory capacity. The range of a possible score is between 0 and 60. Participants' mean task performance (number of letters correctly recalled) was 38 (SD = 11), and their actual scores ranged between 18 and 60 (i.e., between 30% and 100%).

Processing speed

In order to obtain an index of individual participants' processing speed, rather than introducing a new task, we made use of the 'control' trials from the two cognitive

tasks where speed was a built-in task requirement (i.e., the flanker and letter-number tasks). We used a principal component analysis to derive one single speed construct underlying the baseline speed measures from the two tasks. More specifically, this single speed construct was derived from the individual random intercepts in the two speeded tasks (for the congruent condition mapped on the intercept in the flanker task and for the no-switch trials mapped on the intercept in the letter-number task). Factor loadings on the processing speed construct (unrotated factor solution) were 0.83 for both speed measures. Because this measure is based on the two baseline speed measures from the two cognitive tasks (where those who are faster have shorter RTs and hence negative by-participant intercepts), the values of this speed construct were also reversed (i.e., higher values of this speed construct indicate faster processing speed) for a more straightforward interpretation (higher values reflecting 'better' performance).

3.2.4. Speech tasks

The two speech tasks, a DDK task and a tongue-twister task, were set up as maximum performance speech tasks to capture participants' articulatory control ability. In order to provide a more complete picture of speakers' articulatory control ability, Yaruss and Logan (2002) proposed to focus not just on maximum (DDK) rate to quantify (children's) speaking abilities but to also investigate other aspects of DDK performance, such as accuracy. Therefore, we quantified speakers' maximum performance through both rate and accuracy.

DDK task

Task description

A DDK task often contains repetitions of mono- or trisyllabic nonsense words like 'pa' and 'pataka' (Bernthal et al., 2009). Due to the focus of the current study on carrying out alternations, we opted for the sequential motion rate variant of the DDK task, that is, using alternating trisyllabic sequences as task stimuli (e.g., 'pataka'). We made adjustments to the canonical oral DDK task to link to the debate of whether non-word oral DDK is representative of speakers' actual speaking capability (Ben-David & Icht, 2017; Icht & Ben-David, 2015; Maas, 2017; Ziegler, 2002). We therefore specifically included two non-word DDK stimuli, the standard 'pataka' /pataka/ and the reverse-order 'katapa' /katapa/, and two real word DDK stimuli, namely, the two Dutch words that are closest to the non-word sequences: 'pakketten' /pɑˈkɛtə(n)/ (*packages*) and 'kapotte' /kaˈpɔtə/ (*broken*). This allowed us to test whether either type of DDK performance is more strongly associated with executive control. Note that, even though the selected real words were close to the nonsense words in terms of alternating consonants, they also differed from them in multiple respects. For instance, no stress pattern was available for the non-word stimuli, whereas both real

words had lexical stress on the second syllable. Moreover, vowels were full /a/ vowels in the nonsense sequences but were different vowels (different in length and place of articulation and in terms of acoustic reduction due to lexical stress) in the real words.

During the DDK task, each stimulus was always presented in the centre of a full-screen PowerPoint slide. Multiple repetitions of the (nonsense) words were presented in a row, for instance, 'patakapatakapataka...', to elicit repetitive production of the stimulus. Participants were instructed to repeatedly produce the presented stimulus as accurately and as rapidly as possible. A pre-recorded example was played prior to the practice trials to familiarise the participants with the task. A brief line of text reminding them about accuracy and speed of repetition was constantly on display at the top of each slide. A 2 s pause (preparation time) followed by a 75 ms beep tone was used to mark the start of articulation, and each stimulus was to be repeated for around 10 seconds. Additionally, the mono- and disyllabic nonsense stimuli ('pa', 'ta', 'ka', 'pata', 'taka') were presented to participants as practice trials before the experimental trials, such that participants had received extensive task familiarisation, including familiarisation of production of alternating sequences before they moved to the test phase. All DDK trials were presented to the participants in the same fixed order (i.e., practice trials followed by non-word and then by real-word sequences). Note that this implies that we cannot rule out that the fixed order may have contributed to performance differences between non-word and real-word sequences, to which we will come back in the Discussion section below.

Analysis

Maximum performance in terms of rate and accuracy was analysed acoustically in Praat (Boersma & Weenink, 2017). DDK articulation rate (syllables/s) and accuracy (fraction correct) were calculated using the first 7 s time window of the DDK utterance. This 7 s time window was selected because, even though articulation errors and disfluencies already occurred in a 3 s time window for most participants, the number and frequency of errors and disfluencies generally increased in longer time windows. Thus, in order to capture accuracy and articulation rate in a more reliable way, we opted for a relatively long time-window (7 s).

Individual DDK accuracy (fraction correct) was calculated as number of accurate and fluent repetitions divided by number of all repetitions in the 7 s time window (or as close to seven seconds as possible for the repetition counts to be an integer). A repetition was only counted as correct if it did not contain any form of obvious articulation errors (e.g., if a speaker produced 'patapka' or 'katakpa') or disfluencies (e.g., silent pauses longer than 200 ms) within the sequence (Yaruss & Logan, 2002). Individual DDK articulation rate (syllables/s) was calculated by multiplying the total number of

accurate and fluent (non)word repetitions produced by each participant in the same 7 s time window by three (syllables) and divided this number of total syllables by the actual production time (total duration minus erroneous and disfluent repetitions, as well as in-breaths and pauses longer than 200 ms between repetitions).

Tongue-twister task

Task description

Following Wilshire's (1999) tongue-twister paradigm, we selected four Dutch tongue-twister sentences containing a combination of repetition and alternation of word-initial consonants or consonant clusters (e.g., **p**oes **k**otst **p**ostzak, and **fr**its **v**indt **v**is**fr**ietjes). Below are the four tongue-twister sentences that were used as test stimuli with their literal English translations in parentheses (note that the boldface used in the tongue-twister sentences below is only for illustration purpose; the actual stimuli in the task did not have boldface on the similar/contrasting phonemes):

- De **p**oes **k**otst in de **p**ostzak (*The cat puked in the mail bag*)
- **Fr**its **v**indt **v**is**fr**ietjes **vr**eselijk **vi**es (*Frits finds fish-fries terribly gross*)
- Ik **b**ak een **pl**ak **b**ak**bl**oedworst (*I fry a slice of blood-sausage*)
- **P**apa **p**akt de **bl**auwe **pl**atte **b**ak**p**an (*Daddy grabs the blue flat frying pan*)

Prior to the task stimuli, two additional tongue twister sentences were presented as practice stimuli:

- **Sl**imme **Sj**aantje **sl**oeg de **sl**ome **sl**ager (*Smart Sjaantje hit the slow butcher*)
- **B**akker **B**as **b**akt de **b**olle **br**oodjes **br**uin (*Baker Bas bakes the round buns brown*)

Participants were instructed to repeat the tongue-twister sentences as accurately and as rapidly as possible. Similar to the DDK task, each tongue-twister stimulus was also always presented in the centre of a full-screen PowerPoint slide, with a brief line of text reminding participants about accuracy and speed of repetition at the top of each slide. A picture related to one object per tongue-twister sentence (e.g., a blue frying pan) was shown below the printed stimulus on the same slide, and then the picture disappeared after about 2 s of preparation time. Participants were instructed to start repeating the tongue twisters minimally five times as soon as the picture disappeared (note that the picture disappearing only served as a cue to start speaking, whereas the sentence remained on the screen).

Analysis

Participants' maximum performance in terms of accuracy and rate was analysed acoustically in Praat (Boersma & Weenink, 2017). Similar to the accuracy measures in

the DDK task, individual tongue-twister accuracy (fraction correct) was calculated as the number of accurate and fluent repetitions divided by the first five repetitions (five being the number of repetitions speakers minimally produced). A repetition was counted as accurate and fluent if it did not contain any form of perceivable error or disfluency (including silent pauses longer than 200 ms). Tongue-twister articulation rate (syllables/s) was calculated by averaging the articulation rate of the accurate and fluent repetitions of the four tongue-twister stimulus sentences (except for two participants whose overall rate was based on three tongue-twister sentences because they each had an accuracy of '0' in the remaining sentence; in other words, all five repetitions of one of the tongue-twister sentences contained errors). The rate of each accurate and fluent stimulus was measured by dividing the number of syllables in a tongue-twister sentence by the articulation time used for that repetition.

Relating executive control to maximum speech performance

Analysis

In order to investigate how well measures of articulatory control can be predicted by elements of executive control, we analysed our maximum speech performance data with linear mixed-effect regression models (as is the norm in psycholinguistic research). This choice enables us to account for random participant variance and any effects of our fixed predictors (such as cognitive ability indices and lexical status) on speech task performance. Several linear mixed-effects models were fitted for DDK performance (for DDK accuracy and rate separately) and for tongue-twister performance (again one model for accuracy and one for rate). Accuracy and rate were pooled per DDK or tongue-twister item (DDK items being the two real word and two non-word stimuli and tongue-twister items being the four-stimulus sentences), and these pooled item scores (fractions or pooled rates) were analysed as dependent variables.

In the two DDK models, DDK accuracy or rate was entered as numerical dependent variable, with the executive control scores as fixed effects of interest. These included the extracted individual scores of inhibitory control (derived from the by-participant slopes in the flanker task; scores scaled and centred), switching ability (derived from the by-participant slopes in the letter–number task; scores scaled and centred), working memory capacity (derived from the operation span task; scores scaled and centred), and processing speed (from the derived speed construct; scores scaled and centred). Lexicality was also included as a factor in the DDK models (real word vs. non-word stimuli) because we expected participants' DDK performance to differ across real words and non-words and because we wanted to investigate potential interactions between lexicality and cognitive abilities. Additionally, participant was included as a random effect in both DDK models, and we also allowed a random

by-participant slope for the lexicality effect, acknowledging that speakers may be differentially affected by the difference between real words and non-words. DDK item could not be entered as a fixed variable, as this would leave no variance to the model given the item-pooled dependent measure. These full models were then stripped in a stepwise manner to arrive at the most parsimonious model (taking out insignificant interactions, first, and then insignificant effects, starting with the ones with the lowest *t* values). Model comparisons were applied after each removal of the least significant predictor to verify that exclusion of each predictor term did not lead to a significantly different model fit.

Two tongue-twister models were set up as well (one for accuracy and one for rate) with pooled tongue-twister accuracy or rate as numerical dependent variable. Tongue-twister performance was also analysed as a function of the same four cognitive measures used in the two DDK models. Tongue-twister number (four in total) was included as a fixed control predictor (with the first sentence mapped on the intercept), and participant was included as random effect into the model. Similar to the DDK models, the full models were also stripped in a stepwise manner, with model comparisons applied after each removal of the least significant predictor, to arrive at the most parsimonious model.

3.3. Results

Descriptive performance in the two speech tasks in terms of rate (syllables/s) and accuracy (fraction correct) from 78 participants is illustrated in Figures 3.1 and 3.2 below. Rate and accuracy measures averaged over task stimuli were entered as dependent variables in the two linear mixed-effect models for rate and accuracy, respectively. Task (three levels: DDK real word, DDK non-word, and tongue twister) was entered as the fixed effect of interest, with participant as a random effect. As shown in Figures 3.1 and 3.2, maximum performance in the tongue twister task is significantly worse than in DDK non-word repetition (t = -25.17, p < .001 and t = -18.24, p < .001 for rate and accuracy, respectively). Within the DDK task, real word DDK performance is significantly better than non-word DDK performance for both rate (t = 5.45, p < .001) and accuracy (t = 2.71, p < .01).

Our DDK non-word data can be compared to previously established norms for a Dutch nonclinical speaker population (Knuijt et al., 2017). Median maximum repetition rate for 'pataka' in young adults aged 18-29 years in their study was 7.0 syllables/s (range: 4.1-9.0), whereas median performance in our sample was 6.1 (range: 3.8-8.7) for 'pataka' (and median of 5.6 for 'katapa', for which no reference value was

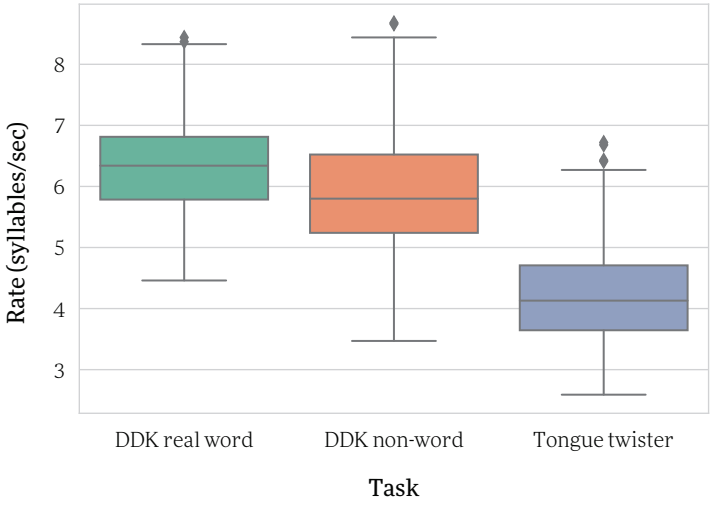


Figure 3.1
The distribution of maximum performance rate in DDK non-word, DDK real word, and tongue-twister tasks. DDK = diadochokinetic

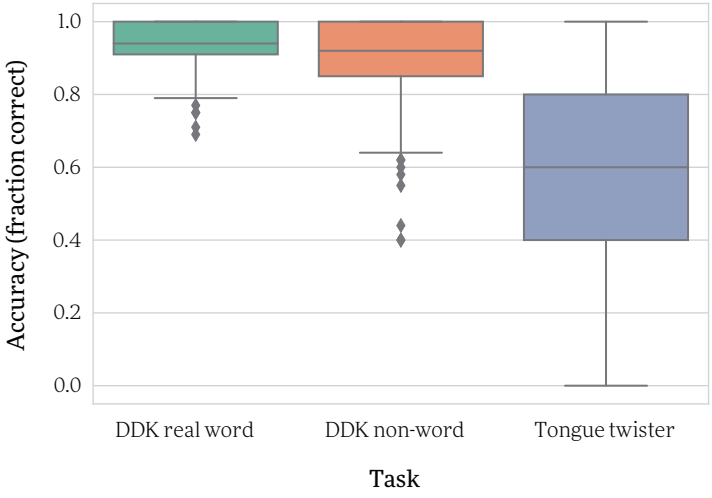


Figure 3.2
The distribution of maximum performance accuracy in DDK non-word, DDK real word, and tongue-twister tasks. DDK = diadochokinetic

available). Differences between samples may be due to differences in the way rate was calculated (in relation to errors and pauses).

We checked for potential speed-accuracy trade-offs in these speech tasks by examining whether rate and accuracy were correlated. Correlations between speech rate and accuracy were not significant for DDK real word (r = -.020, p > .05), DDK non-word (r = -.056, p > 0.1), and tongue twister (r = -.084, p > 0.1).

Before moving on to addressing the research question, we checked intercorrelations between cognitive predictors. Table 1 presents the correlation matrix for our measures of inhibitory control (i.e., flanker task), switching ability (i.e., letter-number task), updating ability (i.e., operation span), and processing speed (based on two speeded measures).

Table 3.1 Correlation matrix of the cognitive measures

	Flanker Inhibitory Control	Letter-number Switching	Operation Span (updating)
Letter-number Switching	0.019		
Operation Span (updating)	-0.112	0.323**	
Processing Speed	0.507***	-0.138	-0.198

Pearson correlation coefficients
*: p < 0.05; **: p < 0.01; ***: p < 0.001

The correlational data presented in Table 3.1 indicate that switching ability (as indexed by the letter–number task performance) was positively linked to updating ability (as indexed by operation span performance), r = .323, p < .01, such that those who are better at switching also have better updating ability. Processing speed is positively related to inhibitory control (as indexed by flanker task performance), such that those who have faster processing speed are also better at inhibiting irrelevant information (r = .507, p < .001). Additionally, updating ability (operation span) did not correlate with inhibitory control ability (flanker performance).

Table 3.2 and Figure 3.3 below summarise the association between executive control (from the most parsimonious model) and DDK accuracy and rate. Note that performance on non-word DDK sequences was mapped on the intercept.

Table 3.2 The coefficient estimates, standard errors, and significance levels of factors involved in diadochokinetic accuracy and rate

	Dependent variable:					
	Accuracy			Rate		
Predictors	Estimates	SE	p	Estimates	SE	p
Intercept	0.888	0.010	<0.001	5.909	0.105	<0.001
Lexical-yes	0.052	0.011	**<0.001**	0.423	0.064	**<0.001**
Letter-number Switching	0.213	0.076	**0.005**			
Lexical-yes: Letter-number Switching	-0.227	0.083	**0.006**			

Note. Effects and interaction that remain significant given an extra conservative alpha level (α = .0125) are shown in boldface.

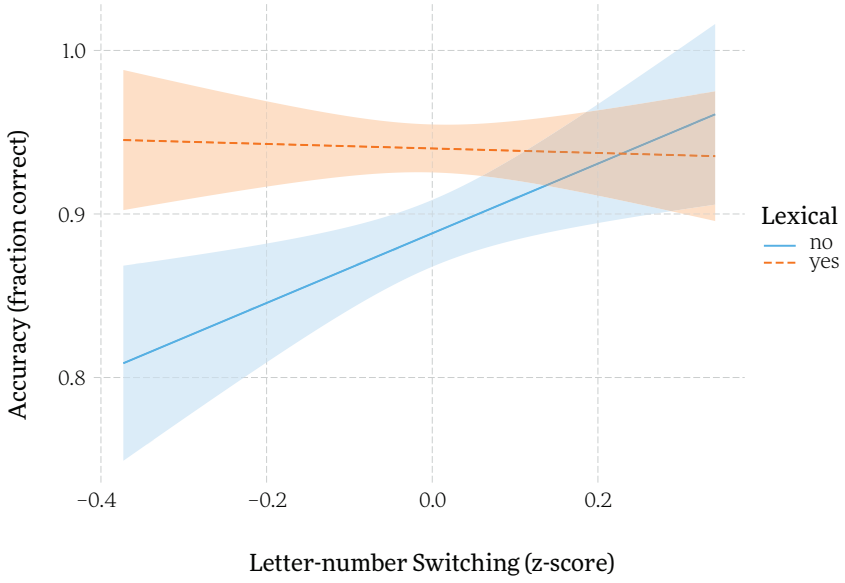


Figure 3.3

Model plot of DDK accuracy in relation to switching ability and lexicality (of the DDK sequences). DDK = diadochokinetic

Table 3.2 shows that DDK accuracy is significantly modulated by lexical status of the DDK stimulus, such that accuracy was higher for real word than non-word sequences (but keep in mind that lexical and non-lexical stimuli also differed on, e.g., stress pattern and order of administration). Furthermore, DDK accuracy was significantly predicted by letter–number switching (b = 0.213, SE = 0.076, t = 2.790), such that participants who were more accurate at switching between the two aspects of the letter-number combination were better able to produce DDK sequences. Additionally, there is an interaction between the lexicality of the DDK stimuli and (letter-number) switching (b = -0.227, SE = 0.083, t = -2.742), indicating that those with better letter-number switching were influenced less by the lexicality of the DDK stimuli. In other words, for participants with good switching ability, the difference between their DDK real word and non-word repetition accuracy was smaller than for those with poorer switching ability.

Similar to DDK accuracy, participants' DDK rate performance differed between real word and non-word sequences, with better (i.e., faster) performance for the real word than non-word sequences. However, DDK rate was not predicted by any of the executive control measures in our study.

Table 3.3 and Figure 3.4 summarise the analysis testing for an association between executive control (from the most parsimonious model) and tongue-twister accuracy and rate.

Table 3.3 The coefficient estimates, standard errors, and significance levels of factors involved in tongue-twister accuracy and rate

	Dependent variable:					
	Accuracy			Rate		
Predictors	Estimates	SE	p	Estimates	SE	p
Intercept	0.675	0.025	<0.001	4.178	0.063	<0.001
Tongue-twister_number2	-0.096	0.030	**0.001**	-0.114	0.053	0.033
Tongue-twister_number3	-0.177	0.030	**<0.001**	-0.653	0.053	**<0.001**
Tongue-twister_number4	-0.075	0.030	**0.012**	0.910	0.053	**<0.001**
Letter-number Switching	0.357	0.126	**0.005**			

Note. Effects that remain significant given an extra conservative alpha level (α = .0125) are shown in boldface.

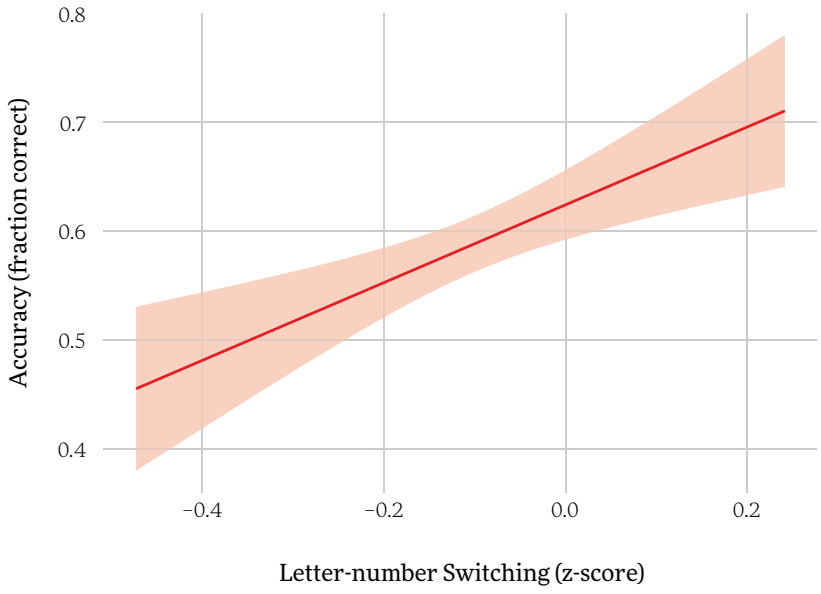


Figure 3.4
Model plot of tongue-twister accuracy in relation to letter–number switching ability

As can be seen in Table 3.3, tongue-twister accuracy differed across the different sentences. Additionally, comparable to DDK accuracy, tongue-twister accuracy was significantly predicted by letter-number switching (b = 0.357, SE = 0.126, t = 2.841), such that those with better switching ability were also more accurate at rapid tongue-twister production. Note that we verified that our results about the link between DDK accuracy or tongue twister accuracy on the one hand and switching on the other also hold if we apply lmer models to the accuracy proportions converted to logits.

Tongue-twister rate, like tongue-twister accuracy, also differed across tongue-twister sentences. As was observed for DDK rate, tongue-twister rate is not predicted by any of the measures of executive control here. As we repeatedly tested for a possible link between aspects of executive control and speech performance (i.e., in four analyses), one can argue that a more conservative alpha level would be appropriate. If we adopt a more conservative alpha level (dividing the critical alpha level by four; $p < .0125$), the relationship between switching ability and DDK accuracy (as well as the lexicality effect and the Switching × Lexicality interaction) and the relationship between switching ability and tongue-twister accuracy remain significant (cf. Tables 3.2 and 3.3).

In summary, participants' cognitive switching ability related to their accuracy in both DDK and tongue-twister tasks. However, performance on the cognitive tasks was not related to participants' maximum rates in the speech tasks.

3.4. Discussion

In this study, we investigated the potential link between maximum speech performance and executive control abilities in a sample of 78 young healthy adults without any language, speech, or hearing impairment. Using two maximum speech performance tasks (i.e., a clinical DDK task and a tongue-twister task), we tapped participants' articulatory control abilities through acoustic (rate) and behavioural (accuracy) data. Both speech tasks require rapid alternation between similar onset consonants or consonant clusters, as we used the sequential version of DDK (repetition of non-word sequences 'pataka' and 'katapa' and real Dutch words 'pakketten' *packages* and 'kapotte' *broken*). Additionally, participants' executive control abilities were assessed by means of three cognitive tasks, that is, a flanker task as an index of inhibitory control, a letter–number task as an index of switching ability, and an operation span task as an index of updating ability.

In general, participants' maximum performance varied for the different types of speech stimuli. More specifically, participants were more accurate and achieved faster speech rates in producing DDK sequences than tongue-twister sentences, possibly due to higher processing load involved in producing the longer 'tongue-twisting' sentences and higher articulatory complexity (involving more complex syllable structures). Within the DDK task, in line with the results obtained from children and healthy older adults (Ben-David & Icht, 2017; Icht & Ben-David, 2015), our young adult speakers performed better in real word than in non-word conditions. That is, they were able to repetitively produce real words more accurately and faster than non-words. We will come back to potential confounds of lexical status with other factors below. More importantly, cognitive switching ability related to both DDK and tongue-twister maximum accuracy, such that individuals with better switching ability were also better able to accurately produce tongue-twister sentences and DDK sequences at a fast rate. This indicates that cognitive switching relates to the rapid production of consecutive alternating speech movements.

Apart from the general effect of cognitive switching on DDK accuracy, an interaction was found between the lexicality of the DDK sequences and cognitive switching ability, such that for participants with better cognitive switching ability, the performance difference between DDK real word and non-word conditions was

smaller, compared to those with poorer switching ability. In other words, those with poorer cognitive switching ability may have benefited more, relatively, from producing familiar sequences such as real words as compared to the relatively unfamiliar and novel non-word sequences. As our results describe relationships from which no causality can be derived, follow-up research with experimental manipulation of cognitive switching load would be required to confirm this. Furthermore, note again that these real word and non-word stimuli differed not only in speakers' familiarity with the required motor programmes but also in their intrinsic prosodic patterns (as also argued in Ziegler, 2002) and in their order of administration in the experimental protocol. For instance, the two real Dutch words both contain one full short vowel (receiving primary stress), one schwa, and one unstressed vowel that could be reduced to a schwa, whereas non-word sequences like 'pataka' or 'katapa' do not have a known stress pattern and contain three long full vowels. Whereas most speakers put primary stress on the initial syllable ('pátaka' and 'kátapa'), we also observed some interspeaker variation in (the consistency of) stress placement and reduction of unstressed syllables to schwa. Uncertainty about the item's stress pattern and about reduction of syllables may contribute to non-word production being more difficult than real word production.

Furthermore, all participants produced the non-word DDK sequences before the real word sequences. Even though participants had had extensive DDK practice before moving on to the critical non-word and real word sequences, having already produced the non-alternating sequences and alternating disyllabic stimuli ('pata' and 'taka') as practice stimuli, we cannot distinguish lexical status effects from order effects on the basis of our design. These confounds may have contributed to the non-word and real word stimuli differing in the amount of executive control required to repetitively produce the sequences.

Our results challenge the idea that 'late stages' of speech-language production are largely automatic (Ferreira & Pashler, 2002; Garrod & Pickering, 2007). Rather, at least when speech production is made as challenging as we did here, executive control seems to relate to speech production, just like it has been shown to relate to language control in bilinguals (e.g., Costa et al., 2006; Rodriguez-Fornells et al., 2006), in noun–phrase production (e.g., Sikora et al., 2016), or in lemma selection (e.g., Piai & Roelofs, 2013; Shao et al., 2012). Our tongue-twister data agree with evidence (McMillan & Corley, 2010) that phoneme production is more error-prone and less target-like in the context of competing phonemes (*kef def def kef*) than when produced in a context without competing phonemes (*kef kef kef kef*). Correct and fluent production of sequences of alternating syllables thus seems to relate to executive control to rapidly alternate between target phonemes. Data by McMillan and Corley (2010) also

suggested that the more similar the competing phoneme to the target phoneme (/t/ being more of a competitor for /k/ than is /d/), the greater the competition effect (McMillan & Corley, 2010). Our DDK stimuli involved switching between highly similar voiceless stops, and accurate DDK performance was indeed also related to switching ability. Our results therefore provide evidence that switching is associated with resolving competition and selection at later stages than lemma selection in speech production (i.e., during phonological and phonetic stages).

Our finding that cognitive switching relates to the production of consecutive alternating speech movements echoes with findings in which cognitive load was manipulated experimentally, such as the findings that cognitive or linguistic load impacted on articulation stability for unimpaired speakers (Dromey & Benson, 2003). Similar findings of cognitive load effects on articulation have also been found among children with specific language impairment (Saletta et al., 2018), as well as for healthy younger and older adults (MacPherson, 2019; Sadagopan & Smith, 2013). In MacPherson's (2019) study, healthy younger and older adults' articulatory control was measured through reading aloud sentences that formed Stroop and non-Stroop conditions. MacPherson found that articulatory motor stability was affected by Stroop interference, and that older adults' speech motor performance was more detrimentally affected in the Stroop condition than that of younger adults. The findings from these studies align with our findings. In our findings, those with poorer executive control were less accurate in rapidly producing alternating sequences, which can be seen as speech motor breakdown. In the MacPherson (2019) study, participants with supposedly poorer executive control due to their older age were more impacted by cognitive stress on their speech motor performance compared to those with supposedly better executive control.

The findings in this study are novel in the sense that we showed evidence of speakers' articulatory control abilities as reflected by maximum speech performance to be related to their executive control abilities, backing up the statement that "speech, or any motor behaviour, is best viewed as a cognitive-motor accomplishment" (Kent, 2004, p. 3). However, note that our approach to articulatory control was maximum performance in terms of rate and accuracy, instead of speech motor stability as in work by, for example, Dromey and colleagues. Thus, further examination of underlying speech motor control through kinematic measures is required to investigate how cognitive ability may relate to articulatory stability. Additionally, our results also broadened the perspective on the relationship between articulatory and executive control by providing data from a young nonclinical rather than a clinical sample (e.g., Nijland et al., 2015; Peter et al., 2018; Shriberg et al., 2012).

In comparison to performance on the DDK task, rapidly producing tongue-twister sentences were shown to be more challenging for our speaker sample, as reflected by the lower and more variable rate and accuracy in performance. As described in Shen and Janse (2019), there may be multiple (methodological) reasons for the difference in performance between the DDK and the tongue-twister tasks. First, compared to DDK sequences, tongue-twister sentences are proper sentences and consequently involve more grammatical and semantic processing. Moving from word or non-word repetition to sentence repetition therefore demands a higher linguistic processing load. Second, some words in the tongue-twister sentences contain consonant clusters in both syllable-onset and -offset positions (e.g., /bl/ for onset and /tst/ for offset, respectively) and more varied phonetic contrasts (e.g., the place of articulation and voicing of the alternating stop consonants /p/, /b/ and consonant clusters /pl/, /bl/), whereas both non-word and real word DDK sequences contain rather simple consonant–vowel structures and less complex phonetic contrasts in their syllable-onset consonants (i.e., only the place of articulation differed in singleton onset consonants /p/, /t/, and /k/). However, despite their differences in linguistic content and linguistic processing load involved, performance in both speech tasks related to (elements of) executive control in that speakers had to switch between similar competing phonemes and had to keep track of where they were in their production of the sequence.

As laid out in the introduction, even though cognitive switching was thought to be most relevant to our specific speech tasks, all three elements of executive control were expected to relate to speech performance to a certain degree. Updating was expected to relate to rapid production of alternating sequences because planning and programming of speech movements need to be constantly updated. Inhibitory control was expected to relate to the suppression of coactivated but incorrect competing phonemes and/or phoneme clusters in the speech stimuli. Our results could indicate that cognitive switching is involved in speech production (in the way speech production was operationalised here) and that inhibitory control and updating are not or to a lesser degree. However, alternative explanations cannot be ruled out. The updating and inhibitory control measures used here could potentially be noisier than the switching measure, in that they were less successful in capturing the target ability. For instance, inhibitory control data of seven participants had to be excluded due to their low accuracy in the flanker task, whereas no one failed to meet the minimum accuracy requirement in the letter–number task that measured switching. Arguably, the level of difficulty of the letter–number task should be higher than that of the flanker task given the complexity of the letter–number task. However, due to a difference in task design, participants did more practice trials in the letter–number task (two blocks of 48 trials) than in the flanker task (one block of 12 trials).

Additionally, participants received feedback on their performance (i.e., number of errors made) after each block in the letter–number task, while no feedback was ever given during the entirety of the flanker task. The feedback in the letter–number task might have motivated participants to pay more attention to the task, resulting in better task performance. We cannot rule out the possibility that we might have gotten a 'purer' measurement of participants' switching ability than their inhibitory control ability, and this could have contributed to observing an effect of switching but not of inhibitory control on performance in the speech tasks.

The existence of various sources of 'noise' in different tasks brings up the issue of the validity of using a single task to measure (a given aspect of) executive control. As Miyake and colleagues described in their studies on measuring the construct of executive control, 'task impurity' was listed as one of the problems when using single tasks to measure aspects of executive control (Miyake & Friedman, 2012; Miyake et al., 2000). *Task impurity* refers to unwanted systematic variance that exists in different cognitive tasks, for example, additional processing of the number being odd or even in the letter–number task. To minimise this task impurity problem, Miyake and colleagues proposed a 'latent variable approach', which is to use multiple tasks that capture the target ability and to extract the commonality across the tasks (a latent variable) as the measure of the targeted executive control (Miyake & Friedman, 2012; Miyake et al., 2000). This latent variable approach, that is, using multiple tasks to measure a construct, requires larger participant sample sizes than used in this study, but it may be an approach to pursue in future research.

Another possible reason why updating and inhibitory control did not predict speech performance could be our analysis method of having all predictors in our initial model. Although updating ability was not found to be significantly involved in the speech performance, it was actually approaching significance ($p = .08$) in the model of DDK accuracy (in addition to the observed interaction between switching and lexicality). Moreover, when updating (rather than switching), lexicality, and their interaction were included as the only fixed effects of interest in a model, both updating and the interaction between updating and lexicality became significant predictors of DDK accuracy (same direction of effects and interaction as observed for switching ability). This finding highlights the collinearity problem of having correlated predictors, even if their correlation does not exceed '.3'. In other words, due to the significant correlation between updating ability and switching ability, inclusion of the stronger predictor (switching in this case) overruled the potential contribution of the weaker predictor (updating). This observation echoes with Miyake and colleagues' arguments on the unity and diversity of executive control measured with simple laboratory tasks: different elements of executive control are correlated yet separable.

As noted in the introduction, updating of the working memory has been shown to be involved in the 'early stage', that is, the formulation stage of the speech production. Our results suggested that aspects of executive control may also relate to the articulatory planning and execution of speech. The significant association between switching ability (and the marginal association between updating ability) and the production of the two maximum performance speech tasks that tax the late speech production processes, therefore, challenges the idea that 'late stages' of speech production are largely automatic (Ferreira & Pashler, 2002; Garrod & Pickering, 2007), at least when speech production is made as challenging as we did here.

Lastly, the four selected tongue-twister sentences tested here turned out to vary in difficulty level, as evidenced by rate and accuracy analyses. The more challenging tongue-twister sentences were (a) 'Frits vindt visfrietjes vreselijk vies' (nine syllables) and (b) 'Ik bak een plak bakbloedworst' (seven syllables), while the relatively easy ones were (c) 'De poes kotst in de postzak' (seven syllables) and (d) 'Papa pakt de blauwe platte bakpan' (10 syllables). The syllable counts already indicate that the difficulty difference is unlikely to be due to sentence length. For a sentence to be a real tongue twister, it should have both repetition and alternation of sounds, leading to facilitation of the repeated consonant and hence interference for any switches in sound (Monaco et al., 2017). The number of repeats and alternations is low in the easy sentence c, but also in the difficult sentence b, and is not low in the easy sentence d. The difference in difficulty level may perhaps relate to the fact that the more challenging sentences contain trisyllabic (compound) words whereas the rather easy ones are mainly composed of bisyllabic simple words. Alternatively, the difficulty difference may relate to easier switching between alternating singleton consonants than alternating singleton consonant onsets and consonant cluster onsets. Only better controlled sentence sets would allow systematic evaluation to determine whether alternating between places of articulation is easier or more difficult to produce than alternating between simple and complex onsets.

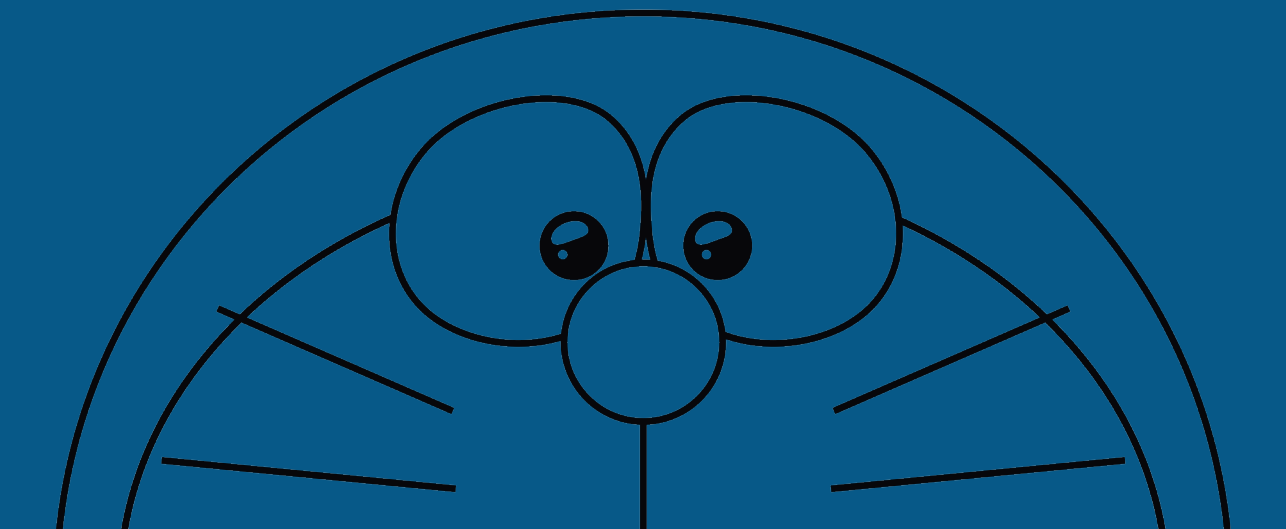
3.5. Clinical implications

When testing articulatory control using the maximum performance DDK task, clinicians can consider administering both non-word and real word stimuli. Our data suggest that, at least for the young healthy adult speakers tested here, the link to executive control (particularly switching) may be stronger for the production of nonsensical DDK sequences than for real word sequences. However, follow-up research is required to establish whether these results generalise to other populations (including clinical populations of different ages) and to better controlled designs and

stimuli (as lexical status in our design was confounded with other factors, see a more detailed explanation above). If our results are found to hold more generally, this stronger link with cognitive switching for non-word compared to real word DDK could be a reason to opt for either type of DDK stimuli, depending on the patient (group) or purpose of the speech assessment. Either way, it may be good practice to use both non-word and real word stimuli in a DDK task to get a more complete picture of participants' speech motor/articulatory control skills. Our results also suggest that DDK may be a challenging task for clinical populations with cognitive impairment. Our analyses also showed that accuracy for DDK performance was more informative than maximum rate itself. This also held for tongue-twister performance, but this observation of accuracy being more informative than rate may have been specific for our young and unimpaired sample.

3.6. Conclusion

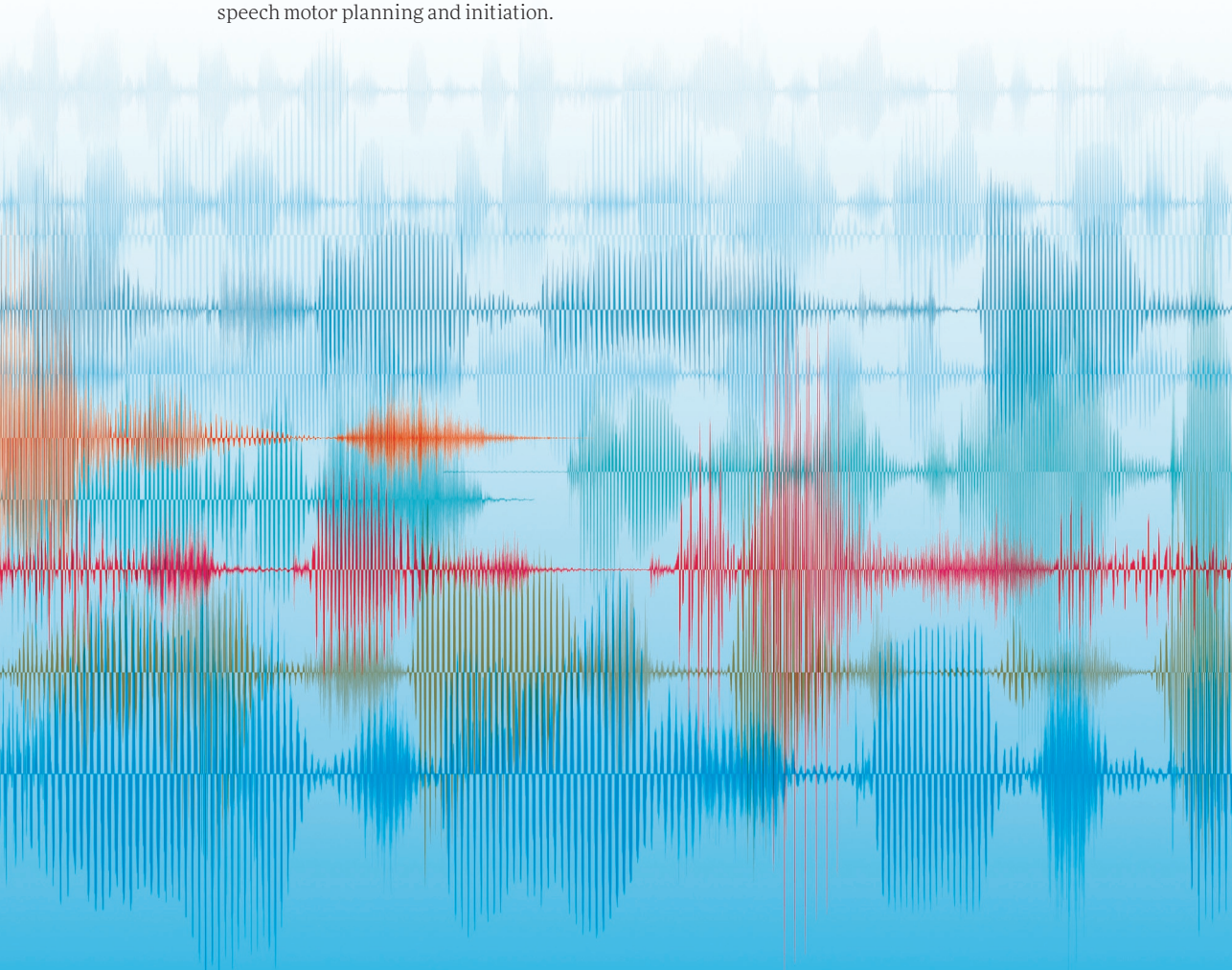
On the basis of our individual differences approach, we conclude that executive control (cognitive switching in particular) relates to speech motor control as quantified with maximum speech performance measures. This finding extends the body of evidence on the link between cognition and language production to late stages of production, such as articulation.



4

Decomposing Age Effects on Speech Motor Planning and Initiation

This chapter is based on the following:
Shen C., Maas E., & Janse E. (in prep). Decomposing age effects on speech motor planning and initiation.



Abstract

Apart from well-known age effects on word finding, speech production in cognitively healthy older adults may also be affected by age-related declines in the planning and/or initiation of speech movements. To find out whether age differences increase if speakers not only need to initiate their response, but also select their response, we adopted a speeded speech-production task in which participants produced a target stimulus in two conditions (prepared vs. unprepared). In the 'prepared' (or simple) condition, participants produced the target stimulus after having prepared it in advance and having waited for the 'go' signal. In the 'unprepared' (or choice) condition, participants could only start preparing for target stimulus production upon receiving a critical-information cue.

Speech production accuracy and latency data were collected from 30 healthy younger and 26 healthy older adults using a speeded speech-production task with monosyllabic and disyllabic targets as stimuli. Our results showed that younger adults were faster than older adults in producing target stimuli in the simple condition, but they slowed down more than older adults in the choice condition, relative to the simple condition (i.e., when they also needed to select and programme their response). Consequently, age differences were only apparent in the simple condition. These results may suggest that ageing affects speech initiation more than speech production processes that can be prepared in advance. Alternatively, these results may mainly reflect strategic response differences between age groups.

4.1. Introduction

Age-related decline in cognitive function has been studied extensively in the past decades (e.g., Glisky, 2007; Murman, 2015; Salami et al., 2012). Changes in the planning and execution of motor movements over the adult life span have also received considerable attention (e.g., Niermeyer et al., 2017; King et al., 2013; Krampe et al., 2005; Sparto et al., 2014). Speech production requires both cognitive functions and speech motor control. The link between speech production and cognitive control has been demonstrated most clearly for early stages of speech production (e.g., formulation and lemma selection), which have been related to cognitive or executive control abilities such as working memory (e.g., Piai & Roelofs, 2013) and sustained attention (e.g., Ferreira & Pashler, 2002). Consequently, given age-related cognitive decline, older adults experience more difficulties in word finding than younger adults (e.g., Bowles & Poon, 1985; Burke et al., 1991; Shafto et al., 2010). Studies on age-related decline in speech production have mainly focused on difficulties in early (formulation) stages of speech production, for instance lexical processes (e.g., Lima et al., 1991; Moers et al., 2017) and semantic processing (e.g., Mayr & Kliegl, 2000). However, less is known about age-related difficulties in later stages of speech production, such as articulation. Recent studies have suggested that late stages (e.g., phonological encoding and articulation) of speech production also relate to executive control abilities (e.g., Jongman et al., 2015; Shen & Janse, 2020, see also Chapter 3), such that age effects can also be anticipated for these stages or processes. A few studies on speech motor performance have revealed that speech production in cognitively healthy older adults may be affected by age-related declines in the planning and execution of speech movements and speech motor performance (e.g., Tremblay et al., 2017; MacPherson, 2019; Mailend et al., 2019; Tremblay et al., 2018). The current study will follow up on these results, in an attempt to unravel which aspect(s) of speech motor planning may be susceptible to age-related decline.

4.1.1. Motor planning and execution

Speech production is a complex process that, according to Tremblay and colleagues (2019), requires "neuromotor mechanisms to implement phonological planning, response selection, sequencing, and timing, contextual adjustments of the motor programmes, as well as action execution and response monitoring" (Tremblay et al., 2019, p. 1). Younger and older adults have been shown to differ in their planning and execution of complex rhythm production (i.e., in conditions with high sequencing demands), and this age difference was argued to be modulated by executive functions (EF) (Krampe et al., 2005). Additionally, in a study on leg movement, Sparto and colleagues examined age differences and effects of varying levels of inhibition control requirements on a voluntary lateral step initiation task (Sparto et al., 2014).

Difficulty of this step-initiation task was manipulated into three levels: basic sensory/motor function, the integration of this function with a choice decision, and the inhibition of a strong sensory-motor association during a choice task. Through analysing postural adjustment errors and step latencies, Sparto and colleagues found that older adults were more variable in step behaviour than younger adults, produced more postural adjustment errors during conditions requiring inhibition, and had larger step initiation latencies that increased more than younger adults as the inhibition requirements of the condition became greater (Sparto et al., 2014). Similarly, using the Stroop interference paradigm by having participants read sentences containing colour words which either matched or mismatched with the print colour, MacPherson (2019) illustrated that age-related cognitive decline in older adults plays a role in their speech motor performance in the way that age differences in speech motor control were enlarged under conditions where inhibitory control was required by the task.

More evidence for a relationship between cognitive control and motor control comes from a study by Niermeyer and colleagues (2017), that used a computerised motor sequencing task (i.e., Push-Turn-Taptap task). Their task evaluated multiple aspects of motor sequencing: action planning, action learning, and motor control speed and accuracy. They manipulated sequence complexity through progressively longer sequences (ranging from two to five movements) to assess the relationship between executive functioning (EF) and motor sequencing among younger and older adults. They found that the action planning aspect of motor sequencing, as indexed by the median time between completion of one sequence and initiation of the next (error-free) sequence, was related to EF for both age groups, whereas the action learning and motor control accuracy aspects were associated with EF for older adults only. Furthermore, while complexity affected older adults' performance more negatively than that of younger adults, the relationship between EF and motor sequencing seen for older adults was regulated by task complexity such that EF was related to performance only for the more complex (longer) sequences.

Speech production tasks do not just require movement planning and control, they may also require participants to respond quickly. In addition to the studies on movement control, researchers have also manipulated the difficulty of button-press response tasks to evaluate age effects on speeded performance tasks. For instance, Der and Deary (2006) investigated relationships between RT parameters and participants' age and sex using a large sample from a national health and lifestyle survey in the UK (7000+ respondents aged between 18 and 94 years). RTs in their analyses were collected in two conditions. Namely, a simple condition where participants only had to press one button upon stimulus presentation, and a choice

condition where participants were instructed to press one of the four buttons depending on the presented stimulus (i.e., simple condition versus 4-choice condition). The authors reported that RTs show different age-related patterns for the simple and the four-choice conditions. Specifically, the simple-RT mean remained more or less unchanged until people reached about 50 years of age while the choice-RT mean increased throughout adulthood (Der & Deary, 2006), such that age differences may be more apparent in the choice than simple condition. Moreover, in a later study using the same simple versus four-choice task paradigm but with three age groups (i.e., at 30, 50, and 69 years), Der and Deary (2017) found that participants' mean RTs were strongly (and negatively) correlated with intelligence (measured through part I of the Alice Heim 4 test of general intelligence, which measures verbal and numeric cognitive abilities) for both simple RT and choice RT, and that these correlations increased with age, especially in the choice RT condition. These results demonstrate age-related slowing of response times in general, and also show that age more strongly affects responses involving a choice between response options than simple responses.

Given the fact that speech production is a highly complex form of motor behaviour, it is reasonable to think that declining motor sequencing mechanisms in older adults may also affect their speech production. For instance, Tremblay and colleagues (2017) studied age differences in the neuromotor control of speech production through a combination of behavioural and functional magnetic resonance imaging analyses. They manipulated non-word production complexity in two different ways. First, sequential complexity was manipulated by comparing repetition of one syllable versus repetition of three alternating syllables. A second (so called 'motor') manipulation involved that of syllabic complexity (i.e., repeating consonant-vowel versus consonant-consonant-vowel syllable structures). Tremblay and colleagues (2017) found that older adults exhibited longer and more variable movement times (MT) than younger adults, especially at high motor and sequence complexity levels. These results indicate that ageing of motor control mechanisms may contribute to age differences in speech production, especially for more complex speech production. In a later study, Tremblay and colleagues (2018) further investigated ageing effects on motor aspects of speech production and the underlying senescence mechanisms. By manipulating syllable frequency and phonological complexity (simple or more complex syllable structures) of the to-be-produced non-words, Tremblay and colleagues wanted to test whether speech motor performance declines with age. Specifically, by varying syllable frequency, they aimed to test whether age affects the online assembly of syllables from phonemes, which is needed for low-frequency (but not high-frequency) syllables as high-frequency syllables would be easily retrievable from a mental syllabary (Levelt et al., 1999). Moreover, the phonological complexity manipulation was implemented to test whether age-related decline particularly

affects the more complex structures. The authors found age-related increases in production error rates and changes in speech timing across the board, indicating a general decline in the planning and execution phases of speech movements. Additionally, age was found to interact with phonological complexity in vocal response duration (RD): more complex syllables exhibited longer RDs, and the interaction reflected a stronger relationship between ageing and RD for complex syllables in comparison to simple syllables. Moreover, response RTs and RDs showed a negative association for the data of younger, but not older adults. This meant that younger adults were found to produce longer RDs with shorter RTs, suggesting that younger adults could slow down their speech execution to complete their planning online. This ability to flexibly adjust planning was not observed in older adults. The authors thus argue that their results provide evidence for age-related decline in speech production. However, because planning in their study included both the retrieval of motor programmes and the organisation of those programmes into a smooth sequence for articulation, their results do not clarify which speech motor process(es) was particularly affected by ageing.

Several tasks have been proposed to allow breaking down the latest stages of speech production. To better understand speech motor planning (impairment), Mailend and colleagues (2019) conducted an experiment to test the nature of the impairment in apraxia of speech (AOS). More specifically, they investigated speech motor planning in light of two competing hypotheses on AOS: The Reduced Buffer Capacity Hypothesis (i.e., people with AOS can only hold one syllable at a time in the speech motor planning buffer) versus the Programme Retrieval Deficit Hypothesis (i.e., people with AOS exhibit difficulty accessing the intended motor programme when several motor programmes are activated simultaneously). They tested RTs and accuracy of single-word production in three groups of participants: speakers with apraxia of speech (AOS), speakers with (only) aphasia, and healthy controls. Through manipulating the initial consonant of the prompt word to be identical, slightly different (in one articulatory feature), or different (in all articulatory features) from the initial consonant of the target word, they were able to differentiate the aspects of retrieving and unpacking of motor programmes from the internal preparation of the required motor programmes. Their results showed that all speaker groups showed the fastest responses in the identical condition, but the AOS speakers, rather than the aphasia speaker group, were slowed down differentially in the condition where prime and target words differed, compared to those trials with the same prime and target words. They therefore concluded that speakers with AOS have difficulty selecting the target motor programme among the competing motor programmes.

The study by Mailend and colleagues (2019) formulated speech motor planning in a psycholinguistic way, where the target motor programme had to be selected from multiple competing motor programmes. This psycholinguistic view followed up on earlier work by Klapp (1995, 2003), who proposed a two-stage model of motor programming that distinguished a pre-programming stage from a sequencing stage. The pre-programming stage prepares the internal (hence the abbreviation of INT) properties of a selected motor programme. This INT process is followed by a sequencing process that assigns serial order to multiple programmes in a sequence (hence the abbreviation of SEQ). The processing load of the INT stage is dependent on the complexity of a single motor programme, while the load of the SEQ depends on the number of motor programmes that should be put in order. To study the effects of processing load on the INT and SEQ stages separately, Klapp (2003) further proposed an experimental simple/choice reaction time (RT) paradigm in which participants needed to produce a pseudoword in prepared or unprepared conditions. In the simple condition, participants are able to pre-select and pre-programme the required motor programme(s) as they have already been told what pseudoword they will need to produce shortly. Consequently, upon receiving the 'go' cue, participants had already performed the INT process, and participants only need to carry out the sequencing process. As a result, the simple RT reflects the time participants needed to carry out those final processes of sequencing and initiation of their production. In the choice condition, participants had to select the target motor programme from a number of possible options and then also sequence the required motor programmes online, as they were asked to speak immediately upon receiving the information cue on which pseudoword to produce. Consequently, the choice RT reflects the time participants need to carry out the INT plus SEQ stages.

Maas and colleagues (2008) examined motor programming in speakers with apraxia of speech (AOS) in the context of the above-mentioned two-stage model, aiming to test whether AOS involves a deficit in the INT (preprogramming) stage of processing and/or the SEQ (online serial ordering) and initiation of movement. Through a series of experiments that involved finger movements and speech movements comparable to the finger movements, they found that speakers with AOS exhibited longer preprogramming but typical sequencing and initiation times compared to healthy controls (for both non-speech and speech movements), as evident from group differences showing up particularly in the choice condition. These results therefore demonstrated that speakers with AOS exhibit a process-specific deficit in the INT stage of motor programming processing rather than a deficit in the serial ordering and initiation of movement in the SEQ stage.

As stated in Maas and Mailend (2012), the choice reaction time condition "requires the participant to respond immediately (i.e., make a choice between alternative possible responses and programme the response)". Thus, the RT measure captures all component processes, such as visual processing of stimuli, phonological planning, and motor planning. The simple RT condition, on the other hand, "allows the participant time to select and prepare the response in advance before a go-signal...[that] cues the response" (Maas & Mailend, 2012, p. S1525). The simple RT, therefore, captures mainly the later stages of motor planning, such as sequencing and initiation. As a result, a comparison of the simple versus choice RT could reveal differences in the early planning/programming and the late sequencing and initiation phases of speech preparation, as illustrated by the results for AOS by Maas and colleagues (2008). As such, the simple/choice paradigm is a robust experimental method to break down speech motor planning into early (motor programme selection/preparation) and late (sequencing and initiation) stages in different speaker groups. We will, therefore, adopt this paradigm in an attempt to unravel which aspect(s) of speech motor planning may be susceptible to age-related decline.

4.1.2. The present study

This study was set up to find out whether adult ageing mainly affects speech production processes that can be prepared, or processes that cannot be prepared in advance. In doing so, we follow related attempts to identify the locus of the speech production problem in clinical populations, e.g., in studies on AOS or dysarthria (Maas et al., 2008; Mailend et al., 2019; Reilly & Spencer, 2013a, 2013b). Specifically, in order to further our understanding of the relationship between ageing and speech motor planning and execution, we used a speeded speech-production task in which participants produced a target stimulus in two conditions (prepared vs. unprepared). In the 'prepared' (or simple) condition, participants produced the target stimulus after having prepared it in advance and having waited for the 'go' signal. In the 'unprepared' (or choice) condition, participants could only start preparing target stimulus production (from a choice of three known alternatives) upon receiving a critical-information cue. Consequently, this choice condition is the more complex condition. Age groups were expected to differ in their speech onset latencies in both conditions, but we aimed to find out whether age differences increased or decreased going from prepared (simple) to unprepared (choice) condition. Earlier research requiring speeded manual (button press) responses generally showed stronger age effects in more complex conditions (e.g., Der & Deary, 2006), which would lead to the expectation of stronger age effects in the unprepared condition than prepared condition in our speech study. Hence, this simple/choice reaction time paradigm can be used to investigate whether age groups differ primarily in speech production processes that cannot be pre-programmed (i.e., those processes evident from simple-

condition RT, such as retrieving and unpacking of motor programmes from a buffer), or in production processes that can be pre-programmed (i.e., internal preparation, as evident from the difference between choice and simple RT).

4.2. Methods

4.2.1. Participants

Speech production data were collected from 30 healthy younger-adult (18 females, *M* = 22;7 years, range = 18 - 32) and 30 healthy older-adult (18 females, *M* = 69;5 years, range = 65 - 77) participants. All participants were recruited online through the participant pool of the Radboud Research Participation System. They were all native speakers of Dutch, and none had any known history of speech, hearing, or reading disabilities, nor past diagnosis of speech pathology or brain injury at the time of testing. Moreover, they all reported to have normal or correct-to-normal vision. In addition to the recruitment criteria, older participants also went through a quick screening. The quick screening was composed of a Montreal Cognitive Assessment (MOCA) test (a screening instrument to test for mild cognitive impairment), and an audiological hearing test testing (air-conduction) hearing thresholds at 500, 1000, and 2000 Hz through an audiometer (Oscilla USB-330). Participants needed to have a MOCA score of 26 or higher for their data to be included in the study, as well as a Fletcher index (hearing threshold averaged over 500, 1000 and 2000 Hz) of 30 dB HL or lower in the better ear. This was done to avoid hearing loss or cognitive impairment effects on participants' production abilities. The study protocol was evaluated and approved by the Ethics Assessment Committee Humanities at Radboud University. All participants gave informed consent for their data to be analysed anonymously, and they received either course credits or gift vouchers as compensation for their time.

After examining the screening results, data from three older female participants had to be excluded from further analysis: one participant failed the MOCA test (scored lower than 26 points) and two failed the hearing test (averaged hearing thresholds over 500, 1000, and 2000 Hz exceeding 30 dB HL). Additionally, data from one older male speaker was also excluded because he told the experimenter that he suffered from a neurodegenerative disease after completing the experiment. After applying these exclusion criteria, data from 30 younger adults and 26 older adults were used for further analyses. Average MOCA score in the resulting older adult sample was *M* = 28.2 (*SD* = 1.3), and the average Fletcher index was *M* = 14.3 (*SD* = 7.2) and *M* = 14.4 (*SD* = 6.4) for the left and right ear respectively.

4.2.2. Simple/choice speech production task

A simple/choice speech production task was designed based on the experiments used in Klapp (1995) and Klapp (2003). More specifically, we selected three monosyllabic and three disyllabic stimuli, staying as close as possible to those used in Klapp (2003). The stimuli all had simple consonant-vowel structures. The monosyllabic stimuli were: /tu/ 'Toe', /ka/ 'Kaa', and /bi/ 'Bie', while the disyllabic stimuli were: /tuka/ 'ToeKaa', /kabi/ 'KaaBie' and /bitu/ 'BieToe' (the orthography represents the Dutch spelling of the stimuli). Note that the disyllabic stimuli were all pseudowords, whereas not all monosyllabic stimuli were pseudowords in Dutch (e.g., 'Toe' is a real Dutch word).

In addition to the difference in stimulus length, there are two conditions in a simple/choice speech production task, namely, a 'prepared' condition and an 'unprepared' condition. In the 'prepared' or 'simple' condition, participants' speech onset was prompted by first visually presenting the initial phoneme of the target stimulus (i.e., either 'K', 'T', or 'B'), followed by a visual GO cue accompanied by a beep tone (at around 4000 Hz) after 500 ms of preparation time. In the 'unprepared' or 'choice' condition, instead of receiving the initial phoneme, participants were presented with a '***' sign first. Speech onset was then prompted 500 ms later by visually presenting the initial phoneme of the target stimulus (i.e., either 'K', 'T', or 'B') as a critical-information cue, accompanied by a beep tone. Participants were instructed to produce the target stimulus as soon as possible upon receiving the visual GO cue on the monitor screen and the simultaneous beep tone via loudspeakers. They were also reminded to *not* produce any speech before the GO cue in the 'simple' condition (for an illustration of the task paradigm, see Figure 4.1 below).

Trials were blocked by stimulus length (mono- or disyllabic) and condition (simple or choice) with each block containing 18 experimental trials (six repetitions of each of the three target stimuli per block). Recall of the disyllabic stimuli was found to be more difficult than that of the monosyllabic stimuli (as shown in pilot data from nine younger participants). Therefore, the experimental block that contained disyllabic stimuli always followed the block containing monosyllabic stimuli during the experiment session. Additionally, different numbers of familiarisation trials were needed for the two length conditions. Prior to the experimental trials, 12 and 18 familiarisation trials were implemented for the monosyllabic and disyllabic blocks respectively. Participants may have needed more familiarisation trials for the disyllabic than monosyllabic stimuli because longer stimuli are more difficult to remember than shorter ones. This easier recall of monosyllabic stimuli can be due to a length effect and/or to a difference in lexical status between monosyllabic and disyllabic stimuli. Thirdly, recall of the longer stimuli was impacted by interference

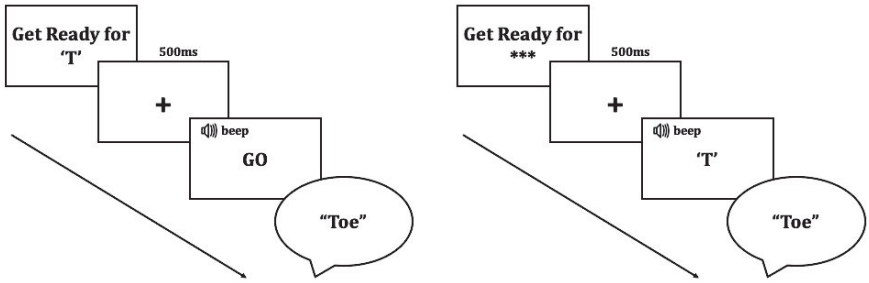


Figure 4.1
Illustration of the speech production task in the 'simple' condition (left) and 'choice' condition (right)

from the shorter ones. The discrepancy in the number of familiarisation trials was applied to compensate for the difference in the level of difficulties posed by the mono- and disyllabic stimuli in the current experiment.

4.2.2.1. Speech recording

Participants completed the simple/choice speech production task in a sound-attenuating recording booth at the Centre for Language Studies Lab of Radboud University. The simple/choice speech production task was programmed in and administered with Presentation software (version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA). All visual stimuli were presented on a 24" full HD monitor, and the beep tone (75 ms in length) was presented through a pair of external loudspeakers each placed on one side of the monitor to the participants. The volume of the beep tone was adjusted per participant to ensure good audibility. One audio recording was made per participant using a Sennheiser ME 64 cardioid capsule microphone on an adjustable table-stand through a pre-amplifier (Audi Ton) onto a Roland R-05 WAVE recorder, to record the beep tone as well as each participant's speech production.

4.2.2.2. Speech production accuracy analysis

We first examined speech production accuracy for both age groups. A response was coded as an error if participants produced an incorrect item. Responses were coded as too fast, and hence invalid, when their RT was less than 200 ms (considered as the minimum time needed to process the auditory and/or visual GO cue). A response threshold was set for 12 correct and valid trials per condition/sub-task (i.e., 1/3 of the total 18 experimental trials). This inclusion of correct and valid responses only was done to ensure that all participants included in further latency analyses had comprehended and performed the tasks at an acceptable level. Data from five younger (two male and three female) participants and one older (female) participant failed to

meet the data inclusion criteria. For three out of the five younger adults, exclusion was due to task compliance issues (i.e., participants producing too many too fast responses). Data of these participants were excluded from further accuracy and latency analyses, leaving 25 younger adults and 25 older adults.

For the accuracy analysis, those 'too-fast' responses were excluded. To analyse accuracy, a generalised mixed-effects (henceforth GLME) logistic regression model was run in RStudio (version 1.2.1335), using the lme4 package (version 1.1-23) (Bates et al., 2015). The model had the binomial family specified and a maximal number of iterations of 100,000 in the optimiser 'bobyqa'. Participants' accuracy data, as the numerical dependent variable, was contrast-coded (1 for correct responses and 0 for errors). Fixed effects of interest were age (younger versus older, with younger mapped on the intercept), condition (simple versus choice, with simple mapped on the intercept), stimulus length (monosyllabic versus disyllabic, with monosyllabic mapped on the intercept), as well as their interactions. Additionally, within-block trial number and initial phoneme of the stimulus ('B', 'T', or 'K', with 'T' mapped on the intercept) were included as control predictors, as well as the interaction between initial phoneme and stimulus length (monosyllabic or disyllabic), and a three-way interaction between within-block trial number, age, and condition. Participant was included as a random effect (Baayen et al., 2008), with condition and stimulus length as random by-participant slopes to capture individual variability amongst participants in the size of the condition and length effect. Inclusion of these two random slopes significantly improved the model fit, thus they were kept in the model. This GLME model was then stripped in a stepwise manner to arrive at the most parsimonious model. Specifically, insignificant interactions were taken out first, and then the insignificant effects were removed, starting with the ones that have the lowest (absolute) z-value in the model output. Model comparisons using the anova() function in R were applied after each removal of the least significant predictor to verify that the removal of each predictor term did not result in significant loss of model fit.

4.2.2.3. Speech onset latency extraction and analysis

Speech onset latency was measured as the time (in ms) between the onset of the beep tone (which served as the cue for participants across conditions to start speaking) and the acoustic onset of their speech production. The onset of the beep tone and the onset of speech were identified and labelled manually in Praat (Boersma & Weenink, 2017) via a TextGrid file using the individual speech recording per participant.

Speech onset latency data from every correct and valid response was then extracted (3460 observations out of the total of 3600, or 96%). In addition, latency values that were outside 3 standard deviation (SD) of each individual participant's mean latency (over those correct and valid responses; 52 observations out of the total of 3460, or 1.5%) were also removed before the statistical analyses. A linear mixed-effects (henceforth LME) model was run in RStudio (version 1.2.1335), using the lme4 package (version 1.1-23) (Bates et al., 2015) to analyse production latencies. Similar to the GLME model for the accuracy data, participants' speech onset latency data (log-transformed to make the latency distribution more normal) was entered as a numerical dependent variable, with age (younger versus older, with younger mapped on the intercept), condition (simple versus choice, with simple mapped on the intercept), stimulus length (monosyllabic versus disyllabic, with monosyllabic mapped on the intercept), as well as their interactions entered as fixed effects of interests. Additionally, initial phoneme of the stimulus ('B', 'T', or 'K', with 'T' mapped on the intercept) and its interaction with stimulus length (monosyllabic or disyllabic) were entered as fixed control predictors, to account for variance that is otherwise left unexplained. We also added within-block trial number as a control predictor to the model, and its potential interactions with age and condition (a three-way interaction), to investigate whether younger and older participants differentially speeded up or slowed down over the course of simple and/or choice experimental blocks.

In addition to the fixed effects, we included participant as random effect (Baayen, 2008), as well as condition and stimulus length as random by-participant slopes to capture individual variability amongst participants in the size of the condition and stimulus length effects. Similar to the GLME model, inclusion of these two random slopes significantly improved the model fit, thus they were kept in the model. This full LME model was then stripped in a stepwise manner to arrive at the most parsimonious model. As described for the accuracy analysis, insignificant interactions were taken out first, followed by removal of insignificant effects, starting with the ones that have the lowest t-values in the model output. Model comparisons (using the anova() function in R) were applied after each removal of the least significant predictor to verify that the removal of each predictor term did not result in significant loss of model fit.

4.3. Results

4.3.1. Descriptive results of production accuracy and speech onset latency

Descriptive results for response accuracy in the two age groups in the simple and choice conditions are summarised in Table 4.1 below.

Table 4.1 Response accuracy (in %) for the two age groups in the two speech conditions

Condition	Younger Adult		Older Adult	
	Monosyllabic	Disyllabic	Monosyllabic	Disyllabic
Simple	99.3%	98.1%	97.3%	95.5%
Choice	99.1%	94.6%	98.9%	96.7%

Additionally, descriptive results of the speech onset latencies per condition are summarised in Table 4.2 and plotted in Figure 4.2 below.

Table 4.2 Speech onset latency results for monosyllabic and disyllabic targets (in ms) from the two age groups in the two conditions

Condition	Younger Adult		Older Adult	
	Monosyllabic	Disyllabic	Monosyllabic	Disyllabic
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Simple	370 (117)	379 (130)	465 (140)	477 (160)
Choice	486 (113)	558 (141)	506 (113)	595 (180)

4.3.2. Speech production accuracy

Results from the GLME model on speakers' production accuracy data are displayed in Table 4.3 and illustrated in Figure 4.3.

As can be seen from Table 4.3, Age significantly influenced production accuracy. That is to say, older adults were less accurate than younger adults at producing the task stimuli in the simple condition. Length of the target stimulus also affected accuracy, with lower accuracy levels for the disyllabic stimuli than monosyllabic stimuli. This held for both age groups, and across conditions. Additionally, Condition and Trial

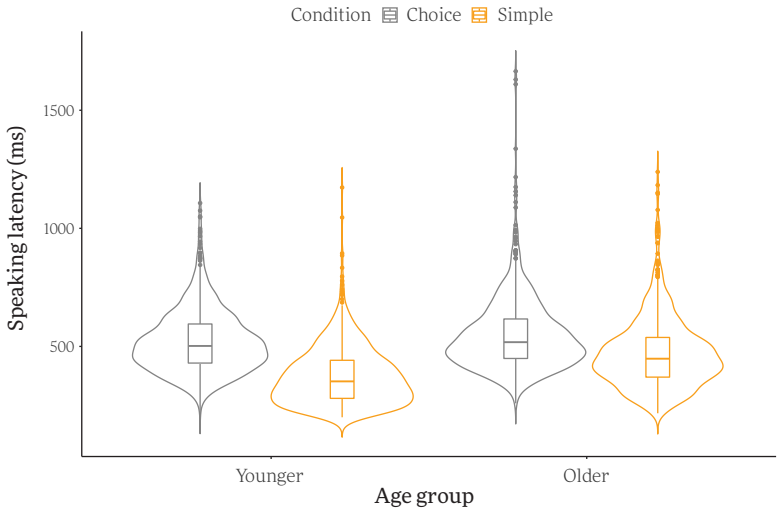


Figure 4.2
Violin plots showing the distribution, median, and range of speech onset latency for the two age groups in the two speech conditions

Table 4.3 Results of the statistical model for speech production accuracy (significant effects printed in boldface)

	Accuracy		
Predictors	Log-Odds	SE	p
(Intercept)	5.92	0.79	**<0.001**
Age [older]	-1.91	0.60	**0.022**
Condition [choice]	-0.64	0.54	0.236
Length [disyllabic]	-1.83	0.65	**0.005**
Initial phoneme [B]	1.26	0.31	**<0.001**
Initial phoneme [K]	0.60	0.25	**0.016**
Trial number	-0.03	0.03	0.286
Age [older] * Condition [choice]	1.47	0.63	**0.020**
Age [older] * Trial number	0.09	0.04	**0.033**
Random Effects (SD)			
Subject (intercept)	1.736		
Condition by Subject	1.245		
Stimulus Length by Subject	1.918		
$N_{Subject}$	50		
Observations	3549		

number influenced younger and older adults differently as evident from the interactions between Age group and Condition, and between Age group and Trial number. More specifically, age groups differ less in accuracy in the choice than simple condition (see also Figure 4.3). Moreover, accuracy also increased more for older adults than younger adults over the course of the experimental trials within a block (see also Figure 4.4).

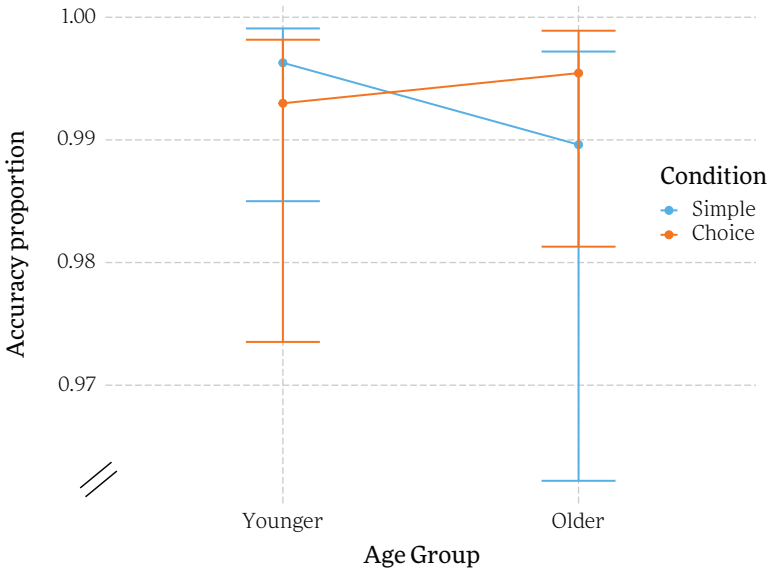


Figure 4.3
Model plot showing effects of age and condition (and their interaction) on production accuracy

Additionally, stimuli with 'B' (98.7% accurate) and with 'K' (97.6% accurate) as the initial phoneme elicited higher accuracy than stimuli with 'T' (96.0%). Moreover, we also verified that production accuracy for stimuli with 'B' as the initial phoneme was higher than for those starting with 'K' (by having initial phoneme 'B' mapped on the intercept).

4.3.3. Speech production latency

Results from the LME model on the speech onset latency data are displayed in Table 4.4 and illustrated in Figures 4.5, 4.6, and 4.7 below.

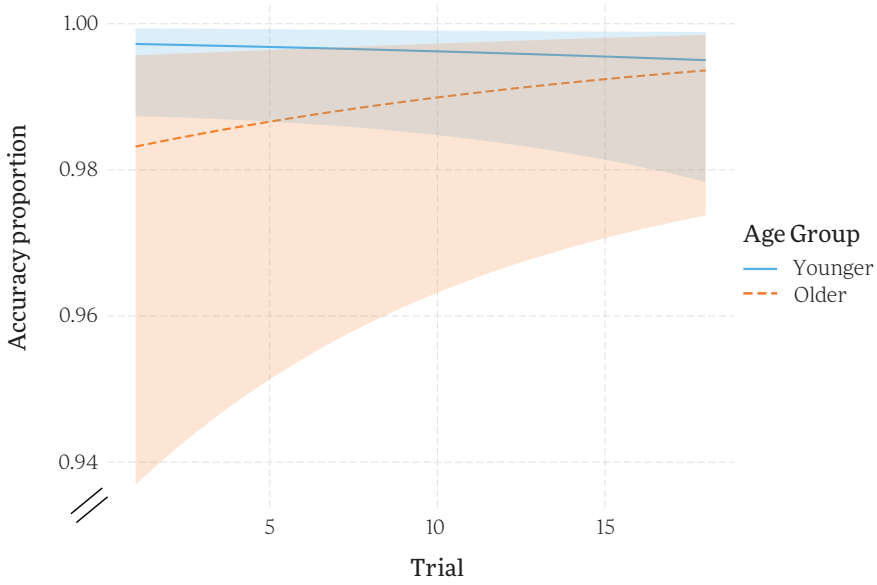


Figure 4.4
Model plot showing the effect of the interaction between age and trial number on production accuracy (shading displays 95% confidence interval)

Table 4.4 shows that older adults had longer speech onset latencies than younger adults (age effect), at least in the simple condition. Additionally, speech production latencies were longer for the choice than the simple condition (condition effect), but this effect was smaller for the older adults (as evident from the age by condition interaction, illustrated in Figure 4.5). Trial number as a simple effect was also significant, indicating that speakers had shorter speech onset latencies towards the end of the experimental blocks, at least in the simple condition. The insignificant interaction between trial and age group (for the simple condition) indicates that this 'practice effect' over the course of the simple condition blocks held for both age groups alike. However, this trial-related 'practice effect' was absent in the choice condition, at least for the younger adults. The three-way interaction between age group, condition, and trial number suggests that the older, but not the younger, adults provided increasingly fast responses over the course of the choice-condition blocks (see also Figure 4.6).

The significant interaction between condition and stimulus length suggests that participants (regardless of age group) slowed down more in the choice condition, compared to the simple condition, for disyllabic stimuli than for monosyllabic stimuli

Table 4.4 Results of the statistical model for speech onset latency (significant effects printed in boldface)

Predictors	Latency		
	Estimates	SE	p
(Intercept)	5.964	0.045	**<0.001**
Age [older]	0.257	0.062	**<0.001**
Condition [choice]	0.256	0.034	**<0.001**
Length [disyllabic]	0.020	0.018	0.266
Initial phoneme [B]	-0.139	0.008	**<0.001**
Initial phoneme [K]	-0.059	0.008	**<0.001**
Trial number	-0.005	0.001	**<0.001**
Age [older] * Condition [choice]	-0.148	0.047	**0.002**
Condition [choice] * Length [disyllabic]	0.121	0.013	**<0.001**
Age [older] * Trial number	-0.001	0.002	0.587
Condition [choice] * Trial	0.005	0.002	**0.006**
Age [older] * Condition [choice] * Trial	-0.005	0.003	**0.033**
Random Effects (SD)			
Subject (Intercept)	0.218		
Condition by Subject	0.141		
Stimulus Length by Subject	0.108		
Residual	0.193		
$N_{Subject}$	50		
Observations	3408		

(see also Figure 4.7). Lastly, stimuli starting with 'B' (*M* = 448 ms, collapsed across stimulus lengths) and 'K' (*M* = 477 ms) were produced faster than those starting with 'T' (*M* = 512 ms). These results echo with the results from production accuracy reported earlier, suggesting that stimuli 'Toe' and 'ToeKaa' were intrinsically more difficult to produce than 'Kaa', 'KaaBie', 'Bie', and 'BieToe'.

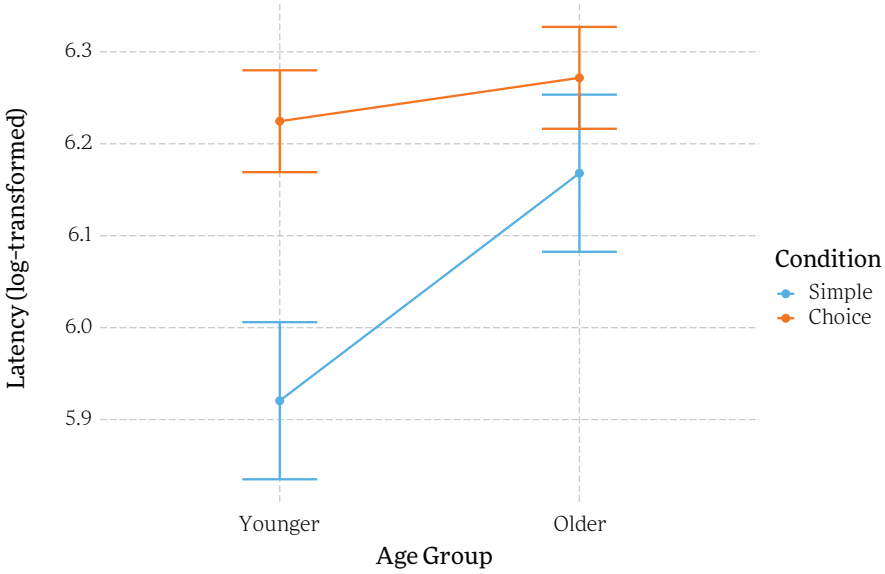


Figure 4.5

Model plot showing effects of age group and speech condition (and their interaction) on speech onset latency

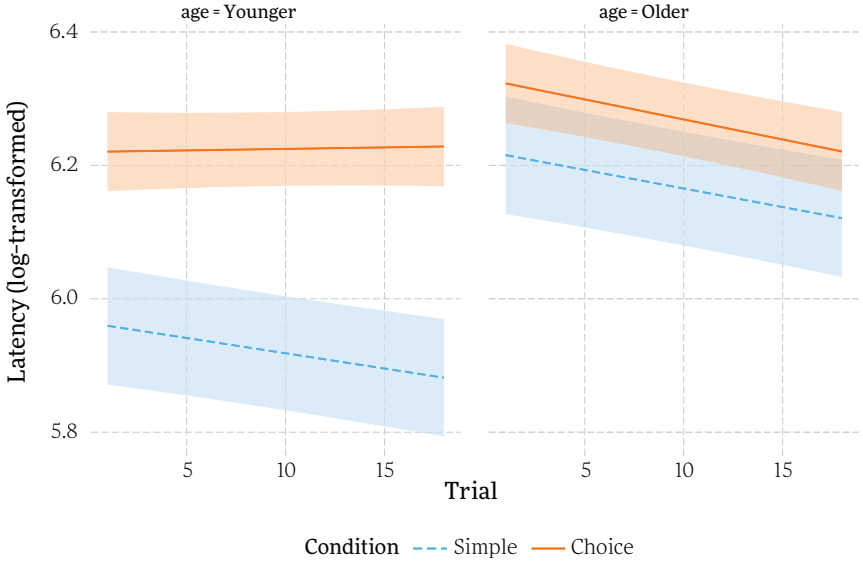


Figure 4.6

Model plot showing the three-way interaction between age, condition, and trial number on production latency (shading displays 95% confidence interval)

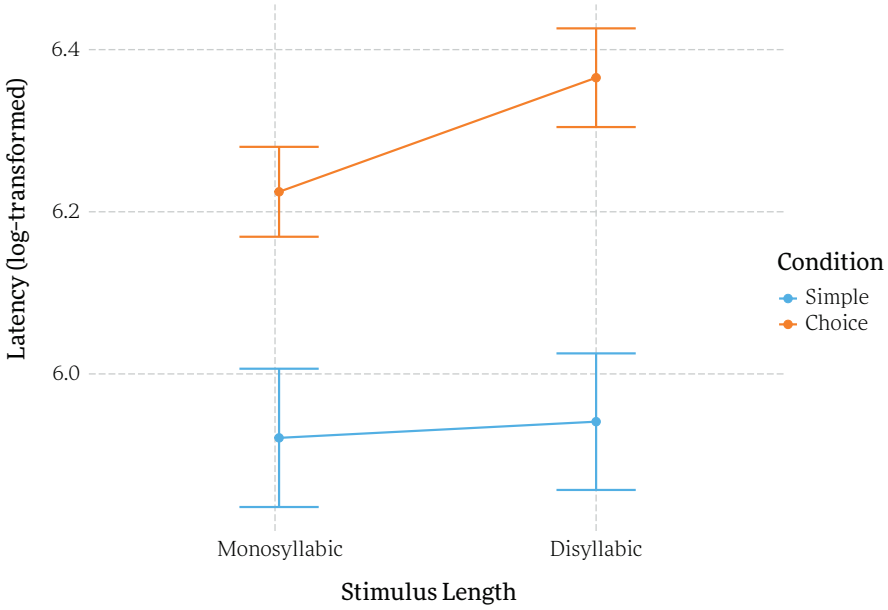


Figure 4.7
Model plot showing effects of condition and stimulus length (and their interaction) on speech onset latency

4.4. Discussion

The present study investigated which aspect(s) of speech motor planning is susceptible to age-related slowing using a speeded simple/choice speech production task. In the simple condition, speakers could have already prepared the stimuli they wanted to produce, and only needed to 'launch' or initiate their production. In the choice condition, speakers only learned upon the presentation of the critical-information cue which of three stimuli they had to produce (immediately). Hence, speakers still needed to select the correct motor programme and then programme their motor movements before they could initiate their production. As earlier work on speeded response tasks had shown stronger age effects in more complex conditions, we also expected larger age differences between younger and older adults in the choice condition than in the simple condition.

Contrary to our expectations, age differences in speech onset latencies were larger in the simple (prepared) than choice (unprepared) condition. Specifically, older adults tested in our study showed a smaller slowing effect in the choice (compared to simple)

condition than younger adults. In both simple and choice speaking conditions, longer RTs supposedly indicate additional processing (Maas & Mailend, 2012). The significant length effect found in choice RT but not in simple RT is consistent with what was found in other speech reaction time studies (e.g., Deger & Ziegler, 2002; Klapp, 2003). In other words, since both INT (i.e., organising the internal structure of a motor programme) and SEQ (i.e., selecting one or more motor programmes to form a smooth sequence) occur during choice RT, choice RT is more dependent on sequence length and complexity. Sequencing and initiating disyllabic sequences, on the other hand, does not seem to take more time than initiating the production of monosyllabic sequences in simple RT. The length effect observed in our study resembles those reported for complexity found in other (age-group comparison) studies on motor sequencing and control (e.g., Niermeyer et al., 2017; Tremblay et al., 2018) with longer/more complex stimuli being more error-prone (in the former study) and requiring more preparation than shorter/less complex stimuli (in both studies).

The significantly longer RTs in the choice rather than the simple condition found in both speaker groups indicate that both groups had to utilise more internal resources to prepare responses in the choice RT (a combination of INT and SEQ stages) than the simple RT condition (mainly reflecting the SEQ stage). These results speak to the findings from the large all-age population study in Der and Deary (2006), where choice RT was shown to be longer than simple RT across the adult age range. The condition effect observed in the latency analysis was not present in the accuracy analysis (for neither age group).

Our results also provide evidence for different learning patterns for the two age groups over the course of the experimental blocks. Whereas both age groups showed a decrease in response RTs over the course of the *simple* experimental blocks, only older adults seemed to speed up over the course of the *choice* condition blocks. This age difference in the 'practice effect' in the choice speaking condition could relate to the fixed order of the two speaking conditions, with the simple condition blocks always preceding the choice blocks (for monosyllabic and disyllabic stimuli alike). In other words, if older adults were to benefit more or benefit longer from practice, this shows up particularly in older adults' improved performance throughout the later (choice) blocks, reflecting that older adults may take longer to familiarise themselves with the target stimuli. Similar effects of prolonged practice effects for older adults are evident in speech production accuracy. Specifically, while accuracy remained more or less stable for younger adults throughout the experimental blocks, older adults actually improved their production accuracy over the course of the experiment blocks (see Figure 4.4). Consequently, the observation that age groups differed most in their latency behaviour in the 'prepared' (i.e., simple) condition may be due, at least

partly, to the fixed order in which conditions were administered. Older adults may have familiarised themselves better with the target stimuli by the time they had to perform the choice condition, such that age group differences had decreased, relative to the simple blocks.

Our primary goal was to find out which aspect of speech production (i.e., planning / programming or initiation of articulation) is more susceptible to age-related slowing. The two speaker groups tested in our study revealed an age group difference in the simple but not the choice speaking condition. Rather than the practice or familiarisation account provided above, this result could also be taken to suggest that older adults suffered more from age-related slowing in their buffer capacity. That is, older adults may be less able to hold the motor programmes of one or more syllables in the speech motor planning buffer before initiation of articulation compared to younger adults. Consequently, older adults seemed less prepared in the simple condition than younger adults or were less efficient in unpacking and launching the prepared motor programmes from the buffer, such that the difference between the *prepared* and *unprepared* condition is smaller for older adults.

A related explanation might be that the fixed planning time provided in the simple condition (500 ms between the information cue and GO cue) provides younger, but not older, adults with ample preparation time. However, this account seems rather unlikely given that older adults take on average 506 ms to start their production of monosyllabic targets in the choice condition, and 595 ms to start producing disyllabic targets. So even though younger adults may be argued to be more 'ready' than older adults to start their production while awaiting the GO cue in the simple condition, it seems unlikely that the 500 ms preparation time would not suffice for older adults. Nevertheless, the suggestion of age differences in speech preparation may relate to a finding mentioned earlier about an age-related change in the relationship between vocal RT and vocal response duration by Tremblay and colleagues (2018). They found that, in contrast to the younger and middle-aged adults who tended to start articulating before motor planning is completed, older adults seemed to need to completely assemble the required motor programme(s) before articulation. Follow-up research could investigate the relationship between speech onset latencies and target stimulus durations for our younger and older adults, to see whether the same age difference in starting to speak before planning is complete is observed in a simple/choice task paradigm.

Another possible account relates to our fast-response cut-off of 200 ms and potential strategic response-behaviour differences between age groups in the simple condition. Possibly, latency data in the simple condition is still influenced by some (young adult)

participants already initiating their response before the GO cue (with acoustic response onset shortly after the 200 ms after the GO cue). As we also removed responses that were 3 SDs faster than (i.e., below) each individual's average RT in the latency analysis, it is unclear which lower RT cut-off would be appropriate to exclude all responses that have been initiated before the GO cue. If older adults are more task-compliant than younger adults, our age results may not (only) speak to 'true' age differences affecting speech production processes, but also, or even mainly, to differences in (strategic) task behaviour. Ultimately, one could argue from our results that there are no age differences in speech production latency if there are no possibilities for strategic response behaviour, i.e., in the choice condition. When speakers have to go through the entire chain of processes (processing the information-bearing letter cue, retrieving the accompanying target stimulus from memory, and preparing it for articulation), younger and older speakers seem to be equally fast. If the total time for all component processes shows no age effect, it is difficult to argue that one of the component processes has been delayed in older adults. That is, the sum of processes could then only show no age effect if age slows down some process, while at the same time speeding up a different process (to the same extent). It is unclear from our results which particular processes would be affected by such hypothetical speeding and slowing.

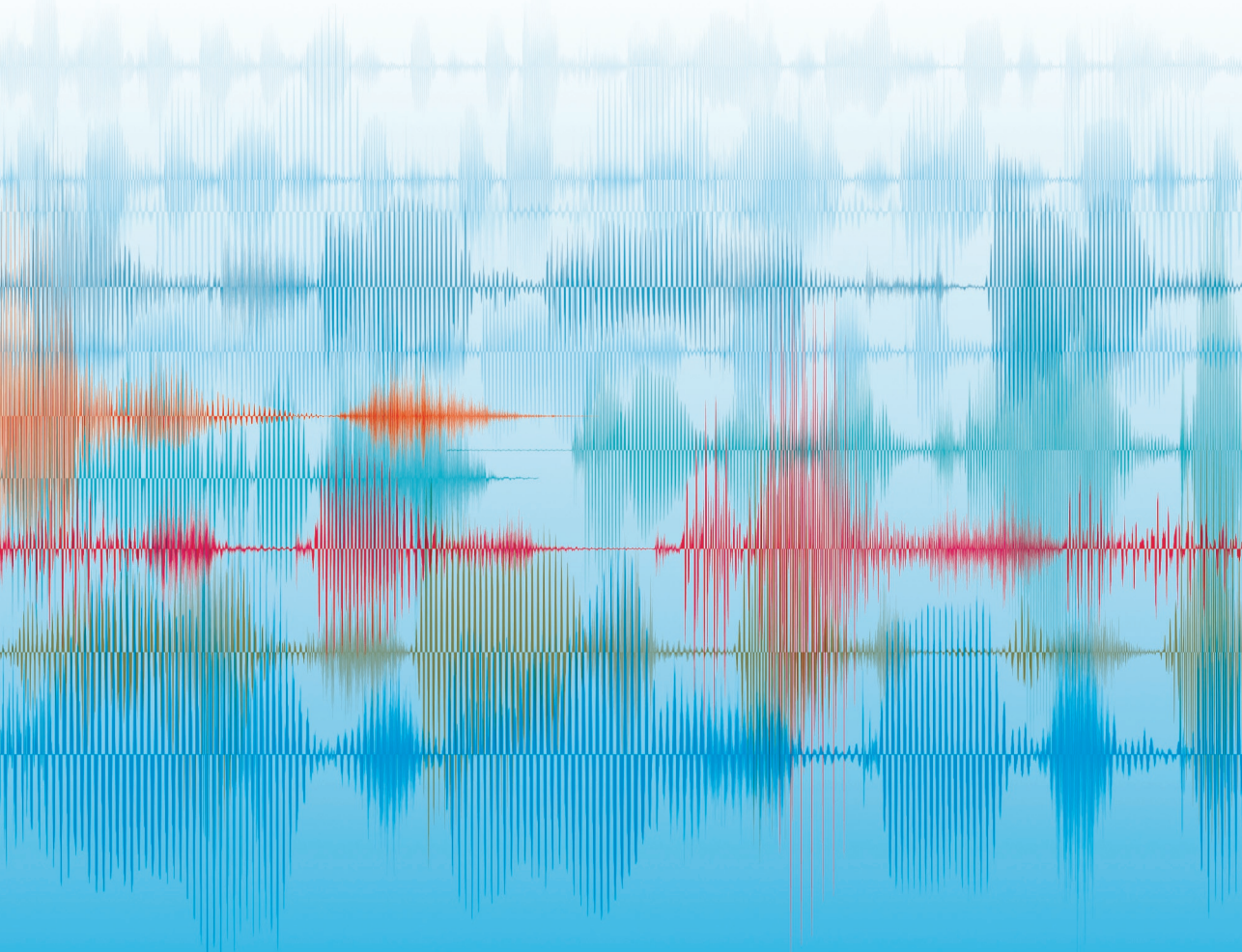
Relatedly, the absence of an age effect in the choice condition could also suggest that our target stimuli were relatively easy to plan and initiate. However, pilot testing had shown that relatively long practice blocks were necessary for participants to be able to quickly recall and produce the disyllabic target items (upon presentation of the initial letter). The task requirement of having to retrieve the target item from memory upon presentation of the letter cue therefore prevented us from including longer or more complex stimuli where age effects might have been more apparent (cf. Tremblay et al., 2019).

In sum, although the primary goal of the present study was to find out which aspect of speech production is more susceptible to age-related slowing, the results from the current study cannot answer this question adequately. The absence of stable age effects could relate to the confound in experimental design with simple blocks always preceding choice condition blocks, by the fact that the task introduced strategic response behaviour in the simple condition, and/or because the target stimuli were relatively easy to plan and produce. Conversely, our choice condition results could also be taken as good news in the sense that programming and initiation of relatively simple target items were not affected by age-related slowing. Follow-up research should preferably look for different experimental tasks and procedures to decompose the articulation processes of speech production.

5

Speaking Clearly in Noise: Consistency Over Time and Inter-speaker Differences

This chapter is based on the following:
Shen C., Cooke M., & Janse E. (in prep). Speaking clearly in noise: Consistency over time and inter-speaker differences.



Abstract

Individual speakers are often able to modify their speech to facilitate communication when necessary (e.g., when speaking in a noisy environment). Such vocal 'enrichments' might include reductions in speech rate or increases in phonetic contrasts. However, it is unclear how consistently speakers enrich their speech over time, and why some speakers seem to be better able to enrich their speech than others. In this study, we used an individual-differences approach to examine inter-speaker variability in the speech enrichment modifications speakers apply changing from a (baseline) habitual speaking style to a clear-Lombard speaking style (i.e., noise-induced Lombard speech with the additional instruction to speak clearly). The first research question was whether differences between the two speaking styles (as captured by different acoustic-phonetic features) would change over sentence trials. Additionally, the second research question was whether an index of speech motor control ability would relate to the speech enrichment modifications speakers apply. Lastly, the third research question was whether model-predicted intelligibility scores (quantified through the high-energy glimpse proportion metric, HEGP) change over sentence trials.

Seventy-eight young-adult participants read out sentences in both the habitual and the clear-Lombard speaking styles. Results showed that acoustic differences between speaking styles generally increased over trials (mostly non-linearly). This suggests that the speakers tested in our study needed some practice before realising their full speech enrichment potential when speaking clearly in noise (with reduced auditory feedback). Additionally, as observed in the acoustic measurements, model-predicted intelligibility scores also changed over sentence trials. Furthermore, speakers' speech motor control ability was found to relate to speakers' general articulation rate, but not to their model-predicted intelligibility scores, nor to their speech enrichment.

5.1. Introduction

Speech communication in our daily life regularly occurs in the presence of various sources of ambient noise at different intensity levels. Speech communication can happen at home with a TV playing in the background, or in a restaurant where many people are talking at the same time. Typically, speakers are able to enrich their speech to facilitate communication when necessary. In this paper, the term speech enrichment refers to any process by which talkers modify their speech to (attempt to) make it easier for listeners to process. While we know that speakers generally enrich their speech in noisy situations, it is not clear how consistently speakers apply these modifications over time, and why some speakers seem to be better able to modify their speech than others. This paper aims to address these two questions.

When speaking in the presence of (loud) noise, apart from linguistic changes in wording (e.g., paraphrasing), speakers often use certain strategies to enrich their speech in order to improve speech intelligibility, so that their interlocutors can better understand them (for a review, see Cooke et al., 2014). Lombard speech, elicited by having speakers speak in the presence of loud noise, is known to have characteristics that reflect speakers' increased vocal effort (e.g., Van Summers et al., 1988; Junqua, 1993). Additionally, Lombard speech exhibits acoustic features such as reduced articulation rate, raised fundamental frequency or F_0, enhanced 1-3 kHz frequency emphasis or flattened spectral tilt, and increased (mainly first and second) formant frequencies that lead to an expanded vowel space (e.g., Junqua, 1993; Bradlow et al., 1996; Cooke & Lu, 2010; Garnier & Henrich, 2014; Lu & Cooke, 2009; Tang et al., 2017; Tuomainen & Hazan, 2016).

There is some debate on whether Lombard speech is produced as some sort of reflex in response to loud noise, and to what extent Lombard speech might reflect intentional changes (e.g., Garnier et al., 2008). Despite this debate on the automatic or intentional nature of producing Lombard speech, studies have reported overlap in the acoustic features of Lombard speech and 'instructed clear speech', with the latter being clearly intended to overcome difficult communication situations (for a review chapter on clear speech, see Uchanski, 2008). More specifically, slower articulation rate and enhanced pitch modulation and vowel articulation (e.g., Smiljanić & Bradlow, 2005) are some of the overlapping features between clear speech and Lombard speech. Some of these acoustic-phonetic modifications of Lombard and clear speech have been shown to have perceptual benefits for hearing-impaired listeners, non-native listeners, as well as normal-hearing listeners, particularly under the listening conditions which the modifications were meant to counter (e.g., Bradlow et al., 1996; Bosker & Cooke, 2020; Ferguson & Morgan, 2018; Uchanski et al., 1996).

There are multiple factors that influence the type and amount of speech modifications speakers apply when speaking (clearly) in noise, such as speaker age (e.g., Amazi & Garber, 1982), type and intensity of background noise (e.g., Cooke & Lu, 2010), and the presence or absence of communicative intent (e.g., reading or monologue versus talking to an actual addressee or dialogue). For instance, Hazan and Baker (2011) investigated speakers' acoustic-phonetic characteristics communicating with a conversation partner under various noise conditions. They found that speakers modified their speech according to the different noise conditions to attend to their interlocutors' needs. In addition, Garnier and colleagues (2010) conducted a series of experiments to examine Lombard speech production in various sound immersion conditions (with different noise types) with and without interactions with a communication partner. They found that speech modifications in noise were greater when speakers were interacting with an interlocutor, suggesting that speech production (by adult speakers) is listener oriented. Results from the above-mentioned studies illustrate that speakers are generally able to modify the type and extent of acoustic-phonetic modifications in their speech under different speaking conditions, thus enriching their speech to make it better intelligible for their listeners.

Despite evidence that speakers can generally enrich their speech to make it more intelligible, large inter-speaker differences in speakers' speech intelligibility and in the amount of speech modifications in their clear/Lombard speech production have also been described (e.g., Junqua, 1993). Clearly, speech production and intelligibility can be influenced by the individual speaker's anatomy and physiology (e.g., the length and shape of the vocal tract), as well as by the linguistic environment that a speaker has grown up and/or lived in, in relation to the listener's (Hughes et al., 2012; Moyer, 2013). Beyond such speaker-specific (anatomical or regional/accentual) characteristics, several studies have also provided acoustic measures to describe what makes a speaker highly intelligible even when speaker characteristics and speaking conditions are highly controlled. For instance, in a study by Bond and Moore (1994), a series of intelligibility tests were used on both native and non-native listeners. One particular speaker out of the five speakers that they tested was consistently less intelligible, as judged by both native and non-native listener groups. Further acoustic-phonetic analyses revealed that, compared to the more intelligible speakers in their sample, the less intelligible speakers spoke more rapidly, exhibited reduced vowel duration as well as vowel space, used minimal cues for consonant contrasts, and varied more on the amplitude of stressed vowels (Bond & Moore, 1994).

A similar larger-scale study by Bradlow and colleagues (1996) investigated the relationship between intelligibility and speaker-specific acoustic-phonetic characteristics based on data from 20 speakers. They found that F_0 range and vowel space were

positively correlated with speech intelligibility (Bradlow et al., 1996). Additionally, Hazan and Markham (2004) found total energy in the 1-3 kHz region and word duration to be important predictors of word intelligibility in noise. Clear speech adaptations are perceived and appreciated by young normal-hearing listeners and by older hearing-impaired listeners, for whom these adaptations were intended (Ferguson & Morgan, 2018). In line with results by Bond and Moore (1994) that intelligibility similarly affects native and non-native listeners, Bradlow and colleagues (2018) showed, at the speakers' end, that speaker-specific traits that help improve speech intelligibility do not depend on the language the speaker uses. In other words, speakers with high intelligibility scores in their first language also had high intelligibility scores in their second language, indicating the existence of some individual-level language-independent articulatory control mechanism in their speech production (Bradlow et al., 2018).

Relatively little is known about how speakers differ in their speech enrichment (e.g., if and how speakers use various enrichment strategies and how effective these are). One counterintuitive finding is that speech produced by speakers with *little* (self-reported) experience communicating with people with hearing loss was observed to be clearer and more intelligible than that of speakers with either *no*, *some* or *frequent* prior experience (Ferguson & Morgan, 2018). This observation questions the learnability of clear speaking style, or speakers' willingness to learn it. More importantly, few studies have looked at the consistency with which speakers apply their speech enrichment over time (but cf. Lee & Baese-Berk, 2020). Do speakers get better at enriching their speech with more practice? Or conversely, do they decrease the size or extent of their modifications over time (e.g., due to fatigue)? To answer these questions, a better understanding of the inter-speaker variability in speech enrichment over time is needed.

This study set out to investigate the consistency with which speakers apply enrichment modifications in a speaking style where they were presented with loud noise through headphones and were explicitly instructed to speak clearly (i.e., a clear-Lombard style) over time. Additionally, it aims to relate speakers' speech characteristics to their speech motor control ability (defined as the agility with which speakers can move their articulators), and to explore whether speech motor control ability relates to the consistency (over sentence-reading trials) of speakers' habitual and clear-Lombard speaking styles.

5.1.1. Consistency of clear-Lombard speech modifications over time

Some studies on prolonged voice usage and vocal fatigue have observed links between increased vocal effort and vocal fatigue and henceforth decreased vocal function (Bottalico et al., 2016; Solomon, 2008). Given that Lombard speech exhibits acoustic- and articulatory-phonetic modifications that require increased vocal effort, the results from those studies on prolonged voice-usage and vocal fatigue could suggest a decrease in the amount of modifications that speakers apply when producing the effortful Lombard speech over an extended period of time. On the other hand, many speakers are able to maintain an extended conversation in noisy environments such as busy bars or restaurants. Thus rather than showing vocal fatigue and slackening of attention in the Lombard speaking style, speakers may show some learning or practice effect over time instead. If speakers improve with practice, this also raises the interesting question of what information they actually use to improve the clarity of their own speech, if they cannot hear their own speech well because of loud background noise (e.g., played over over-ear headphones, such that auditory feedback is drastically reduced). Little research has directly measured the consistency of the acoustic- and articulatory-phonetic changes in Lombard speech over a certain period of time, or has related it to the consistency of speakers' habitual speaking style. However, a recent study investigated speakers' use and *maintenance* of clear speech in an interactive speech task with interlocutors (Lee & Baese-Berk, 2020). In this naturalistic 'spot-the-difference' (Diapix) task, where speakers have to describe their version of a pictured scene to their interlocutor who has a slightly different version of the picture, speakers were not explicitly instructed to use clear speech. Nevertheless, native-English speakers' speech was found to be more intelligible when their interlocutors were non-native (rather than native) English listeners, and was found to be more intelligible in the early portion of each conversation. Speakers in their study were found to 'reset' to clear speech whenever they started their description of a new picture. Thus, Lee and Baese-Berk (2020) concluded that the initiation of clear speech could be listener oriented while the (lack of) maintenance of clear speech is perhaps speaker driven, as speakers can gradually spend less articulatory effort over the course of their conversation to still be understood and then 'reset' their speech clarity at topic boundaries.

In our study, we investigated the consistency aspect of speakers' speech enrichment. In order to have comparable speech materials (in terms of content and amount of materials) across speakers, we used a sentence reading task with instructed clear speech production rather than spontaneous (less controlled) speech. Specifically, we analysed the acoustic- and articulatory-phonetic modifications that speakers apply moving from a habitual reading style to a clear-Lombard reading style. Our main research question was whether the acoustic difference between habitual and

clear-Lombard speech would increase or decrease over the course of an experimental session (i.e., over a sentence list). In line with Lee and Baese-Berk (2020)'s results, we could find that speakers would decrease their clarity gradually. Alternatively, as our speech task may not constitute a discourse context, speakers may maintain their clear-Lombard speech throughout the experimental session.

5.1.2. Relationship between speech motor control and speech enrichment

As mentioned above, even homogeneous groups of speakers differ in how intelligible their speech can be, due to speaker-specific speech characteristics such as slower articulation rate and less reduced vowels. This variability may relate to speakers' inherent articulatory/speech motor control abilities.

For instance, Lively and colleagues (1993) found considerable inter-speaker variability in individuals' speech acoustics (e.g., amplitude, spectral tilt, and speaking rate) when speaking under cognitive workload (i.e., a compensatory visual tracking task), whereas these speakers differed less in their speech acoustics in a control condition. Lively and colleagues observed differences in intelligibility across speakers as a result of the variable acoustic modifications that speakers applied when speaking under cognitive workload. This suggests that increased (cognitive) demands affect speech intelligibility of some speakers more than that of others. Additionally, Walsh and Smith (2002) examined the development of speech motor control in adolescents using a measure that reflects consistency of trajectories of various articulators (e.g., the lips and jaw) during speech production. They found that articulatory trajectories differed among their adolescent as well as young-adult participants. The results from these studies suggest that speakers' speech production (at least in a young non-clinical population) may be modulated by some underlying (idiosyncratic) motor control mechanism. We therefore further explored the potential role speech motor control plays in our healthy young adult speakers' speech enrichment modifications.

Speech motor control (or articulatory control) is a term often used to refer to "the systems and strategies that regulate the production of speech, including the planning and preparation of movements and the execution of movement plans to result in muscle contraction and structural displacement" (Kent, 2000, p. 391). One straightforward way to quantify speech motor control in clinical settings is through maximum performance speech tasks, for instance, using a so-called diadochokinetic or DDK task. Through measuring maximum repetition rates (and sometimes also accuracy) in sequential (e.g., papapa...) and/or alternating (e.g., patakapatakapataka...) non-words or real words, this task captures speakers' speech motor limitations (e.g., Maas, 2017), henceforth providing an index of speakers' speech motor control. Although DDK is

typically used in clinical settings, comparing patient populations to controls, we have shown in previous work that young adult speakers vary considerably in their DDK performance, both in terms of rate and accuracy (Shen & Janse, 2020, see also Chapter 3). As a first indication that DDK performance (as an index of articulatory skills) relates to speech behaviour, de Jong and Mora (2017) observed a relationship between speakers' DDK accuracy and their speech fluency in both of their L1 and L2 speech. In this study, we investigated whether these maximum speech performance measures relate to (flexibility of) speech behaviour.

Through exploring the potential role speech motor control plays in speakers' speech enrichment modifications, we aim to test whether those with better speech motor control (as indexed by higher DDK maximum speech rates and better DDK accuracy scores) apply larger and/or more consistent clear-Lombard acoustic speech modifications when changing from a habitual to a clear-Lombard speaking style.

Additionally, gender differences in speakers' speech enrichment success (as reflected by intelligibility) have been observed in multiple studies across different speaking styles (e.g., Junqua, 1993; Bradlow et al., 1996). More specifically, female speakers have been demonstrated to exhibit larger clear speech benefits than male speakers when their speech was presented in noise, as judged by both children (Bradlow et al., 2003) and adult listeners (Bradlow & Bent, 2002). Although not the main focus of the current study, we examined possible effects of gender on speech enrichment as well. However, given our gender-imbalanced sample, gender effects should be interpreted with caution.

5.1.3. HEGP model-predicted speech intelligibility scores

Apart from analysing the acoustic changes in speakers' speech enrichment modifications, it is also critical to understand how different speech enrichment modifications contribute to speech intelligibility. One way of obtaining speech intelligibility scores is to collect listeners' subjective responses. However, the process of acquiring subjective intelligibility responses can be quite time-consuming and resource-demanding, especially when the number of speakers to be assessed is very large. Given the constraints on human (subjective) intelligibility responses, several objective intelligibility measures have been proposed. For instance, the articulation index (Kryter, 1962a, 1962b) and the speech-transmission index (Steeneken & Houtgast, 1980) were the most commonly used metrics in earlier studies. More recently, a glimpse-based speech perception model was proposed (Cooke, 2006). This model uses an internal automatic speech recognition component to recognise the speech-domain spectro-temporal regions, or 'glimpses', in speech that survives energetic (noise) masking, attempting to model human speech perception in noise (Cooke, 2006). Subsequently, several studies have used the output of the initial stage of the glimpsing

model, which is the amount of supra-threshold target speech surviving energetic masking, or 'glimpse proportions', as a proxy for intelligibility, aiming to predict speech intelligibility without the need to construct an automatic speech recognition system for each task (e.g., Tang & Cooke, 2012; Valentini-Botinhao et al., 2012). Four glimpse-based metrics have been evaluated to compare their capability of accounting for the subjective intelligibility of a variety of speech styles in a range of masker types (Tang & Cooke, 2016). Of the four, the high-energy glimpse proportion (HEGP) metric had the highest correlations with human listeners' judgement (r values between .87 and .92) across the tested datasets. It is important to note that the HEGP metric (like many other acoustic metrics) only measures the contribution of energetic masking to intelligibility. Any higher-level linguistic clarification that speakers may apply in their clear-Lombard speech (in the form of enhanced phonetic contrasts) are thus not captured by these metrics.

Given the large number of speakers (N = 78) involved in our current study, we opted for this robust automatic speech intelligibility-in-noise metric, HEGP, to obtain objective intelligibility ratings of both habitual and clear-Lombard speech that our speakers produced. The HEGP-model predicted intelligibility ratings will be used to explore the relationship between speech intelligibility (i.e., intelligibility in relation to release from energetic masking) and the acoustic changes speakers apply in their speech enrichment modifications. Additionally, we will examine how consistent these model-predicted intelligibility scores are over sentence trials, and whether speaker characteristics, particularly speech motor control, relate to (the consistency of) their speech intelligibility.

5.1.4. The present study

In this study, we used an individual-differences approach to examine inter-speaker variability in the speech enrichment modifications speakers apply going from baseline (habitual) to clear-Lombard speech production. More specifically, through analysing changes in four acoustic-phonetic features (i.e., articulation rate, median F_0, F_0 range, and spectral balance), we investigated how consistently speakers apply their speech enrichment modifications over sentence trials (RQ 1). We then investigated whether speaker characteristics, such speech motor control ability, relate to the speech enrichment modifications they apply, and whether these speaker characteristics predict (the consistency of) speakers' speech modifications over sentence trials and in the two speaking styles (habitual versus clear-Lombard) (RQ 2). Lastly, we investigated model-predicted intelligibility of speakers' habitual and clear-Lombard speech through an intelligibility metric (the high-energy glimpse proportion, or HEGP) that attempts to model human speech perception in noise. We examined the relationship between the acoustic changes in speakers' (enriched)

speech and model-predicted speech intelligibility, the consistency of the model-predicted intelligibility scores over sentence trials, and the association of speaker characteristics, especially speech motor control with their model-predicted speech intelligibility scores (RQ 3).

5.2. Methods

5.2.1. Participants

A total of 78 native Dutch speakers (age: *M* = 22;7 [years;months], *SD* = 2;10, 61 females) were recruited online through the Radboud Research Participation System. Note that these 78 participants were recruited for an extensive data collection session containing multiple tasks to generate data for multiple studies including the current study (Shen & Janse, 2019, see also Chapter 2; and Shen & Janse, 2020, see also Chapter 3). All of the participants were university students or had graduated from university at the time of the experiment. Participants did not report any known history of speech, hearing, or reading disabilities, nor past diagnosis of speech pathology or brain injury at the time of testing. Additionally, they all reported to have normal or corrected-to-normal vision. The study protocol had been evaluated and approved by the Ethics Assessment Committee Humanities at Radboud University. All participants gave informed consent for their data to be analysed anonymously, and they received either course credits or gift vouchers as compensation for their time.

5.2.2. Speech tasks

All 78 speakers performed a sentence-reading task and a speech motor control task (a Diadochokinesis or DDK task) in a sound-attenuating recording booth at the Centre for Language Studies Lab of Radboud University. Written stimuli for both tasks were presented on a 24" full HD monitor placed on a table in front of the participant. The presentation of the task stimuli was controlled real-time by the experimenter on the stimulus computer outside the recording booth. Speech recordings were made using a Sennheiser ME 64 cardioid capsule microphone placed around 15 cm away from the speaker's mouth through a pre-amplifier (Audi Ton) onto a Roland R-05 WAVE recorder. The sampling rate of the resulting .wav format recordings that were used for acoustic analysis was 44.1 kHz with 16-bit resolution.

5.2.2.1. Sentence reading task - Radboud Habitual and Lombard Speech corpus (RaLoCo)

A multi-talker speech corpus – the Radboud Habitual and Lombard Speech Corpus (RaLoCo) was created for the purpose of this study (cf. Appendix B). The RaLoCo corpus contains Dutch sentence reading material read by the 78 native Dutch speakers, with 96 sentences per speaker (i.e., 7488 sentences in total). Each participant read out 48 unique sentences twice: once in a habitual style in which they were only told to read out the sentences fluently (i.e., habitual speaking style), and once in a style where they were instructed to read the sentences out as clearly as possible while hearing speech-shaped noise continuously over headphones (i.e., clear-Lombard style). The speech-shaped noise file they heard was created based on an average speech spectrum envelope across a balanced mix of male and female voices. The total duration of the noise file was around 25 minutes, which proved long enough to cover the noise-condition part of the sentence-reading session. Note that there was a brief break between reading of the two conditions, as participants needed to read through the instructions for the clear-Lombard style. After the participant indicated that they had understood the clear-Lombard instructions, the experimenter played the noise file at 78 dB SPL (as calibrated using a Brüel & Kjær Type 4153 artificial ear), through a HP Probook laptop over a pair of closed headphones (Sennheiser HD 215 MKII DJ), to the participant.

Of the 48 unique sentences, half (i.e., 24 sentences) had a keyword noun containing one of the three corner vowels (i.e., /i/, /u/, /a/) embedded in the sentence. An example keyword sentence is 'Mijn opa had de *piep* jammer genoeg niet meer gehoord (translation: Unfortunately, my grandfather hadn't heard the *beep* anymore), in which 'piep (beep)' is the corner-vowel target word. Speakers were presented with 48 unique sentences in one of eight random orders, followed by the same 48 sentences in a different random order. During the reading task, the habitual style always preceded Lombard style for all participants to avoid potential spill-over effects from Lombard to habitual speech production. The four long lists (differing only in order) were rotated over participants such that trial effects could be isolated from sentence (i.e., item) effects.

The production of the sentences was live monitored for hesitations, misarticulations, and other disfluencies. If the experimenter noticed a disfluency, they would ask the speaker to repeat the target sentence (by re-presenting the sentence stimulus in the slide show). Overall, participants were able to read out all sentences fluently, with only occasional repetitions of one or two sentences for some speakers.

5.2.2.2. Speech motor control task

To capture individuals' speech motor control capabilities, we used the DDK task that was set up as a maximum performance speech task using alternating non-word sequences (Duffy, 2013; Wang et al., 2004). The non-word sequences were the standard 'pataka' /pataka/ (the standard alternating sequence in clinical assessment), and the reverse-order 'katapa' /katapa/. Target sequences were presented to participants at the centre of the screen. Participants were instructed to repeatedly produce the presented sequence as accurately and as rapidly as possible for about 10 seconds. Before the task began, a short audio clip of a pre-recorded example was played to each participant. Participants were then presented with three mono-syllabic stimuli ('pa', 'ta', 'ka') and two di-syllabic stimuli ('pata', 'taka') as practice stimuli. The task requirements of accuracy and speed of repetition were constantly shown as a reminder on top of each stimulus slide.

Participants' speech motor control ability was quantified through their maximum performance in (averaged) rate and accuracy in the DDK task, averaged over the 'pataka' and 'katapa' sequences (Shen & Janse, 2019, see also Chapter 2). More specifically, individual DDK articulation rate (syllables/sec) was calculated by multiplying the total number of accurate and fluent non-word repetitions produced by each participant in the first 7-second (or as close to 7-second as possible for the repetition counts to be an integer) time window of the DDK utterance, and then dividing this number of total syllables by the actual production time (total duration minus erroneous and disfluent repetitions, as well as in-breaths and pauses longer than 200 ms between repetitions). DDK accuracy (fraction correct) was calculated as number of accurate and fluent repetitions divided by number of all repetitions in the same 7-second time window. A repetition was only counted as correct if it did not contain any form of articulation errors or disfluencies (e.g., pauses longer than 200 ms) within the sequence.

5.2.3. Acoustic measures

In order to analyse the type of modifications speakers employed producing the two types of speech, we used standard Lombard speech acoustic measures (e.g., articulation rate, pitch/F_0 measures, spectral balance measures, and vowel space measures) as reported in previous studies (e.g., Garnier et al., 2010; Cooke & Lu, 2010; Lu & Cooke, 2008). The long audio recordings from the sentence-reading task were labelled and extracted as separate sentence-length audio files using a customised Praat script. Long silences (>200 ms) in the extracted sentences were manually labelled and then removed in Praat for sentence-level acoustic measures.

5.2.3.1. Sentence-level acoustic measures

Articulation rate, F_0 measures (median F_0 and F_0 range), and spectral balance were calculated using the sentence-length audio files. Articulation rate was calculated as syllables per second, through dividing the number of syllables in the orthographic transcription of a sentence by the actual production duration of that sentence (excluding pauses that are longer than 200ms). Higher values index faster rates.

F_0 measures were obtained using a customised script in Praat (Boersma & Weenink, 2017) that calculates estimated F_0 values at 10 ms intervals in each individual sentence. The pitch floor was set at 75 Hz and 60 Hz for female and male speakers respectively, while the pitch ceiling was 500 Hz and 300 Hz respectively. The script coded unvoiced parts of the audio as '-1', and these values were subsequently excluded from further analyses. 'Doubling' or 'halving' errors in pitch tracking (shown as sudden jumps in the estimated F_0 values) were corrected using a customised Python script that detects and deletes values that are above or below a factor of 1.5 compared to the penultimate value (cf. Marcoux & Ernestus, 2019). Median F_0 and F_0 range measures were then calculated per sentence based on the remaining (cleaned) pitch values (i.e., on the remaining 96.32% of the data points). Note that instead of using the 100% range for calculating F_0 range, we opted for the range between the 10th and the 90th percentile from the original F_0 values to avoid extremely high and low values caused by erroneous pitch values that might not have been excluded by the Python script. The observed Hz values for median F_0 and F_0 range per sentence were then converted to semitones with 1 Hz as the reference using this formula: s =12 * log2(targetHz) (e.g., Hazan & Baker, 2011; Dichter et al., 2018). Higher values for median F_0 index higher or raised voice, and higher numbers for F_0 range indicate expansion of F_0 range.

Spectral balance was calculated using the Hammarberg Index (Hammarberg et al., 1980), at the level of individual sentence. The Hammarberg index captures the relative difference between the energy maxima in the low frequency range (0-2000 Hz) and the high frequency range (2000-5000 Hz). The Hammarberg values were automatically extracted using a customised Praat script by extracting the long-term average spectrum (LTAS) of each sentence with a filter bandwidth of 100 Hz, and subtracting energy maxima in the low frequency range from that in the high frequency range. Higher values indicate a steeper spectral roll-off, which implies less vocal effort.

5.2.3.2. Speaker-level vowel space measures

As mentioned earlier, out of the 48 unique sentences read out by speakers in habitual and clear-Lombard style, 24 sentences contained a monosyllabic target (noun) keyword with one of the three Dutch corner vowels /i/, /a/, and /u/ (six keywords/sentences per vowel). All vowels were selected to have a cV(c)c structure with 'c'

being obstruent consonants to keep vowel segmentation relatively straightforward. Additionally, the 24 keywords were matched for frequency and phonological neighbours. The corner vowels in those 24 keywords were segmented manually in Praat following segmentation rules recommended by Machač and Skarnitzl (2009) and the acoustic properties of vowels and consonants described in Ladefoged and Disner (2012). For instance, to identify the onset and the offset of the target vowel, special attention was paid to select the relatively stable parts of the first two formants. This was done to minimise potential influence from coarticulation with the surrounding consonants. After segmentation, the first two formants of the vowels were extracted at three time points using a customised Praat script: one point at the vowel midpoint, one point at 15 ms before, and one point at 15 ms after the midpoint. For the formant measurement settings, the maximum formant was set at 5500 Hz with 5 formants and 5000 Hz with 5 formants for female and male speakers respectively. The window length was set at 25 ms, and the time step at 10 ms. The extracted formant frequencies in Hz were then transformed into the perceptually motivated Bark scale (cf. Traunmüller, 1990). The converted Bark values were used for speakers' vowel space calculations. Specifically, the convex hull vowel area encompassing all vowel tokens per speaking style per speaker was calculated using the 'phonR' package (version 1.0-7) (McCloy, 2016) to capture changes in vowel space associated with a potential change in articulatory behaviour from habitual to clear-Lombard speaking style. Larger numbers indicate an expansion in the vowel space.

5.2.4. HEGP model-predicted intelligibility

Having analysed the acoustic features of the speech produced in both habitual and clear-Lombard speaking styles, we further examined the (model-predicted) intelligibility of the two types of speech produced by the speakers.

HEGP-metric predicted intelligibility scores were obtained per sentence (for the 48 unique sentences) in both speaking styles (habitual and Lombard) for all speakers using a customised script from Tang and Cooke (2016). The same speech-shaped noise used for clear-Lombard speech elicitation was used as the added noise-masker for the HEGP calculation at -5 dB signal-to-noise ratio. This SNR was chosen because the differences in intelligibility between habitual and Lombard speech were largest around this SNR level. Following Tang and Cooke (2016), HEGP scores were calculated by first computing all raw 'glimpses' defined as time-frequency regions where the energy of the target speech exceeds that of the masker, then selecting the subset of 'high-energy' glimpses, defined as those whose energy exceeds the mean speech-plus-masker energy, measured independently for each frequency region. The HEGP scores lie between 0 and 1, with higher numbers indicating a higher glimpse proportion escaping energetic masking, thus suggesting higher predicted intelligibility.

In order to verify that the HEGP-predicted intelligibility ratings would produce reliable objective ratings for our speech material, we compared the HEGP-predicted ratings to human listening effort ratings. While intelligibility, even if derived from human transcription of the speech, is not the same as listening effort, research has shown that intelligibility and listening effort (both as perceived by human listeners) are highly related (Krueger, Schulte, Zokoll et al., 2017). To verify the relationship between model-predicted intelligibility and human ratings of listening effort for our materials, we selected a subset of eight speakers (four females and four males) who had a relatively 'neutral' accent in Dutch to limit the possible influence that noticeable/unfamiliar accent might have on listeners' listening effort judgement.

32 normal-hearing young (18 - 30 years) native-Dutch female listeners (who had not previously participated in the main speech production experiment) completed the listening experiment using the online survey software Qualtrics (Qualtrics, Provo, UT). Each listener rated a total of 48 unique sentences produced by the eight selected speakers (3 unique stimulus sentences per speaker), in the two speaking styles (24 sentences from the habitual and 24 other sentences from the clear-Lombard style). Four counter-balanced speaker-lists were rotated over the 32 listeners, and as a result, a total of 1536 ratings were obtained. The sentences presented to the listeners were embedded in the same speech-shaped noise that was used to elicit the Lombard speech at -6 dB SNR (note that although there is a slight (1 dB) difference in SNRs between this perception experiment and the aforementioned HEGP-metric predicted intelligibility ratings, the differences between -5 dB and -6 dB SNR in HEGP scores are small).

The listeners were instructed to rate the amount of listening effort that they experienced listening to the speech stimulus presented in noise. Ratings were given on a scale from 1 (no effort) to 7 (extreme effort) using an adapted version of the Adaptive Listening Effort Tests developed by Krueger and colleagues (Krueger, Schulte, Brand et al., 2017). For each of the available speaker-style combination, an average human rating was calculated based on ratings from eight human listeners.

Pearson correlation coefficients were calculated (at individual item level, i.e., for each combination of speaker, sentence, and speaking style) between the average subjective ratings (obtained at -6 dB SNR) and the HEGP-predicted intelligibility scores (obtained at -5 dB SNR). A significant and high correlation was found between the subjective ratings and the HEGP scores (r = -.81, p < .001). In other words, higher predicted intelligibility was associated with lower degrees of listening effort, thus arguing for the use of HEGP-model predicted intelligibility scores as a proxy of speech intelligibility in noise.

5.2.5. Consistency of speaking style over time and the role of speech motor control

To answer the main research questions set out for the current study, a number of linear mixed-effects (henceforth LME) models were run in RStudio (version 1.2.1335), using the lme4 package (version 1.1-21) (Bates et al., 2015). Detailed data modelling procedures are described below.

The consistency of the acoustic modifications that each speaker applied throughout the habitual and clear-Lombard speaking styles over the 48 sentences or trials (i.e., trial represents the sentence's position in the sentence list/experimental session) could only be investigated for sentence-level acoustic measures, as averaging (across sentences) was applied for the vowel space measure. Consistency of speaking style was investigated for each sentence-level acoustic measure, i.e., articulation rate, median F_0, F_0 range, and spectral balance, separately. Additionally, in order to check for potential non-linear trial effects, we added the quadratic trial term (i.e., trial squared) to each of the four LME models (Bruce & Bruce, 2017). Specifically, for each dependent acoustic measure, one LME model with and one without a quadratic trial term were set up (always in addition to a linear trial effect). For the models with quadratic trial terms, trial, trial squared (i.e., quadratic term of trial), and speaking style (habitual and Lombard) were entered as fixed effects of interest. Additionally, speakers' individual speech motor control abilities (as indexed by DDK rate and accuracy) were entered as fixed effects of interest to explore the potential association between speech motor control and speakers' speech enrichment. Gender of the speaker (female or male) was included as a fixed control predictor. As for the models without the quadratic term of trial, trial squared was removed while all the other predictors remained unaltered.

Interactions between speaking style and trial, and if applicable also between speaking style and trial squared, were included to answer RQ 1 on the consistency of speech enrichment applied by speakers when changing from habitual to clear-Lombard speech over the experimental session. Additionally, interactions between speaking style and speaker characteristics (DDK rate, DDK accuracy, and gender) were also included to answer RQ 2 on the association of speaker characteristics (i.e., speech motor control ability and gender) and the size of speaker's speech enrichment applied moving from habitual to clear-Lombard speech. Across all models, participant and item (i.e., sentence) were included as two random effects. we also allowed a random by-participant slope for speaking style and trial, acknowledging that individual participants may differ in the acoustic changes moving from their habitual to clear-Lombard speaking style and may differ in their behaviour over sentence trials.

Model comparisons using the anova() function in R between the full models with and without the quadratic trial term (and its interactions) were applied. The full models with better fit were then selected for stepwise model stripping to arrive at the most parsimonious models. More specifically, we took out insignificant interactions first, and then removed insignificant effects, starting with the ones with the lowest t-values in the model outputs. Model comparisons were applied after each removal of the least significant predictor or interaction term to verify that the removal of each predictor/interaction term did not result in significant loss of model fit.

Additionally, one simple linear regression model was run to test whether individual speech motor control related to the extent to which speakers clarified their speech in terms of vowel space. Specifically, the averaged vowel space measures per speaker were included as a dependent variable, with speaking style (habitual and Lombard) and DDK performance (non-word rate and accuracy) entered as fixed effects of interest. Gender of the speaker (female and male) was entered as a fixed control predictor. Similarly, model-stripping was applied in a stepwise manner to arrive at the most parsimonious model, with model comparisons applied after each removal of the least significant predictor.

5.2.6. Consistency of HEGP-model predicted intelligibility over time and the role of speech motor control

Having verified the relationship between HEGP-metric predicted intelligibility and human listeners' subjective ratings of listening effort for a subsample of our materials, we now return to our third research question on the consistency of sentence-level HEGP-model predicted intelligibility scores over the course of our experimental session, and the potential relationship between HEGP-predicted intelligibility and speakers' speech motor control abilities. One LME model was used to test the two elements of this question. Prior to fitting the model, the raw HEGP scores for each unique sentence token were converted to logits using the following equation: logit = ln (p/1-p) (Jaeger, 2008). Similar to the set-up of the LME models for the four sentence-level acoustic measures above, the quadratic term of trial (i.e., trial squared) was also added to the LME model for HEGP logit scores. Model comparison between the full models with and without the quadratic term of trial was applied. The full model with better fit was then selected for stepwise model stripping to arrive at the most parsimonious model. Specifically, we included sentence-level HEGP-model predicted intelligibility scores (in logits) as dependent variable. For the model with the quadratic term of trial: trial (i.e., sentence number), trial squared (i.e., quadratic term of trial), and style (habitual and Lombard) were entered as fixed effects of interest. Speakers' individual speech motor control abilities (as indexed by DDK performance) were also included as fixed effects of interest to investigate the

association between speakers' speech motor control and speakers' speech intelligibility. Gender of the speaker (female and male) was included as a fixed control predictor. Again, interactions between speaking style and trial, and (if applicable) between speaking style and trial squared were included to investigate the consistency of the model-predicted intelligibility in habitual and clear-Lombard speech over trials. Additionally, interactions between speaking style and speaker characteristics (DDK rate and accuracy and speaker gender) were also included to test whether speakers' gender and speech motor control abilities influenced consistency and the differences in predicted intelligibility scores in different speaking styles. For the random structure, we allowed participant and item (i.e., sentence) as two random effects, plus random by-participant slopes for speaking style and trial. The full model with better fit was then stripped in a stepwise manner to arrive at the most parsimonious model, with model comparisons applied after each removal of the least significant predictor.

5.3. Results

5.3.1. Descriptive data of the acoustic measures and the HEGP predicted intelligibility

Descriptive data for articulation rate, pitch measures (median F_0 and F_0 range in semitones), spectral balance, vowel space measures, and HEGP-predicted intelligibility scores split by speaking style and speaker gender, are summarised in Table 5.1. Additionally, speakers' overall speech enrichment success (as predicted by the HEGP-intelligibility model) is shown in Figure 5.1.

Table 5.1 Summary of the five acoustic measures divided by speaking style (Habitual and Lombard) and speaker gender

	Habitual		Lombard	
	Female	Male	Female	Male
Measurement	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Articulation rate (syll/sec)	5.53 (0.70)	5.98 (0.95)	4.53 (0.69)	5.19 (0.81)
Median F_0 (semitone)	93.03 (1.95)	82.92 (2.91)	95.09 (2.07)	86.40 (2.75)
F_0 range (semitone)	6.38 (1.90)	6.37 (2.12)	8.02 (2.04)	8.47 (1.86)
Spectral balance (dB)	19.96 (4.05)	19.39 (3.39)	12.17 (3.57)	16.34 (4.04)
Vowel space (convex hull)	22.56 (4.38)	15.95 (2.93)	23.29 (6.49)	17.78 (4.53)
HEGP intelligibility scores	0.44 (0.04)	0.44 (0.04)	0.56 (0.04)	0.50 (0.06)

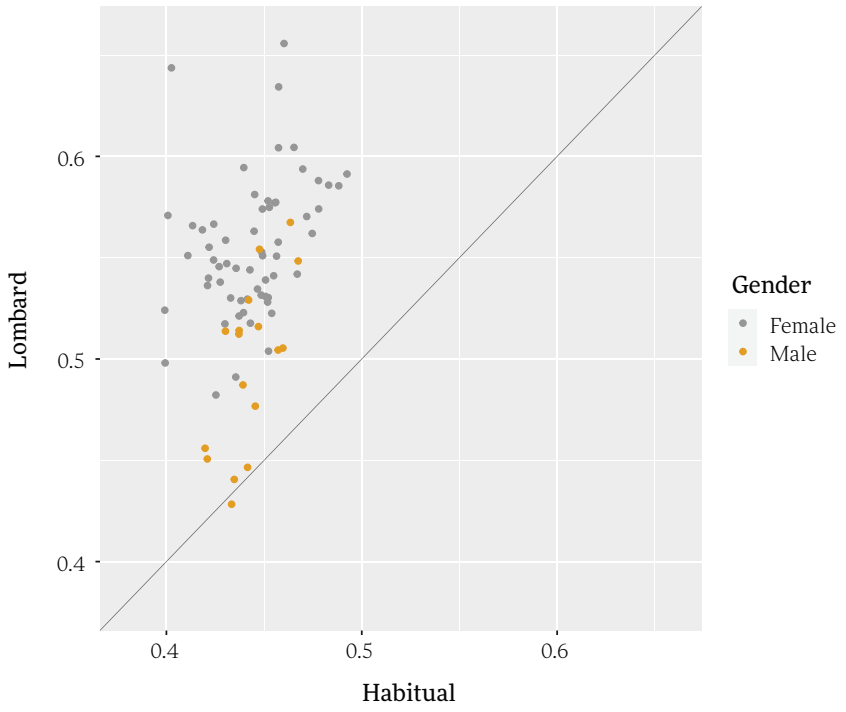


Figure 5.1
Scatter plot illustrating the mean HEGP intelligibility scores (proportion) averaged over sentences per speaker in the two speaking styles (habitual and Lombard) colour-coded by speaker gender

5.3.2. Correlations between acoustic and intelligibility measures

In order to obtain a general view of relationships between different measures in each speaking style, we first explored the correlations between the acoustic measures and the HEGP intelligibility scores through Pearson correlation coefficients. In addition, we examined how these intercorrelations may differ between speaking styles.

The correlation tables 5.2 and 5.3 show that in both the habitual and clear-Lombard speaking styles, the strongest correlations were found between HEGP intelligibility scores and spectral balance (r = -.66 and r = -.64 respectively). Note that we adopted a conservative alpha level for our correlation measures (as we present 30 correlations in total, we opted for the strict alpha level of $p < 0.001$). In the clear-Lombard speaking style, except for F_0 range, all other acoustic measures exhibited significant correlations with each other. HEGP intelligibility score was associated with articulation rate, median F_0, and F_0 range (r = -.59, r = .49, and r = .11) in clear-Lombard

Table 5.2 Pearson correlation coefficients of the four sentence-level acoustic measures and the HEGP intelligibility scores in the habitual speaking style, with boldface representing significant correlations after Bonferroni correction for multiple testing

Parameter	F_0 range	Median F_0	Articulation rate	Spectral balance
HEGP	**0.09**	0.01	**-0.13**	**-0.66**
Spectral balance	**-0.16**	**0.06**	0.02	
Articulation rate	**-0.08**	**-0.24**		
Median F_0	0.05			

Table 5.3 Pearson correlation coefficients of the four sentence-level acoustic measures and the HEGP intelligibility scores in the clear-Lombard speaking style, with boldface representing significant correlations after Bonferroni correction for multiple testing

Parameter	F_0 range	Median F_0	Articulation rate	Spectral balance
HEGP	**0.11**	**0.49**	**-0.59**	**-0.64**
Spectral balance	-0.05	**-0.45**	**0.31**	
Articulation rate	**-0.09**	**-0.39**		
Median F_0	0.02			

style, but note that F_0 is gender-dependent (see section 3.3.1). If we split the (clear-Lombard) data by gender, we saw that the correlation between HEGP scores and median F_0 was stronger in the female (r = .27) than in the male speakers (r = .16).

Tables 5.4 presents the correlation coefficients for the association between the speaker-level vowel-space measures and the five measures (four sentence-level acoustic measures and the HEGP intelligibility measure), per speaking style. Table 5.4 shows that in the habitual speaking style, vowel space is correlated with median F_0 (r = .49) and articulation rate (r = -.37). In the Lombard speaking style, vowel space is correlated with articulation rate (r = -.40), and HEGP scores (r = .40). These results indicate that these acoustic features change hand-in-hand in (modified) speech production, and that the four sentence-level acoustic and the speaker-level phonetic features are associated with HEGP-model predicted intelligibility scores, particularly with predicted intelligibility for the clear-Lombard style.

Table 5.4 Correlation coefficients for the relationship between vowel space and five (sentence-level) measures: four acoustic measures and predicted intelligibility HEGP. Boldface represents significant correlations, after Bonferroni correction for multiple testing

Parameter	F_0 range	Median F_0	Articulation rate	Spectral balance	HEGP
Vowel space (Habitual style)	0.13	**0.49**	**-0.37**	-0.10	0.19
Vowel space (Clear-Lombard style)	0.15	0.34	**-0.40**	-0.36	**0.40**

5.3.3. Consistency of speaking style over time and the role of speech motor control

Outcomes of our statistical modelling procedures will be discussed below per dependent variable (sentence-level acoustic measures and HEGP-model predicted intelligibility scores first, followed by the aggregated speaker-level vowel-space measure).

5.3.3.1. Acoustic measures

Articulation rate

For articulation rate, the model with the quadratic term of trial had better fit than the one without. Table 5.5 shows that articulation rate is modulated, for the habitual speaking style mapped on the intercept, by trial, trial squared, speaking style, speaker gender, and speech motor control (as indexed by DDK non-word rate). Additionally, there is a significant interaction between speaking style and trial, and between speaking style and trial squared. These results address our RQ 1 on consistency over trials by showing that speakers initially speeded up over trials in the habitual speaking style, while slowing down in the Lombard speaking style (see Figure 5.2). As these effects again levelled off at later trials, the overall rate difference between the two speaking styles was largest half-way the experimental trials and then decreased again. Moreover, female speakers had, in general, slower articulation rate than male speakers.

The fact that articulation rate is also significantly predicted by DDK rate indicates that speakers who were faster at repeatedly producing DDK sequences were also faster at sentence reading, across both speaking styles. This indicates, in relation to our RQ 2 on predictors of speaking style, that there is a link between speakers'

maximum speech rate (as indexed by alternating DDK task using non-word stimuli) and their habitual as well as clear-Lombard articulation rates.

Table 5.5 The coefficient estimates, standard errors, and p-values of factors as well as random effects involved in articulation rate measures. Boldface denotes significant results

	Articulation rate		
Predictors	Estimates	Std. Error	p
(Intercept)	5.29041	0.08530	**<0.001**
Style [Lombard]	-0.62492	0.06906	**<0.001**
Trial^2	-0.00037	0.00004	**<0.001**
Trial	0.02197	0.00199	**<0.001**
DDK non-word rate	0.18178	0.05609	**0.001**
Gender [Male]	0.42520	0.13586	**0.002**
Style [Lombard] * Trial^2	0.00054	0.00005	**<0.001**
Style [Lombard] * Trial	-0.03111	0.00272	**<0.001**
Random Effects (SD)			
Subject (Intercept)	0.52414		
Speech Style by Subject	0.55781		
Trial by Subject	0.00509		
Sentence (Intercept)	0.33563		
Residual	0.29302		
$N_{subject}$	78		
$N_{sentence}$	48		
Observations	7488		

Median F_0

For median F_0, the model with the quadratic term of trial had a better fit than the one with only a linear trial effect. Table 5.6 shows that, for the habitual speaking style mapped on the intercept, changes in median F_0 (in semitones) are related to trial, trial squared, speaking style, and speaker gender. Additionally, speaking style interacted with trial (linear trial term only) and speaker gender. These results relate to our RQ 1 on consistency over trials by illustrating that speakers raised their median F_0 throughout the experiment session in both speaking styles, and that the increase in median F_0 in the clear-Lombard style was larger than the increase in the habitual style (see Figure 5.3). As a result, the difference in median F_0 between the two

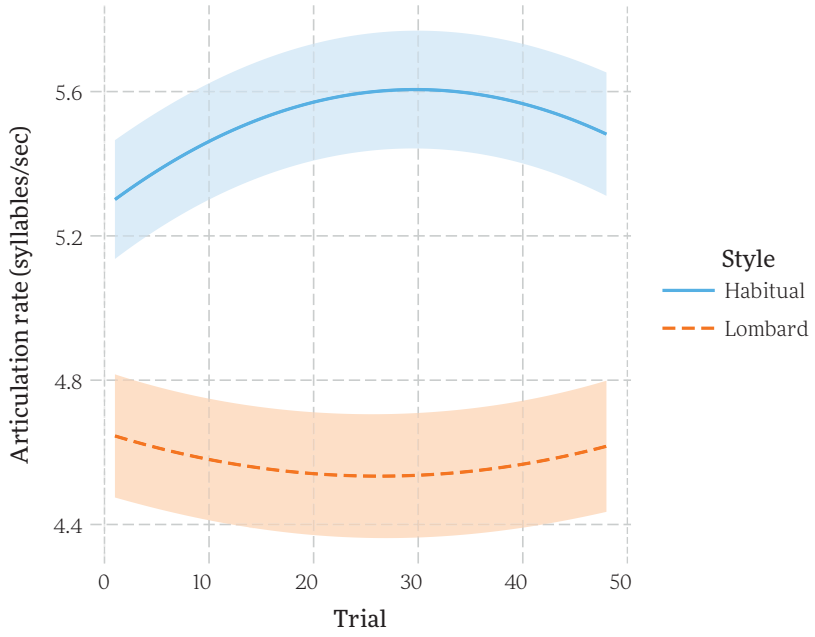


Figure 5.2

Model plot illustrating how articulation rate is associated with speaking style and trial (shading represents 95 percent confidence interval)

speaking styles was enlarged throughout the experiment session. Additionally, female speakers had higher median F_0 than male speakers, and Lombard speaking style exhibited higher median F_0 than habitual speaking style. The significant interaction between speaker gender and speaking style reflects that the increase in median F_0 in the Lombard style was larger for male speakers than for female speakers. Speech motor control was not linked to changes in median F_0 measured in our task here, which informs our RQ 2 on predictors of speaking style or speech enrichment.

Table 5.6 The coefficient estimates, standard errors, and p-values of factors as well as random effects involved in median F_0 measures. Boldface denotes significant results

	Median F_0		
Predictors	Estimates	Std. Error	p
(Intercept)	92.80396	0.23695	**<0.001**
Style [Lombard]	1.76779	0.16635	**<0.001**
Trial	0.01922	0.00602	**0.001**
Gender [Male]	-10.08096	0.46593	**<0.001**
Trial^2	-0.00032	0.00012	**0.006**
Style [Lombard] * Trial	0.01351	0.00260	**<0.001**
Style [Lombard] * Gender [Male]	1.24737	0.31678	**<0.001**
Random Effects (SD)			
Subject (Intercept)	1.72523		
Speech Style by Subject	1.18495		
Trial by Subject	0.01152		
Sentence (Intercept)	0.36666		
Residual	1.26991		
$N_{subject}$	78		
$N_{sentence}$	48		
Observations	7488		

F_0 range

For F_0 range, the model *without* the quadratic term of trial had better model fit, leaving only a linear trial effect. Table 5.7 shows that, for the habitual speaking style mapped on the intercept, F_0 range measures (in semitones) decreased over trials. The significant speaking style effect suggested that speakers exhibit larger F_0 range in Lombard than in habitual speaking style. Related to our RQ 1, the interaction between speaking style and trial indicates that the difference in F_0 range between Lombard and habitual speaking style actually increased over trials, indicating that speakers applied more acoustic modification of F_0 range to their Lombard speech towards the end of the experiment (see Figure 5.4). A lack of a gender effect in F_0 range suggest that females and males had similar F_0 ranges (in semitones), and that both female and male speakers increased their F_0 range to similar degrees when changing speaking styles. Moreover, in relation to our RQ 2 on predictors of speaking style or speech enrichment, speech motor control was not linked to changes in F_0 range here.

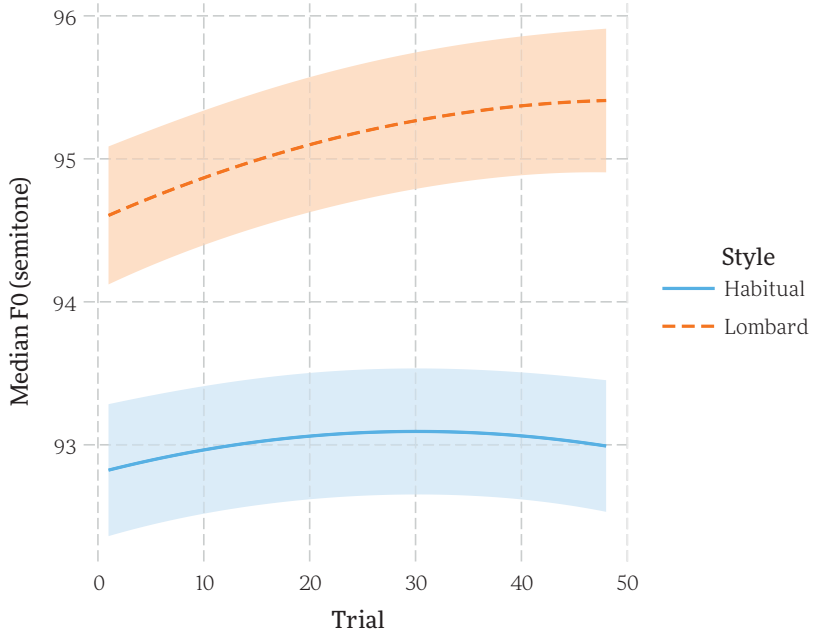


Figure 5.3

Model plot illustrating how median F_0 is associated with speaking style and trial (shading represents 95 percent confidence interval)

Spectral balance

Lastly, for spectral balance, the model with the quadratic trial term had better fit than the one without. Table 5.8 displays that, relative to the habitual speaking style mapped on the intercept, changes in spectral balance are related to speaking style. As lower values in spectral balance indicate louder voices or more vocal effort/energy, these results suggest that speakers increased their vocal effort/energy when changing from habitual to Lombard speaking style (shown as negative values in Table 5.8). Although trial and trial squared were not significant predictors for the habitual speaking style mapped on the intercept, they significantly interacted with speaking style. These results suggest the over-trial increase in vocal effort/energy was present in the clear-Lombard speaking style, and nonlinearly so. In relation to our RQ 1 on consistency of speaking style over trials, instead of decreasing vocal effort towards the end of the clear-Lombard reading session, speakers in our experiment actually increased their vocal effort/energy more towards the end (see Figure 5.5).

Table 5.7 The coefficient estimates, standard errors, and p-values of factors involved in F_0 range measures. Boldface denotes significant results

	F_0 range		
Predictors	Estimates	Std. Error	p
(Intercept)	6.60742	0.18523	**<0.001**
Style [Lombard]	1.43206	0.13447	**<0.001**
Trial	-0.00932	0.00261	**<0.001**
Style [Lombard] * Trial	0.01257	0.00267	**<0.001**
Random Effects (SD)			
Subject (Intercept)	1.42795		
Speech Style by Subject	1.00339		
Trial by Subject	0.01267		
Sentence (Intercept)	0.48325		
Residual	1.30463		
$N_{subject}$	78		
$N_{sentence}$	48		
Observations	7488		

The effect of gender was not significant for the habitual style, but it interacted with speaking style, with males increasing their vocal effort/energy less than females when changing from habitual to Lombard speaking style. Speech motor control, again, was not linked to changes in spectral balance measured here. Thus, the data here do not provide evidence that vocal effort/energy per se, or changes in vocal effort/energy moving from habitual to Lombard speech, were related to individual speech motor control.

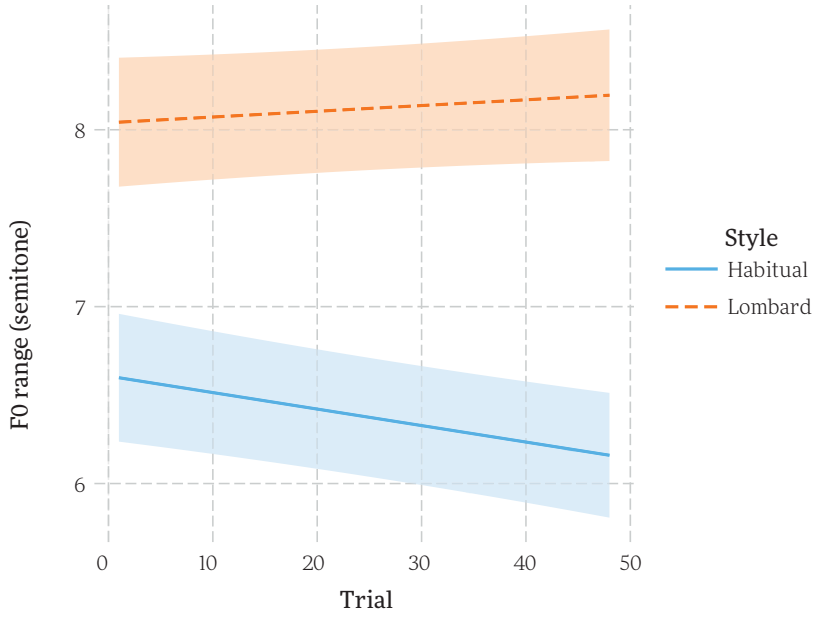


Figure 5.4
Model plot illustrating how F_0 range is associated with speaking style and trial (shading represents 95 percent confidence interval)

Vowel space

After examining the relationship between speech motor control and sentence-level acoustic measures, we investigated the relationship between speech motor control and the vowel space measure. Note again that this vowel space measure is an aggregated measure over multiple sentences, such that it cannot be analysed for trial effects. Table 5.9 summarises that, changes in vowel space measure were related to speaker gender, with smaller vowel spaces for male speakers. However, changes in vowels as measured in our study were not associated with speech motor control (as indexed by maximum DDK performance), nor speaking style.

HEGP-model predicted intelligibility

For HEGP-model predicted intelligibility scores, the model with the quadratic term of trial had better fit than the one without. Table 5.10 shows that HEGP-model predicted intelligibility scores for the habitual speaking style (mapped on the intercept) are significantly related to the non-linear term of trial. Predicted intelligibility increased non-linearly over trials throughout the experiment. The speaking style effect

Table 5.8 The coefficient estimates, standard errors, and p-values of factors involved in spectral balance measures. Boldface denotes significant results

	Spectral balance		
Predictors	Estimates	Std. Error	p
(Intercept)	20.33556	0.42400	**<0.001**
Style [Lombard]	-7.87021	0.37244	**<0.001**
Trial^2	0.00030	0.00030	0.321
Trial	-0.02493	0.01560	0.110
Gender [Male]	-0.58505	0.66721	0.381
Style [Lombard] * Trial^2	-0.00129	0.00042	**0.002**
Style [Lombard] * Trial	0.04572	0.02106	**0.030**
Style [Lombard] * Gender [Male]	4.66022	0.65872	**<0.001**
Random Effects			
Subject (Intercept)	2.45818		
Speech Style by Subject	2.35476		
Trial by Subject	0.02058		
Sentence (Intercept)	1.61563		
Residual	2.34300		
$N_{subject}$	78		
$N_{sentence}$	48		
Observations	7488		

confirmed that clear-Lombard speech indeed had a higher predicted intelligibility in noise than habitual speech. Furthermore, speaking style interacted with trial, suggesting that at least for female speakers (mapped on the intercept), the speaking style effect (or Lombard-intelligibility gain) increased over trials (see Figure 5.6).

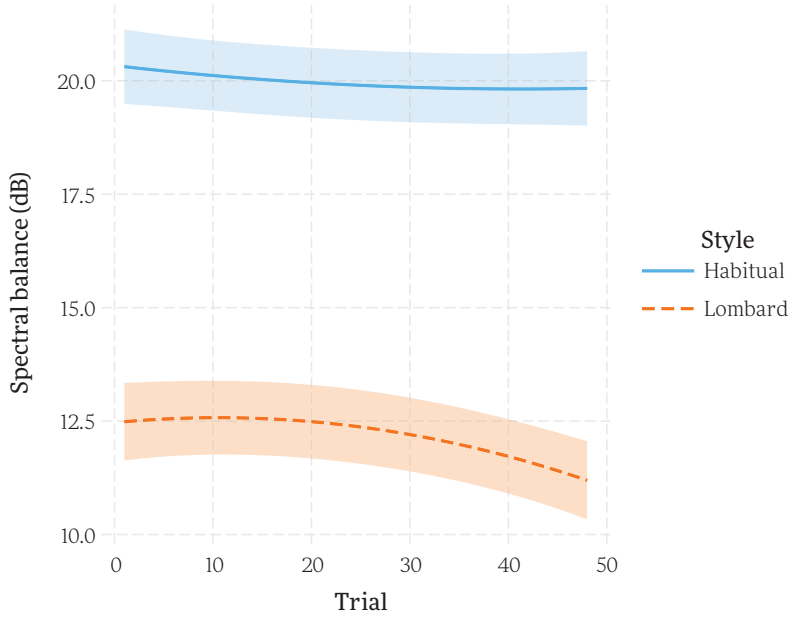


Figure 5.5
Model plot illustrating how spectral balance is associated with speaking style and trial (shading represents 95 percent confidence interval)

Table 5.9 The coefficient estimates, standard errors, and p-values of factors involved in factors involved in vowel space measures. Boldface denotes significant results

	Vowel space		
Predictors	Estimates	Std. Error	p
(Intercept)	22.92455	0.47248	**<0.001**
Gender [Male]	-6.06022	1.01206	**<0.001**
Observations	156		

Table 5.10 The coefficient estimates, standard errors, and p-values of factors involved in HEGP-model predicted intelligibility scores (in logits). Boldface denotes significant results

	HEGP (logits)		
Predictors	Estimates	Std. Error	p
(Intercept)	-0.22252	0.01736	**<0.001**
Style [Lombard]	0.43396	0.01821	**<0.001**
Trial	-0.00091	0.00046	0.051
Gender [Male]	0.00119	0.02214	0.957
$Trial^2$	0.00002	0.00001	**0.015**
Style [Lombard] * Trial	0.00055	0.00020	**0.005**
Style [Lombard] * Gender [Male]	-0.21864	0.03736	**<0.001**
Random Effects (SD)			
Subject (Intercept)	0.08055		
Speech Style by Subject	0.13611		
Trial by Subject	0.00111		
Sentence (Intercept)	0.08950		
Residual	0.09386		
$N_{subject}$	78		
$N_{sentence}$	48		
Observations	7488		

The interaction between speaking style and gender indicates that male speakers had a smaller predicted Lombard-intelligibility gain compared to female speakers. This result echoes with our previous results in the consistency of speech modifications that speakers apply over trials. Concerning our RQ 3 on the consistency of predicted intelligibility over trials, speakers generally increased their Lombard speech modifications over trials, resulting in improved (predicted) intelligibility gains in noise over the course of the experiment (i.e., sentence list). Additionally, speech motor control as measured in our study did not relate to predicted intelligibility (RQ 2).

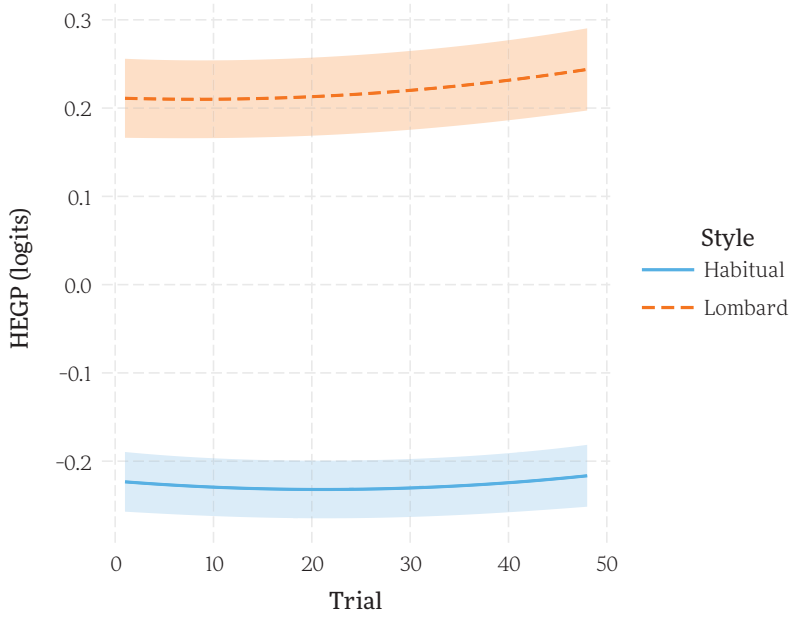


Figure 5.6

Model plot illustrating how HEGP-model predicted intelligibility scores is associated with speaking style and trial (shading represents 95 percent confidence interval)

5.4. Discussion

In this study, we investigated variability in the speech enrichment modifications speakers apply when changing from baseline (habitual) speech to clear-Lombard speech production. More specifically, we analysed how consistently speakers produced enriched speech over trials, how similar their habitual speech production was across trials, and whether speaker characteristics such as speech motor control related to this consistency and to the overall speech enrichment modifications that speakers applied (RQs 1 and 2). Additionally, we looked into the relationship between the acoustic features of speakers' (enriched) speech on the one hand, and the model-predicted speech intelligibility on the other (RQ 3). In general, albeit to a varying degree, virtually all speakers applied enrichment modifications in their clear-Lombard speech compared to their habitual style speech (see Figure 5.1 in the result section). These enrichment modifications are evident from the four sentence-level acoustic-phonetic measures, namely articulation rate, median F_0, F_0 range, and spectral balance, but not from the speaker-level vowel space measure

(which we will further discuss later in this section). This confirms that speakers generally enriched their speaking style through the clear-Lombard speech elicitation technique, which enables us to address the research questions for this study.

5.4.1. Consistency of speaking style over time and the role of speech motor control

Most of the sentence-level acoustic measures (i.e., articulation rate, median F_0, and F_0 range) showed changes during the course of the experiment in the clear-Lombard speaking style. Regardless of whether these changes were linear or non-linear, these findings suggest that speakers adapted their speech enrichment modifications constantly throughout the experiment session. Similarly, speech acoustics also change during the habitual speaking style, such that the intelligibility benefit of the clear-Lombard style is far from static. Different from Lee and Baese-Berk (2020), speakers in our study exhibited mostly increased speech enrichment modifications and hence improved intelligibility ratings throughout the experiment. Given the differences in task set-up, clear speech production in the more spontaneous conversations as in Lee and Baese-Berk's study could be more effortful to begin with, which may explain the differences in clarity/intelligibility changes over time. Possibly, if speakers read, rather than formulate spontaneously, they can allocate more attention to the clarity of their pronunciation.

The enrichment modifications in articulation rate that speakers applied in their clear-Lombard speech (i.e., slowing down over Lombard reading trials) exhibited an opposite pattern compared to the rate changes in their habitual style, making the differences in articulation rates in the two speaking styles at its maximum halfway through the experiment. Thus, for articulation rate, speakers initially took some time to adjust their articulation rate for the clear speaking style, and would then gradually return to rates that they are supposedly more comfortable with.

Concerning pitch measures, clear-Lombard speech exhibited significantly higher median F_0 and wider F_0 range. Additionally, speakers generally increased their pitch enrichment modifications throughout the experiment session. Specifically, they continuously increased their median F_0 and F_0 range in the clear-Lombard style. These results show that our speakers were able to continuously raise their pitch and, in a way, exaggerate their articulation through expanded F_0 range throughout the experiment session, which again provides evidence that speakers need time or practice to achieve their maximally clear style. While speakers cannot keep raising their pitch and expanding their F_0 range, they may do so until a certain 'asymptote' is reached, similar to that observed in articulation rate. However, without empirical data, we cannot know the maximum amount of pitch modifications possible for our

young adult speakers tested here. It seems, from comparing the rate and pitch patterns over trials, that speakers need more time or practice for pitch adaptation than rate adaptation in clear-Lombard speech production. For future work, it could be informative to test speakers' maximum pitch modifications capacity in clear-Lombard speech production using longer experimental sessions.

For spectral balance, changes over trials were greater in the clear-Lombard style than in the habitual style, especially towards the end of the experiment session. These results suggest that speakers were able to continuously apply this speech enrichment adaptation in their clear-Lombard speech production, and that some practice is needed for them to realise the more 'enriched' speech. Whereas speakers seemed to exert less effort over trials in their habitual style, as evident from changes in pitch and rate, their vocal effort remained more or less constant in their habitual reading style. However, in order to test speakers' maximum capacity in speech enrichment modifications, a longer experiment session needs to be implemented, given that vocal fatigue has generally been observed after prolonged periods of voice use (e.g., Novak et al., 1991; Gelfer et al., 1991).

These speech enrichment modifications that speakers applied in the clear-Lombard speaking style are consistent with previous literature (e.g., Van Summers et al., 1988; Cooke & Lu, 2010; Garnier & Henrich, 2014; Junqua, 1993). The novel results on the consistency of speakers' speech enrichment modifications in an experiment session indicate that speakers (at least the healthy and young adult speakers tested in our study) may need some practice to reach their full potential in producing clear and/or Lombard speech. This result challenges the idea that the Lombard reflex is an automatic change: even though speakers often immediately speak up when presented with loud noise, they do improve their speech enrichment modifications with prolonged practice. This 'practice effect', or the fact that speakers would need some time to adjust their speech production has also been found in a study when their auditory feedback was altered (e.g., Purcell & Munhall, 2006), and in a study where their articulation was disturbed (e.g., Fowler & Turvey, 1980). Note again that in our study, speakers adapted their speaking style, while being exposed to loud noise played through headphones, which strongly reduced their auditory feedback. Possibly, upon receiving such limited feedback on their production, speakers still realised that they could do more, and thus kept trying to overcome the noise in order to monitor their own speech better by extracting useful (auditory and/or somatosensory) information about how well their articulation targets are being met. Thus, our finding that speakers became generally better at enriching their speech over time without actually being able to hear their own speech well, raises the interesting question of what cues speakers used to improve their speech clarity for

future studies. Possibly, this very limited auditory feedback may have caused speakers to be relatively slow at adjusting their speech output during the clear-Lombard speech production, compared to a condition where they would have heard their own speech better. Future studies are needed to address this.

As mentioned earlier, vowel space did not differ between the two speaking styles in our study. This could be caused by a number of factors. Firstly, previous studies that examined vowel-space expansion in clear-speech and/or Lombard speech have shown inconsistent findings. For instance, a few studies have reported vowel-space expansion in Lombard speech in English (Bond et al., 1989), infant-directed clear speech as well as Lombard speech in Mandarin Chinese (Tang et al., 2017), and hyper-articulated clear speech in both English and Croatian (Smiljanić & Bradlow, 2005). However, some other studies have reported no clear vowel-space expansion in speech produced in noise (e.g., Cooke & Lu, 2010; Kim & Davis, 2014). Our results were more similar to the findings in the latter two studies. Secondly, different studies that investigated vowel space expansion have made rather different methodological choices in, for instance, the number of (corner) vowels and the number of tokens per vowel included in their speech stimuli. In addition, the noise level used in this study (78 dB SPL) was lower than in some Lombard studies (e.g., 95 dB SPL in Bond et al., 1989). Admittedly, our vowel space measure was a rather crude one given that we only had three corner vowels with eight tokens per vowel. Future studies using more vowels with a larger number of tokens per vowel than the ones employed in the current study could yield more informative insights into vowel space expansion in Lombard speech. Alternatively, hyperarticulated clear speech may differ from speech produced in noise in terms of the extent of vowel space expansion.

We also investigated how speaker characteristics (e.g., speech motor control and speaker gender) related to (enriched) speech production. Although DDK non-word rate did not relate to a speaker's rate consistency over trials (nor did DDK accuracy), DDK rate did predict articulation rate in both speaking styles. This link between the maximum rate at which speakers can move their articulators and their habitual as well as clear-Lombard articulation rate suggests that there may be a speaker-intrinsic aspect to rate control in their speech production. Moreover, de Jong and Mora (2017) found DDK accuracy to be related to speakers' speech fluency in their L1 and L2 speech, providing similar evidence for an association between maximum speech performance and observed (natural) speech performance. However, note that speech motor control has been measured in various other ways than the maximum performance measure of DDK we employed here. Other motor control studies focused on variability of articulatory movement (e.g., Sadagopan & Smith, 2013; Terband & Maassen, 2010) over repeated productions, where speaker groups with

less variability are seen as having a more stable motor system. It is conceivable that potential relationship between speech motor control and other (acoustic) aspects of speech production may be found if, for instance, the consistency aspect of speech motor control (quantified by e.g., the spatial-temporal variability index) was measured. Future research using multiple tasks that tap various aspects of speech motor control could help further understand the relationship between speech motor control (either quantified through maximum performance or stability measures) and enriched speech production.

Like our index of speech motor control, speaker gender did not relate to the consistency over trials of clear-Lombard speech modifications either. However, gender did play a role in changes of median F_0 and spectral balance moving from habitual to clear-Lombard speaking style, in line with earlier findings (e.g., Junqua, 1993; Bradlow et al., 1996).

5.4.2. Consistency of HEGP-model predicted intelligibility over time and the role of speech motor control

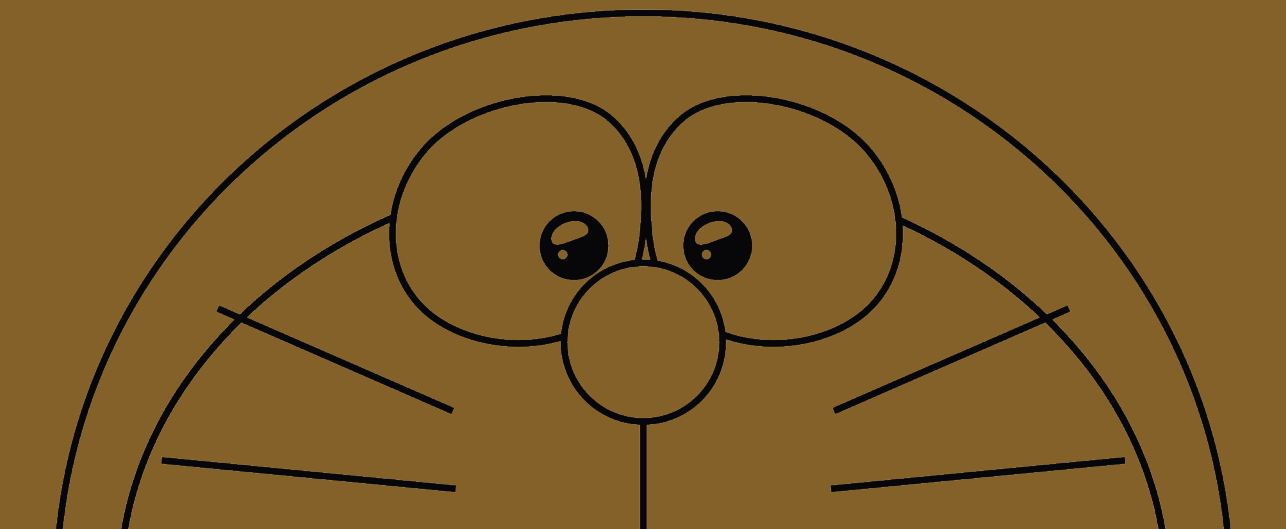
Before discussing predicted intelligibility, we first discuss the relationship between HEGP-model predicted intelligibility scores and the acoustic features observed in speakers' habitual and clear-Lombard speaking styles. Generally speaking, higher HEGP scores, particularly in the clear-Lombard style, were associated with reduced articulation rate, raised median F_0, expanded F_0 range, and reduced spectral balance (i.e., increased vocal effort/energy). Amongst these correlations, HEGP intelligibility scores and spectral balance measures displayed the strongest relationship, suggesting that increased vocal effort/energy was the most salient contributor to higher (predicted) intelligibility in noise.

In line with the enlarged acoustic differences (over trials) between speaking styles, the predicted intelligibility difference between the two speaking styles also increased over sentence trials. This enlarged difference was mainly driven by speakers' increased HEGP scores over trials in the clear-Lombard speaking style, which are likely related to the changes in spectral balance and articulation rate over the course of the Lombard reading session (see also Table 5.3 with correlational data). Although the HEGP model does not 'perceive' speech intelligibility the same way as human listeners, given the model's focus on 'low-level' energetic masking (or release thereof) while ignoring segmental changes that could be beneficial to listeners, the model-predicted intelligibility scores did correlate highly with human listeners' listening effort ratings on a sub-sample of speakers. It seems therefore reasonable to expect that speakers' speech intelligibility increased over trials in the clear-Lombard speaking style.

Interestingly, gender differences in speakers' speech enrichment success, as reflected by HEGP-model predicted intelligibility, were also found in our speaker group. Female speakers had a larger (predicted) Lombard-intelligibility gain compared to male speakers (see also e.g., Ferguson & Morgan, 2018; Junqua, 1993; Bradlow et al., 1996). This is in line with the acoustic finding (see section 4.4.1) that female speakers increased their vocal effort/energy more than male speakers in their clear-Lombard speech. Male speakers raised their pitch more than female speakers, but did not increase their vocal effort as much as female speakers when enriching their speech for the clear-Lombard style. These results also tie in with the observation that vocal effort/energy (as indexed by spectral balance) has the strongest correlation with HEGP intelligibility scores. The differences in speech enrichment modifications between female and male speakers may have contributed to the higher (HEGP-model predicted) intelligibility scores for female speakers, particularly in the clear-Lombard speaking style. Additionally, given the limited association between speech motor control and acoustic patterns over the two speaking styles, it is not surprising that the HEGP metric is not related to speech motor control either.

5.5. Conclusion

Results showed that young adult speakers may need some practice to reach their full speech enrichment potential when asked to speak clearly in the presence of loud background noise which drastically reduced their auditory feedback. Additionally, changes in speakers' speech enrichment modifications over trials (as evident from the acoustic measures) were also reflected by the HEGP-model predictions of their intelligibility in noise. Lastly, speakers' speech motor control ability (as indexed by maximum-rate-performance using a non-word DDK task) was shown to relate to speakers' habitual and Lombard rates. This suggests an underlying rate control mechanism in speakers' (enriched) speech production.



6

General Discussion

In the act of speaking, people are generally capable of adjusting their speech according to task requirements (e.g., speeding up articulation rate when asked for rapid repetition of particular phrases) or communicative needs (e.g., speaking more clearly in loud background noise). Yet, speakers may also differ in the extent to which they adjust their speech according to the speech situation at hand. This thesis investigated speakers' individual differences in speech behaviour through three different studies. Throughout these studies, the focus was on the so-called 'late stages' of speech production, defined as the stages involved in speech production after lemma selection, according to Levelt's model for speaking (Levelt, 1989). More specifically, I first explored individual differences in young adult speakers' speech motor control abilities, as indexed by two maximum-performance speech tasks. In these maximum-performance tasks, speakers were required to accurately alternate between syllable sequences at a fast speed (i.e., at maximum performance levels). In the first empirical study, I investigated whether speakers' speech motor control ability (as indexed by maximum speech performance tasks) relates to their executive control abilities. Then, in the second empirical study, I looked into which processes of speech motor planning (i.e., motor programming and response initiation) were affected by adult ageing. Following these investigations of individual and age differences in speech motor control, I then looked into whether and how individual differences in speech motor control (as indexed by maximum speech performance) relate to how speakers communicate in noise. More specifically, in the third study, I examined intra- and inter-speaker variability in speakers' speech enrichment strategies and success while speaking in a noisy background. That is to say, I investigated the extent to which speakers adjusted their speech acoustics when speaking in noise, and how these adjustments changed over sentence trials.

The structure of this chapter is as follows. First, the two speech corpora that resulted from the research presented in this thesis are described. Then, the main findings of the empirical chapters are summarised and discussed in light of frameworks of speech production. Possible directions for future studies as well as the methodological choices and the limitations of the current studies are also addressed in the themed discussion of the relevant findings.

6.1. Speech corpora

Two open-access corpora (published together with this thesis) were developed based on speech data collected for **Chapters 2, 3,** and **5** of this thesis. Specifically, one corpus is the Radboud Maximum Speech Performance Corpus (RaMax) and the other corpus is the Radboud Lombard Corpus (RaLoCo).

6.1.1. The Radboud Maximum Speech Performance (RaMax) Corpus

The first speech corpus, RaMax (Radboud Maximum Speech Performance), was collected to serve as a maximum speech performance dataset for indexing speakers' speech motor control (Shen & Janse, 2019, 2020, see also **Chapters 2** and **3**). The RaMax corpus contains two speech tasks, i.e., a diadochokinetic (or DDK) task and a tongue twister task. Speech motor control or articulatory control has often been measured in clinical settings using a maximum speech performance task: the DDK task (Bernthal et al., 2009; Duffy, 2013). During this task, speakers are encouraged to repeat one syllable or a nonsensical sequence of multiple syllables at maximum speed, while maintaining accuracy. Articulatory control has been assessed in a different way in psycholinguistic studies. Psycholinguistic studies on late stages of speech production have often adopted the tongue twister task to elicit phoneme selection errors as a means to tax speech performance (e.g., Goldrick & Blumstein, 2006; Wilshire, 1999). Speech data from 78 native Dutch speakers performing these two speech tasks then form two sub-corpora. In the tongue twister task, the 78 young adult speakers each produced four Dutch language tongue-twister sentences at their maximum capacity (in terms of production rate and accuracy) with five to eight repetitions per sentence. An example of the tongue twister sentence (and its translation in English) is: 'Ik bak een plak bakbloedworst' *(I fry a slice of blood-sausage)*, in which syllable onset 'b' /b/ alternates with syllable onsets 'pl' /pl/ and 'bl' /bl/.

Additionally, for the DDK task, the same 78 speakers produced a total of seven repeating and alternating DDK stimuli. The three repeating DDK stimuli are: 'papapa...', 'tatata...', and 'kakaka...' while the four alternating DDK stimuli are: 'pataka...', 'katapa...', 'kapotte...', and 'paketten...'. Note for the alternating stimuli that the first stimulus, 'pataka', is the standard alternating DDK stimulus used by the vast majority of other DDK research, while the other three stimuli were selected to fit the research question of the thesis. That is to say, the second DDK stimulus, 'katapa', is the reverse-ordered 'pataka' and the last two stimuli are two real Dutch words that are closest to the non-word sequences, as we wanted to relate speakers' executive control abilities to their maximum speech performance on both lexical and non-lexical items. Again, the 78 speakers repeatedly produced these stimuli (following a fixed presentation order) at their maximum rate and accuracy for up to 10 seconds. For more detailed information about the contents of this RaMax corpus, including more participant information and technical details of the recording equipment, readers are referred to Appendix A.

6.1.2. The Radboud Lombard (RaLoCo) Corpus

The second corpus is the RaLoCo (Radboud Lombard Corpus) corpus, consisting of speech of the same 78 young adult native Dutch speakers as in the RaMax corpus. The RaLoCo corpus is composed of two datasets, namely a speech dataset and an accompanying rating dataset. The speech dataset of the RaLoCo corpus was collected to facilitate answering the research questions addressed in **Chapter 5**. The speech dataset of the RaLoCo corpus contains Dutch sentence-reading materials of the 78 speakers mentioned above in two conditions: a habitual condition in which the speakers were instructed to read out 48 sentences fluently; and a clear-Lombard condition where they were instructed to read out the same 48 sentences as clearly as possible while hearing loud speech-shaped noise (at 78 dB SPL) via headphones. The sentence-reading materials hence amount to 7488 sentences in total: 78 speakers x 48 sentences x 2 speech conditions. Of the 48 unique sentences, half (i.e., 24 sentences) had a keyword noun containing one of the three corner vowels (i.e., /i/, /u/, /a/) embedded in the sentence. An example keyword sentence is 'Mijn opa had de piep jammer genoeg niet meer gehoord (translation: Unfortunately, my grandfather hadn't heard the beep anymore)', in which 'piep (*beep*)' is the corner-vowel target word.

The rating dataset of the RaLoCo corpus contains two types of rating data: 1) predicted intelligibility ratings based on an acoustic metric (the high-energy glimpse proportion, or HEGP) for the entirety of the speech dataset (i.e., 7488 ratings in total with one rating per utterance), which have been discussed in **Chapter 5**; and 2) listening effort ratings from 231 human subjects for a subset of the speech dataset (i.e., for 48 out of the total of 78 speakers, or a total of 4608 unique utterances). Please note that collection of these human ratings was not part of this thesis, and hence has not been discussed in the thesis. The HEGP rating, based on an acoustic glimpse-based metric, indexes the contribution of the high-energy glimpses to intelligibility surviving energetic masking from noise (at a signal-to-noise ratio, or SNR, of -5 dB). For the listening effort rating, each of the unique sentences (in both speech conditions) from the selected 48 speakers was rated by two to four human listeners. For these ratings, utterances were embedded in the same speech-shaped noise as had been used for the elicitation of the clear-Lombard materials, presented to raters at -6 dB SNR. Upon presentation of the speech fragment, raters were asked to indicate (on a scale from 1 to 7) how much effort they needed to spend to understand the content of the speech, with 1 indicating 'no effort at all to understand' and 7 indicating 'extremely effortful to understand'.

The human ratings were collected via an online listening experiment conducted by a student assistant at Centre for Language Studies, Radboud University. The rating dataset of the RaLoCo corpus includes two CSV files, one with *predicted-intelligibility*

ratings based on the HEGP metric (of all unique utterances in the speech corpus), and one with data of the human listeners' listening effort ratings based on a subset of the speech corpus. For more information on the contents of this RaLoCo corpus, including technical details of the recording equipment and the distribution of speech materials over human raters and the rating procedure, the reader is referred to Appendix B.

6.2. Summary of main findings

The clinical and psycholinguistic research fields have used different methods to elicit maximum speech performance to index articulatory control. Both the clinical DDK and the psycholinguistic tongue twister tasks require speakers to rapidly alternate between similar syllables. In **Chapter 2**, a methodological chapter, I aimed to explore the link between individual speakers' performance on the two tasks, especially when both tasks are utilised as maximum speech performance tasks. The main finding regarding the link between participants' maximum performance in the DDK and the tongue twister task was that maximum speech rates for the two tasks, but not accuracy levels, were moderately positively correlated ($r > .5$). Further analyses revealed differences in task performance both within a task between different stimuli, and across tasks. Specifically, maximum rate and accuracy of the DDK task was considerably higher than those of the tongue twister task. Additionally, as expected, participants were faster and more accurate in producing the real word DDK sequences compared to the non-word ones (cf. section 6.3 for discussion of the lexicality effect). These results suggest that there may be intrinsic differences in task and/or stimuli difficulty levels, although better experimental control over task stimuli and presentation order are needed to further verify this. The rate correlation between the two speech tasks suggests that both tasks contain elements reflecting participants' ability to plan and execute similar articulatory programmes, regardless of differences between tasks in terms of the type and length of stimuli and the involvement of (sentence-level) linguistic processing.

Following up on the observed correlation between DDK and tongue twister performance, **Chapter 3** took participants' performance on these two tasks as indices of their articulatory control, aiming to investigate the relationship between articulatory control and executive control abilities. While some studies have argued that late stages of speech production are largely automatic (e.g., Ferreira & Pashler, 2002; Garrod & Pickering, 2007), several other studies have demonstrated links between participants' executive control and articulatory control abilities (e.g., Dromey & Benson, 2003; Nijland et al., 2015). **Chapter 3** addressed these mixed

findings on the relationship between articulation and executive control, in order to shed some light on our understanding of the processes of speaking.

Executive control or executive functions (EF) have been proposed to cover three core elements: inhibition, switching, and updating of working memory (Miyake et al., 2000). These three elements are related, yet distinct. In **Chapter 3**, participants' executive control abilities were captured using three different cognitive tasks, each measuring one of these three main elements of EF. Specifically, a Flanker interference task was used to index inhibitory control, a Letter-number switching task was used to index cognitive switching, and an Operation Span task was administered to index updating of working memory. The main finding based on the results from 78 young adult participants (the same participants as those in **Chapter 2**) was that, although none of the executive control abilities related to the maximum rates at which participants performed the two speech tasks, their cognitive switching ability predicted their speech accuracy in the two speech tasks. In other words, participants with better cognitive switching ability were also better able to accurately repeat DDK sequences and tongue twister phrases at a fast rate. Additionally, participants' speech production accuracy differed across different DDK task stimuli: accuracy was higher for real word than non-word DDK sequences (as already shown in **Chapter 2**), and the size of this performance difference between the two types of stimuli was found to be modulated by speakers' cognitive switching ability. These results underscored an association between maximum speech accuracy and executive control (cognitive switching in particular), suggesting that late stages of speech production are also associated with cognitive control, as has been observed for earlier stages of speech production.

If articulatory control relates to (at least) the cognitive switching aspect of executive control, it is conceivable that articulatory control may change when speakers age. Adult ageing may affect the control of speech movements through age-related cognitive decline (Glisky, 2007; Salami et al., 2012), as has been observed for limb movement control (Krampe et al., 2005; Niermeyer et al., 2017; Sparto et al., 2014). **Chapter 4** looked into ageing effects on the late stages of speech production, particularly on speech motor planning and initiation. In doing so, **Chapter 4** aimed to better understand *which* late stages of speech production might be vulnerable to age effects. Earlier studies had developed paradigms to decompose the planning and the initiation stages of speech motor programming (e.g., Klapp, 2003; Maas et al., 2008; Maas & Mailend, 2012). Following up on those paradigms, an age group comparison study was conducted in **Chapter 4**. Younger and older adults participated in a speeded simple/choice speech production paradigm with which the processes of speech motor planning and initiation could be distinguished. More specifically, in the simple or

prepared condition, participants could (internally) prepare the target item they were asked to produce before initiating its production. In the choice or unprepared condition, participants only learned which of three target stimuli to produce upon receiving the critical information cue, after which they immediately had to produce the target. Thus, in the choice condition, participants still needed to select the right motor programme and then programme the corresponding motor movements before they could initiate their production. The reaction time (RT) difference between the simple and choice conditions can thus be taken to index the additional time needed to select and prepare the internal properties of a motor programme. Results from the age group comparison showed that age differences in speech onset latencies were larger in the simple than the choice condition, such that older adults only had longer speech onset latencies than younger adults in the simple (prepared) condition, but were equally fast in the unprepared condition. These results suggest that older adults were less prepared in the simple condition than younger adults, or were less efficient in unpacking and launching the prepared motor programmes from the speech motor planning buffer. Conversely, when there was no need to temporarily withhold a spoken response, there was no evidence of any age-related slowing in speech preparation.

Speakers often can flexibly modify their speech behaviour to meet various communicative needs (e.g., when faced with a hearing-impaired interlocutor). In **Chapter 5**, I examined speech produced in the presence of loud background noise, aiming to investigate individual speakers' speech characteristics when producing the so-called 'Lombard speech'. More specifically, I investigated how consistent speakers' acoustic Lombard-speech modifications are over a period of time, and how individuals' articulatory control ability may influence their general speech behaviour, and their speech enrichment consistency and success. Seventy-eight young adult participants (the same speaker sample as those in **Chapters 2** and **3**) read out stimulus sentences in both their habitual speaking style, and in a condition where they were instructed to speak clearly while hearing loud speech-shaped noise over headphones (i.e., clear-Lombard style). A maximum performance speech task (non-word DDK) was used to quantify speakers' articulatory control. Individuals' predicted speech intelligibility in both speaking styles was quantified using an acoustic glimpse-based metric (high-energy glimpse proportion, or HEGP) introduced above. This metric thus represents how intelligible the speaker's speech would have been if their speech had been presented in noise at a defined speech-to-noise ratio.

Acoustic analyses of speakers' habitual as well as clear-Lombard speech focused on speakers' articulation rate, median fundamental frequency (F_0), F_0 range, spectral balance, and vowel space. Results indicated that, despite the higher vocal and

articulatory effort that is required to produce clear-Lombard speech, speakers tested in this study were generally able to not only maintain but even enhance their clear-Lombard speech modifications over the course of the sentence list. More specifically, while producing clear-Lombard speech, speakers tested in the current thesis exhibited reduced articulation rate, increased median F_0, expanded F_0 range, and increased vocal effort (as reflected by lower values in spectral balance) over the course of the experiment. Relatedly, speakers' HEGP-model predicted intelligibility also increased (non-linearly) over sentence trials, especially in the clear-Lombard condition. This signals the flexibility of speakers' speech modification strategies in enriched speech production, and provides evidence for a non-static intelligibility benefit of clear-Lombard speech. Additionally, speakers' articulatory control (maximum speech rate) was associated with their articulation rate in both speaking styles. Note, however, that this association did not show up in **Chapter 2**, where it was tested with a simple correlation test. In **Chapter 5**, more elaborate modelling was applied, bringing out the rate association between maximum and habitual speech rate. To sum up, this study displays acoustic evidence for a non-static Lombard effect over trials. Moreover, speakers differ in the size of this speech enrichment effect, though this difference may not be directly regulated by their articulatory control ability (as quantified here).

6.3. Executive control and articulatory control in late stages of speech production

Whereas the 'early stages' of speech production including conceptualisation and formulation have been shown to relate to cognitive resources such as executive control (Piai & Roelofs, 2013; Shao et al., 2012; Sikora et al., 2016), past psycholinguistic studies have provided mixed results regarding whether processes involved in late stages of speech production require executive control or not (Ferreira & Pashler, 2002; Garrod & Pickering, 2007; Jongman et al., 2015). At the same time, studies on articulatory control have shown links between participants' executive control and articulatory control abilities using experimental (e.g., Bailey & Dromey, 2015; Dromey & Benson, 2003) and correlational methods (e.g., Nijland et al., 2015). Likewise, for people with Parkinson's disease, a link was found between their oral DDK rate ('pataka') and their performance on a Trail Making task (Barbosa et al., 2017), with strong negative correlations between DDK rate on the one hand, and the time they needed to complete Trail Making part A (indexing general cognitive speed) and Trail Making part B (tapping cognitive switching ability) on the other. The findings in **Chapter 3** on the connection between speakers' cognitive switching ability and their maximum speech performance are in line with results obtained in articulatory

control studies (Bailey & Dromey, 2015; Dromey & Benson, 2003; MacPherson, 2019) and in studies involving patient populations with known motor disorders (Barbosa et al., 2017; Nijland et al., 2015). Consequently, the findings in **Chapter 3** provide evidence for the association between executive control and the articulation phase of speech production, also in healthy young speakers without motor disorders. Importantly, there are indications that the association between articulation and executive control is not restricted to maximum performance tasks. Whereas the findings here can be criticised for the unnaturalistic task which involves repetition of somewhat nonsensical sequences or for its focus on maximum speed, similar results have been observed recently for young children. For instance, Netelenbos et al. (2018) investigated executive function (via parental report of the Behaviour Rating Inventory of Executive Function inventory) and fine-grained speech articulation abilities (as indexed by distinctions between two word-initial phonemes: 's' and 'sh' in e.g., *sit* and *ship* respectively) in 4- to 6-year-olds. Their results showed that children with better executive function as measured in their study were also better at distinguishing the /s/ and /ʃ/ sounds in their production, henceforth exhibiting better articulatory abilities. Developmental studies could follow up on our and Netelenbos and colleagues' results to investigate how the association between articulatory and executive control may change during language acquisition throughout childhood and into (older) adulthood.

The maximum speech performance tasks employed in this thesis mainly concern late stages of speech production as no conceptualisation nor lemma selection is needed when reading out DDK sequences or tongue twister sentences on repeat. However, it remains unclear *which* late stages are actually associated with cognitive switching. The fact that participants' production accuracy is shown to be linked to cognitive switching in both maximum performance speech tasks in **Chapter 3** provides evidence for the association between switching and resolving competition and selection at later stages than lemma selection in speech production (i.e., during phonological and phonetic encoding stages). Note, however, the relationship between cognitive switching and late stages of speech production may possibly mainly concern speakers' speech monitoring processes. Speakers can monitor their covert and overt speech, i.e., monitoring for speech errors they are either about to produce, or to repair errors they have just made (see e.g., Nooteboom & Quené, 2017). As laid out by McMillan and Corley (2010) and Goldrick and Blumstein (2006), conflicts during phonological encoding that have not been resolved may result in partial activation of both the target and the competing articulatory gestures. This in turn may lead to potential conflict between multiple highly activated sounds and motor programmes for the same syllable onset slot. Self-monitoring of overt and covert speech may be driven by neural signatures that detect such conflict and by the comparison between

target speech and realised speech (Acheson & Hagoort, 2014; Gauvin & Hartsuiker, 2020; Nozari et al., 2011; Pickering & Garrod, 2013). Self-monitoring has been associated with cognitive control in both language production (Nozari & Novick, 2017), and comprehension (Musz & Thompson-Schill, 2017). Further investigation looking into speech production errors and repairs, and the timing of error detection and repair, could better pinpoint which exact production processes or components are associated with executive control than the current study.

Furthermore, for the non-pathological young adults tested in this thesis, task difficulty seemed to have influenced their articulatory control ability (as indexed by their maximum speech accuracy). Maximum performance of DDK non-words is arguably more demanding due to novel combinations of syllables that are less practiced by participants compared to real words. Previous studies looking into speech performance and executive control have found the effect of task and/or stimulus difficulty on participants' speech motor performance (Tremblay, et al., 2018, Tremblay et al., 2017). Their results revealed that the more demanding the speech task, the more cognitive resources are required to successfully execute the intended speech plans, especially for older adults. The results in this thesis also seemed to suggest this was the case, given the stronger association between cognitive switching and non-word DDK than between switching and real word DDK. However, bear in mind that alternative explanations for these results concerning the order of task administration and the intrinsically different prosodic patterns associated with different (non-word versus real word) task stimuli cannot be ruled out (see discussion in **Chapter 3**). If the results presented here are found to hold more generally and extend to older adults and groups with language or speech disorders, this stronger link of cognitive switching with non-word DDK compared to real word DDK could be a reason to opt for either type of DDK stimuli, depending on the purpose of the speech assessment. Either way, it may be good practice to use both non-word and real word stimuli in a DDK task to get a more complete picture of participants' speech motor/articulatory control skills.

One obvious limitation of the results on the association between cognitive switching and speech performance is that they are correlational in nature, such that no causal relationships can be deduced from them. One study investigated the effect of cognitive load on speech production during simulator flights (Huttunen et al., 2011). The cognitive load induced during different phases of the simulator flights was rated by their flight instructor using video-recorded flight data. The authors found correlations between cognitive load and speech production in both articulation rate and vowel formant measures. It is not entirely clear whether the cognitive load variation during the simulator flight in the Huttunen et al. (2011) study could be

viewed as a proxy for cognitive switching demand. Follow-up research with more controlled experimental manipulation of speech task/stimulus difficulty and of cognitive switching load (i.e., through using speech tasks that require more or less switching) is needed to further explore the relationships between executive control and articulatory control.

A further step to follow up on these switching results presented in **Chapter 3** is to analyse whether the association between cognitive switching and performance on *alternating* DDK sequences is indeed stronger than that between cognitive switching and performance on monosyllabic *(non-alternating)* DDK, as these non-alternating sequences (such as 'tatata' and 'kakaka') were also included in the DDK task of this study as practice trials. Recent findings from a master's degree thesis in our research group (De Kerf, 2021) suggest that this is indeed the case: cognitive switching ability predicts DDK accuracy for non-alternating DDK sequences, but it is more strongly associated with accuracy performance for alternating than non-alternating DDK sequences. This finding supports the assumption that participants' performance on the letter-number switching task is associated with resolving competition and selection during speech planning or speech monitoring stages.

The way in which articulatory control was quantified in **Chapters 2** and **3** also needs some further consideration. Production errors in both the DDK and the tongue twister tasks in this thesis were restricted to clearly audible errors, i.e., based on binary perceptual judgements (i.e., accurate or error). Several studies have shown that speech errors can occur at more fine-grained acoustic and/or articulatory levels than the segment level (Goldrick & Blumstein, 2006; Goldstein et al., 2007; McMillan & Corley, 2010). For instance, when producing tongue twisters containing syllable-initial stop consonants that only differ in voicing ('keff geff geff keff'), speakers tested in Goldrick and Blumstein (2006) exhibited longer voice-onset time of 'g' /g/ for 'k' /k/ errors than correctly produced 'g' tokens. Hence, the coarse way errors were coded in this thesis entails that subtler acoustic and/or articulatory errors or blends in participants' speech production were not taken into account. Future studies, regardless of whether investigating maximum speech performance or not, could consider analysing articulation itself, or articulation in combination with the acoustic/perceptual signal, to gauge speakers' articulatory control abilities and speech errors in more detail. Additionally, speech motor control research also has a research tradition of focusing on (group differences in) articulatory stability over speakers' repetitions of an utterance (Smith et al., 2000; Smith & Goffman, 1998; Van Brenk & Lowit, 2012). Hence, acoustic or kinematic indices of token-to-token stability (or, conversely, of repetition variability) could also be investigated (e.g., by detailed acoustic analysis of the RaMax corpus materials, cf. section 6.1.1), and related to executive control, instead of the accuracy and rate measures employed in this thesis.

6.4. Age effects in late stages of speech production

A few recent studies have suggested that (late stages) of speech production in cognitively healthy adults may be affected by age-related declines in the planning and execution of speech movements and speech motor performance (Tremblay, et al., 2018, Tremblay et al., 2017). However, as speech planning in Tremblay et al. (2017) and Tremblay et al. (2018) involved both retrieval and sequencing of motor programmes, it remains unclear which of these speech planning processes are subject to adult ageing. It is unclear whether adult ageing mainly affects retrieval of motor programmes, or the organisation of those programmes into a smooth sequence for articulation, or in fact both. The study presented in **Chapter 4** employed a simple/choice task reaction paradigm in which younger and older adults produced nonsensical target words either in a simple (prepared) condition or in a choice (unprepared) condition to see which speech condition would show the larger latency difference between age groups. Results from **Chapter 4** demonstrated that healthy older and younger adults differed more in the condition where pre-programming was possible, than in the condition where pre-programming was not possible. This finding was somewhat unexpected, as earlier work on speeded response tasks had shown stronger age effects in more complex conditions (e.g., which would arguably be the choice condition in the current study) (Der & Deary, 2006; Niermeyer et al., 2017). These results then suggest that if these results provide any evidence for age-related decline in speech production at all, this decline seems to affect the buffer capacity from which articulatory programmes are launched just prior to execution.

Similar to the results discussed previously for **Chapter 3**, some confounds may provide alternative explanations for the unexpected findings of **Chapter 4**. The age differences observed in the simple/choice reaction time paradigm could be due to the fixed presentation order employed in the experiment. As simple condition blocks always preceded the choice condition blocks, age differences may have been larger in the earlier (i.e., simple condition) blocks if older adults needed more practice to familiarise themselves with the stimuli. Another factor that could have contributed to the unexpected results may be age-related differences in task/motivational strategies or in task compliance (cf. e.g., Freund, 2006). For instance, younger and older participants may have made different choices in which aspects of task performance to focus on. Even though the results did not provide evidence for speed-accuracy trade-offs, younger adults could have been more focused on response speed than older adults. This was apparent in the prepared condition where more younger than older adults provided responses before they were actually allowed to. Follow-up research into potential age differences in the latest stages of speech planning may need to change the simple/choice paradigm to avoid such task or

motivational differences between age groups. One possibility to minimise such confounds might be to change the task instruction in the simple or prepared condition from waiting for a GO cue to follow a cue to either speak or not.

If the observation that younger adults are somehow more 'ready' than older adults to start their production in the simple condition is a 'real' finding, which is not due to age group or experimental design confounds, this may mean that buffering of speech motor programmes is the only age-related problem for which this thesis provides some evidence. This buffering problem in ageing does not arise in the choice condition, as the utterance is still relatively short and simple and needs to be produced immediately. Obvious follow-up research could therefore manipulate utterance length to see whether age differences indeed increase with longer utterance length where buffering can no longer be avoided. The 'buffer' finding may also relate to age-related changes in the relationship between vocal RT and vocal response duration as observed by Tremblay and colleagues (2018). Tremblay et al. (2018) found that younger adults tended to start articulating before motor planning was completed and to slow down during articulation to complete their motor planning. Conversely, older adults seemed to need to completely assemble the required motor programme(s) (of maximally three syllables long in Tremblay et al., 2018) before articulation. Future research could shed more light on age differences in the specific processes of speech motor planning and the possible overlap between subprocesses by looking into response duration of utterances in both prepared and unprepared conditions with counterbalanced presentation order, and by taking precautions to minimise age group differences in task strategies.

6.5. Inter- and intra-speaker differences in speech enrichment strategies and success

Speakers have been demonstrated to modify their speech in different communicative settings. Speakers also differ in the extent to which they can clarity their speech upon request. When speaking in (loud) noise, such 'enhanced' speech often requires extra articulatory and vocal effort from speakers (Hazan & Baker, 2011; Junqua, 1993; Picheny et al., 1986; Van Summers et al., 1988). Increased vocal effort has been linked to vocal fatigue and decreased vocal function (Bottalico et al., 2016; Solomon, 2008). Thus, in line with predictions of the H&H theory (Lindblom, 1990), speakers may be inclined to maintain speech clarity while spending as little vocal effort as possible. **Chapter 5** looked into speakers' speech enrichment capabilities in terms of consistency (i.e., the consistency of their habitual and clear-Lombard speaking styles over a list of sentences within an experimental session) and enrichment success

(i.e., the acoustic and intelligibility gains of the enriched speech as compared to plain speech).

Findings from **Chapter 5** speak to the H&H theory of speech production by showing that when producing enriched (i.e., clear-Lombard) speech, speakers displayed rather dynamic adaptation patterns, instead of showing a linear decrease or increase of speech enrichment modification over time. That is to say, speakers continuously adapted their speech enrichment modifications over the course of the experimental session, according to their perceived need to overcome the loud background noise (possibly for their own speech monitoring purpose, and/or for the benefit of their imaginary interlocutor), and to reduce effort of maintaining their enriched speech. For instance, for articulation rate, speakers initially speeded up over experimental trials in the habitual speaking style, while slowing down in the clear-Lombard style. These rate adjustments then levelled off at later trials in both speaking styles, making the overall rate difference between the two styles the largest at a point half-way in the sentence list. Additionally, both median F_0 (i.e., pitch) and vocal effort/energy (as indexed by spectral balance) exhibited nonlinear increase patterns over the clear-Lombard trials. These acoustic modifications over trials may have also contributed to the improvement over trials in high-energy glimpse proportion (HEGP) model-predicted intelligibility, as suggested by the correlations between acoustic parameters and predicted intelligibility in the clear-Lombard condition (cf. Table 5.3 in **Chapter 5**).

Similar evidence that speech enrichment is far from static was obtained in a recent study by Lee and Baese-Berk (2020), in which speakers' maintenance of clear speech production was investigated. Using an interactive speech task (i.e., Diapix), native English speakers in their study were found to be more intelligible in the early rather than the late portions of the conversation. Moreover, speakers in their study were found to be more intelligible at the beginning of each conversation, as if they 'reset' their speech to clear speech when a new conversation begins. However, different from Lee and Baese-Berk (2020), the results here showed that the (predicted) intelligibility increased continuously over the course of the experiment or until a tipping point halfway through the experiment. Note that the time course of changes cannot really be compared between the two studies. Still, maintaining a clear speaking style may be more effortful during spontaneous conversation than in the sentence-reading task implemented in this thesis. Linking back to Levelt's speech production model (Levelt, 1989; Levelt et al., 1999), spontaneous conversations involve all three main stages of speech production: i.e., from conceptualisation, to formulation, and then to articulation. A sentence-reading task, on the other hand, does not require speakers to think about what to say, such that more cognitive resources may be left for careful

articulation and monitoring of one's own speaking style. However, cognitive demand may not be the factor that influences speakers' speech enrichment behaviour. A very recent study by Tuomainen and colleagues (2021) investigated speech modification speakers applied when communicating in quiet and noisy (i.e., non-speech and background speech noise) conditions. The authors found that both older adults and younger children increased their vocal effort the most in the more distracting background-speech noise condition. This finding suggests that despite the higher cognitive load required to inhibit interfering background speech noise, speakers were still trying hard and spending more vocal effort in order to maintain communicative success.

This speech behaviour fits in with the 'communication effort' framework proposed by Beechey et al. (2020). In this framework, speakers constantly monitor their own speech output as well as the feedback from their interlocutors to optimise communicative success. Coming back to the dynamic nature of speech enhancement, as observed in this thesis, and following the 'communication effort' framework, speakers may be (immediately) able to use their internal knowledge and experience of what adjustments can make speech better intelligible for their interlocutor(s), yet still be able to improve their attempt at producing clear speech through practice. Future studies might develop ways to systematically compare speech enhancement, and changes thereof over time, between conditions that do involve formulation stages with conditions that do not (such as scripted speech conditions), and between conditions that do or do not involve a real interlocutor. Such systematic comparisons would be needed to investigate interactions between cognitive demands for formulation and articulation during spontaneous speech in communicative settings.

The findings reported in **Chapter 5** also indicate that speakers (at least the healthy young adults tested here) may need some practice to reach their full potential in producing the enriched clear-Lombard speech. This, in turn, challenges the automatic reflex view of Lombard speech production, but note again that our 'Lombard' speech condition was a mixture of Lombard and instructed clear speech. Although the speakers tested here immediately spoke more loudly, slowly, and clearly in the clear-Lombard condition, they were able to keep improving their speech enrichment modifications with prolonged practice. This may be good news for those with hearing impairment, as it seems to suggest that clear speaking style is, at least to some extent, trainable (cf. Ferguson & Morgan, 2018).

We know from the various speech production models that speakers are able to monitor and adapt their speech output through auditory (and somatosensory) feedback (Guenther et al., 2006; Guenther, 2016; Levelt, 1989; Levelt et al., 1999;

Tourville & Guenther, 2011). Studies on the effects of altered auditory and/or somatosensory feedback have found a 'practice effect' in speakers' adapted speech production (e.g., McFarland et al., 1996; Purcell & Munhall, 2006), with speakers applying larger changes over time. The fact that speakers would need some time to adjust their speech production could be caused by limited auditory feedback in this study (as opposed to typical altered auditory feedback studies where speakers hear their own manipulated speech well). This limited auditory feedback entails that, speakers had to adapt their speaking style while being exposed to loud noise played through headphones. In the current experimental setting, it is possible that speakers kept on adjusting their speech output based on either limited auditory or somatosensory information about how well their articulation targets were being met, or both.

This reliance on auditory or somatosensory information relates to a recent study on Lombard speech production in L1 and L2 (Cai et al., 2021). Speakers in the Cai et al. (2021) study exhibited larger Lombard effects when speaking in their L2 English than in their native language (or L1, i.e., Chinese), in both weak and strong noise conditions. Their results suggest that L2 speech motor control relies on auditory feedback to a larger extent than L1, as motor commands for L1 speech sounds are more rehearsed and 'entrenched'. Consequently, this would mean that the native Dutch speakers in our study may have mainly relied on somatosensory information for their clear-Lombard adjustments, which was reinforced by the use of closed headphones. Future studies would need to implement a condition where speakers can hear their own speech adjustments better. Only such an experimental manipulation of the availability of auditory feedback could test the role of auditory feedback on the time course of native speakers' acoustic clear-Lombard adaptation in noisy communicative settings.

Even though there was ample evidence of speaker variability in predicted intelligibility of speakers' habitual and clear-Lombard speech, and in the degree of speakers' clear-speech enhancement, there was no clear relationship between intelligibility or speech enhancement on the one hand, and articulatory control ability on the other. Note that intelligibility here refers to model-predicted intelligibility as intelligibility was not assessed through ratings or sentence identification tests with human listeners (although pilot data referred to in **Chapter 4** suggest a high correlation between model-predicted intelligibility and human ratings of listening effort; and see the RaLoCo human ratings that have been collected outside of the scope of this thesis; cf. section 6.1.2). Consequently, future investigation can show whether speakers' intelligibility as rated by listeners correlates with measures of their articulatory control.

Speakers' individual speech motor control ability (as indexed by DDK rate) was found to only predict articulation rate (in both habitual and clear-Lombard style). This result indicates a link between the maximum rate at which speakers can move their articulators and their 'normal' articulation rate during sentence reading. As such, speaking rate may be considered an individual speech 'trait'. One potential reason for why the DDK measure of speech motor control did not relate to other aspects of speakers' speech enrichment modifications could be due to its maximum performance nature: stressing maximum speed rather than intelligibility or communicative intent. As mentioned earlier, other aspects of speech motor control such as stability and consistency could be measured in future studies, which could be used to further explore the relationship between speech motor control and (enriched) speech production in the two speech corpora that this thesis compiled (cf. section 6.1). Additionally, individual differences in 'helpful' speech behaviour may link to other factors than motor control. Speakers have been shown to take listeners' mental state and their access to (shared) information into consideration during communication (cf. the H&H theory mentioned earlier and also in Lindblom, 1990). Speakers' ability to take listeners' perspective into account has been related to 'theory of mind' (ToM), also in studies on phonetic phenomena (Turnbull, 2019). Turnbull explored the role of ToM in phonetic reduction, aiming to test for the relationship between phonetic reduction upon second mention in discourse and individual variation in theory of mind. He found, however, that ToM was not systematically correlated with phonetic reduction. Similar future experiments could shed more light on whether speakers' ToM relates to their speech enhancement behaviour in communication. Lastly, investigation on aspects of speakers' anatomical and physiological constraints especially in (prolonged) enriched speech (e.g., being able to sustain high vocal effort in noisy communication settings) production could also help further explore the nature of adaptive behaviour in speech production.

6.6. Conclusion

Individual speakers exhibit wide variability in their speech during communication. This doctoral thesis provided evidence that articulation, similar to earlier stages of speech production such as conceptualisation and formulation, is associated with cognitive control, at least when articulation is quantified as maximum speech performance. This thesis also demonstrated the overlap between clinical measures of maximum speech performance and psycholinguistic measures of speech performance. Additionally, speakers were found to show a dynamic pattern of acoustic speech adjustment in noisy speech conditions, following both speaker- and listener-oriented patterns proposed by the hyper- and hypo-articulation theory. In order to implement

these acoustic adjustments, speakers may have relied on somatosensory and the (sometimes limited) auditory self-monitoring of their speech output, as well as on their internal knowledge and experience of what adjustments can make speech better intelligible for their interlocutor. The speech corpora that are published with this thesis open up opportunities for answering more questions about between-speaker differences in diverse speaking conditions, and their effects on listeners.

References

Acheson, D. J., & Hagoort, P. (2014). Twisting tongues to test for conflict-monitoring in speech production. *Frontiers in Human Neuroscience*, *8*, 1–16. https://doi.org/10.3389/fnhum.2014.00206

Acheson, D. J., & MacDonald, M. C. (2009a). Twisting tongues and memories: Explorations of the relationship between language production and verbal working memory. *Journal of Memory and Language*, *60*(3), 329–350. https://doi.org/10.1016/j.jml.2008.12.002

Acheson, D. J., & MacDonald, M. C. (2009b). Verbal working memory and language production: common approaches to the serial ordering of verbal information. *Psychological Bulletin*, *135*(1), 50–68. https://doi.org/10.1037/a0014411

Amazi, D., & Garber, S. (1982). The Lombard sign as a function of age and task. *Journal of Speech, Language, and Hearing Research*, *25*(4), 581–585. https://doi.org/10.1044/jshr.2504.581

Baayen, H. R. (2013). *languageR: Data sets and functions with "Analyzing linguistic data: A practical introduction to statistics"*. https://cran.r-project.org/package=languageR

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

Baddeley, A. D., & Della Sala, S. (1996). Working memory and executive control. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1346), 1397–1404. https://doi.org/10.1098/rstb.1996.0123

Bailey, D. J., & Dromey, C. (2015). Bidirectional interference between speech and nonspeech tasks in younger, middle-aged, and older adults. *Journal of Speech, Language, and Hearing Research*, *58*(6), 1637–1653. https://doi.org/10.1044/2015_JSLHR-S-14-0083

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Barbosa, A. F., Voos, M. C., Chen, J., Francato, D. C. V., Souza, C. D. O., Barbosa, E. R., ... & Mansur, L. L. (2017). Cognitive or cognitive-motor executive function tasks? Evaluating verbal fluency measures in people with Parkinson's disease. *BioMed Research International*, 2017. https://doi.org/10.1155/2017/7893975

Beechey, T., Buchholz, J. M., & Keidser, G. (2020). Hearing impairment increases communication effort during conversations in noise. *Journal of Speech, Language, and Hearing Research*, *63*(1), 305–320. https://doi.org/10.1044/2019_JSLHR-19-00201

Ben-David, B. M., & Icht, M. (2017). Oral-diadochokinetic rates for Hebrew-speaking healthy ageing population: non-word versus real-word repetition. *International Journal of Language and Communication Disorders*, *52*(3), 301–310. https://doi.org/10.1111/1460-6984.12272

Bernthal, J. E., Bankson, N. W., & Flipsen, P. Jr. (2009). *Articulation and phonological disorders: Speech sound disorders in children* (6th ed.). Pearson/Allyn & Bacon.

Boersma, P., & Weenink, D. (2017). *Praat: doing phonetics by computer [Computer program]. Version 6.0.36*. http://www.praat.org/

Bond, Z. S., & Moore, T. J. (1994). A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech Communication*, *14*(4), 325–337. https://doi.org/10.1016/0167-6393(94)90026-4

Bond, Z. S., Moore, T. J., & Gable, B. (1989). Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask. *The Journal of the Acoustical Society of America*, *85*(2), 907–912. https://doi.org/10.1121/1.397563

Bosker, H. R., & Cooke, M. (2020). Enhanced amplitude modulations contribute to the Lombard intelligibility benefit: Evidence from the Nijmegen Corpus of Lombard Speech. *The Journal of the Acoustical Society of America*, *147*(2), 721–730. https://doi.org/10.1121/10.0000646

Bottalico, P., Graetzer, S., & Hunter, E. J. (2016). Effects of speech style, room acoustics, and vocal fatigue on vocal effort. *The Journal of the Acoustical Society of America*, *139*(5), 2870–2879. https://doi.org/10.1121/1.4950812

Bowles, N. L., & Poon, L. W. (1985). Aging and retrieval of words in semantic memory. *Journal of Gerontology*, *40*(1), 71–77. https://doi.org/10.1093/geronj/40.1.71

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America*, *112*(1), 272–284. https://doi.org/10.1121/1.1487837

Bradlow, A. R., Blasingame, M., & Lee, K. (2018). Language-independent talker-specificity in bilingual speech intelligibility: Individual traits persist across first-language and second-language speech. *Laboratory Phonology*, *9*(1), p.17. https://doi.org/10.5334/labphon.137

Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research*, *46*(1), 80–97. https://doi.org/10.1044/1092-4388(2003/007)

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, *20*(3-4), 255–272. https://doi.org/10.1016/S0167-6393(96)00063-5

Brocklehurst, P. H., & Corley, M. (2011). Investigating the inner speech of people who stutter: Evidence for (and against) the Covert Repair Hypothesis. *Journal of Communication Disorders*, *44*(2), 246–260. https://doi.org/10.1016/j.jcomdis.2010.11.004

Bruce, P., & Bruce, A. (2017). *Practical statistics for data scientists*. O'Reilly Media.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language*, *30*(5), 542–579. https://doi.org/10.1016/0749-596X(91)90026-G

Cai, X., Yin, Y., & Zhang, Q. (2021). Online control of voice intensity in late bilinguals' first and second language speech production: Evidence from unexpected and brief noise masking. *Journal of Speech, Language, and Hearing Research*, *64*(5), 1471-1489. https://doi.org/10.1044/2021_JSLHR-20-00330

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, *119*(3), 1562–1573. https://doi.org/10.1121/1.2166600

Cooke, M., & Lecumberri, M. L. G. G. (2012). The intelligibility of Lombard speech for non-native listeners. *The Journal of the Acoustical Society of America*, *132*(2), 1120–1129. https://doi.org/10.1121/1.4732062

Cooke, M., & Lu, Y. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *The Journal of the Acoustical Society of America*, *128*(4), 2059–2069. https://doi.org/10.1121/1.3478775

Cooke, M., King, S., Garnier, M., & Aubanel, V. (2014). The listening talker: A review of human and algorithmic context-induced modifications of speech. *Computer Speech and Language*, *28*(2), 543–571. https://doi.org/10.1016/j.csl.2013.08.003

Costa, A., Santesteban, M., & Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1057–1074. https://doi.org/10.1037/0278-7393.32.5.1057

De Kerf, E. (2021). *Het verband tussen het cognitieve switchingvermogen en articulatorische controle bij het snel produceren van alternerende en niet alternerende syllabereeksen door jongvolwassen sprekers* [Unpublished Master thesis]. Radboud University.

Deger, K., & Ziegler, W. (2002). Speech motor programming in apraxia of speech. *Journal of Phonetics*, *30*(3), 321–335. https://doi.org/10.1006/jpho.2001.0163

Dell, G. S. (1986). A spreading-activation model of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

De Jong, N. H., & Mora, J. C. (2019). Does having good articulatory skills lead to more fluent speech in first and second languages? *Studies in Second Language Acquisition*, *41*(1), 227–239. https://doi.org/10.1017/S0272263117000389

Der, G., & Deary, I. J. (2006). Age and sex differences in reaction time in adulthood: Results from the United Kingdom health and lifestyle survey. *Psychology and Aging*, *21*(1), 62–73. https://doi.org/10.1037/0882-7974.21.1.62

Der, G., & Deary, I. J. (2017). The relationship between intelligence and reaction time varies with age: Results from three representative narrow-age age cohorts at 30, 50 and 69 years. *Intelligence*, *64*(July), 89–97. https://doi.org/10.1016/j.intell.2017.08.001

Dichter, B. K., Breshears, J. D., Leonard, M. K., & Chang, E. F. (2018). The control of vocal pitch in human laryngeal motor cortex. *Cell*, *174*(1), 21–31.e9. https://doi.org/10.1016/j.cell.2018.05.016

Dromey, C., & Benson, A. (2003). Effects of concurrent motor, linguistic, or cognitive tasks on speech motor performance. *Journal of Speech, Language, and Hearing Research*, *46*(5), 1234–1246. https://doi.org/10.1044/1092-4388(2003/096)

Duffy, J. R. (2013). *Motor speech disorders: Substrates, differential diagnosis, and management* (3rd ed.). Elsevier.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *The Journal of the Acoustical Society of America*, *116*(4), 2365–2373. https://doi.org/10.1121/1.1788730

Ferguson, S. H. (2012). Talker differences in clear and conversational speech: Vowel intelligibility for older adults with hearing loss. *Journal of Speech, Language, and Hearing Research*, *55*(3), 779–790. https://doi.org/10.1044/1092-4388(2011/10-0342)

Ferguson, S. H., & Morgan, S. D. (2018). Acoustic and perceptual correlates of subjectively rated sentence clarity in clear and conversational speech. *Journal of Speech, Language, and Hearing Research*, *61*(1), 159–173. https://doi.org/10.1044/2017_JSLHR-H-17-0082

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1187–1199. https://doi.org/10.1037//0278-7393.28.6.1187

Fletcher, S. G. (1972). Time-by-count measurement of Diadochokinetic syllable rate. *Journal of Speech, Language, and Hearing Research*, *15*(4), 763–771. https://doi.org/10.1044/jshr.1504.763

Fowler, C. A., & Turvey, M. T. (1980). Immediate compensation in bite-block speech. *Phonetica*, *37*(5-6), 306–326. https://doi.org/10.1159/000260000

Freund, A. M. (2006). Age-differential motivational consequences of optimization versus compensation focus in younger and older adults. *Psychology and Aging*, *21*(2), 240–252. https://doi.org/10.1037/0882-7974.21.2.240

Garnier, M., & Henrich, N. (2014). Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Computer Speech and Language*, *28*(2), 580–597. https://doi.org/10.1016/j.csl.2013.07.005

Garnier, M., Dohen, M., Loevenbruck, H., Welby, P., & Bailly, L. (2008). The Lombard Effect: a physiological reflex or a controlled intelligibility enhancement? *7th International Seminar on Speech Production*, 255–262. https://doi.org/10.1044/1092-4388(2009/08-0138)

Garnier, M., Henrich, N., & Dubois, D. (2010). Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language, and Hearing Research*, *53*(3), 588–608. https://doi.org/10.1044/1092-4388(2009/08-0138)

Garrod, S., & Pickering, M. J. (2007). Automaticity of language production in monologue and dialogue. In A. Meyer, L. Wheeldon, & A. Krott (Eds.), *Automaticity and control in language processing* (1st ed., pp. 1–20). Psychology Press. https://doi.org/10.4324/9780203968512

Gauvin, H. S., & Hartsuiker, R. J. (2020). Towards a new model of verbal monitoring. *Journal of Cognition*, *3*(1), 1–37. https://doi.org/10.5334/joc.81

Gelfer, M. P., Andrews, M. L., & Schmidt, C. P. (1991). Effects of prolonged loud reading on selected measures of vocal function in trained and untrained singers. *Journal of Voice*, *5*(2), 158–167. https://doi.org/10.1016/S0892-1997(05)80179-1

Gilbert, S. J., & Burgess, P. W. (2008). Executive function. *Current Biology*, *18*(3), R110–114. https://doi.org/10.1016/j.cub.2007.12.014

Glisky, E. L. (2007). Changes in cognitive function in human aging. In D. R. Riddle (Ed.), *Brain aging: Models, methods, and mechanisms*. CRC Press Press/Taylor & Francis.

Goldrick, M., & Blumstein, S. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(6), 649–683. https://doi.org/10.1080/01690960500181332

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, *103*(3), 386–412. https://doi.org/10.1016/j.cognition.2006.05.010

Guenther, F. H. (2016). *Neural control of speech*. MIT Press.

Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, *25*(5), 408–422. https://doi.org/10.1016/j.jneuroling.2009.08.006

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*(3), 280–301. https://doi.org/10.1016/j.bandl.2005.06.001

Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J., & Wedin, L. (1980). Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngology*, *90*(5-6), 441–451.

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustical Society of America*, *130*(4), 2139–2152. https://doi.org/10.1121/1.3623753

Hazan, V., & Markham, D. (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *The Journal of the Acoustical Society of America*, *116*(5), 3108–3118. https://doi.org/10.1121/1.1806826

Hazan, V., Grynpas, J., & Baker, R. (2012). Is clear speech tailored to counter the effect of specific adverse listening conditions? *The Journal of the Acoustical Society of America*, *132*(5), EL371-EL377. https://doi.org/10.1121/1.4757698

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), nrn3158. https://doi.org/10.1038/nrn3158

Huettig, F., & Hartsuiker, R. J. (2010). Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. *Language and Cognitive Processes*, *25*(3), 347–374. https://doi.org/10.1080/01690960903046926

Hughes, A., Trudgill, P., & Watt, D. (2012). *English accents and dialects: an introduction to social and regional varieties of English in the British Isles* (5th ed.). Routledge. https://doi.org/10.4324/9780203784440

Huttunen, K. H., Keränen, H. I., Pääkkönen, R. J., Päivikki Eskelinen-Rönkä, R., & Leino, T. K. (2011). Effect of cognitive load on articulation rate and formant frequencies during simulator flights. *The Journal of the Acoustical Society of America*, *129*(3), 1580-1593. https://doi.org/10.1121/1.3543948

Icht, M., & Ben-David, B. M. (2015). Oral-diadochokinetic rates for Hebrew-speaking school-age children: Real words vs. non-words repetition. *Clinical Linguistics and Phonetics*, *29*(2), 102–114. https://doi.org/10.3109/02699206.2014.961650

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1-2), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

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

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology*, *14*, 5–82.

Jongman, S. R., Roelofs, A., & Meyer, A. S. (2015). Sustained attention in language production: An individual differences investigation. *Quarterly Journal of Experimental Psychology*, *68*(4), 710–730. https://doi.org/10.1080/17470218.2014.964736

Junqua, J. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *The Journal of the Acoustical Society of America*, *93*(1), 510–524. https://doi.org/10.1121/1.405631

Kent, R. D. (2000). Research on speech motor control and its disorders: A review and prospective. *Journal of Communication Disorders*, *33*(5), 391–427. https://doi.org/10.1016/S0021-9924(00)00023-X

Kent, R. D. (2004). Models of speech motor control: implications from recent developments in neurophysiological and neurobehavioral science. In B. Maassen, R. Kent, H. Peters, P. van Lieshout, & W. Hulstijn (Eds.), *Speech motor control in normal and disordered speech* (1st ed., pp. 3–28). Oxford University Press.

Kent, R. D., Kent, J. F., & Rosenbek, J. C. (1987). Maximum performance tests of speech production. *Journal of Speech and Hearing Disorders*, *52*(4), 367–387. https://doi.org/10.1044/jshd.5204.367

Kim, J., & Davis, C. (2014). Comparing the consistency and distinctiveness of speech produced in quiet and in noise. *Computer Speech and Language*, *28*(2), 598–606. https://doi.org/10.1016/j.csl.2013.02.002

King, B. R., Fogel, S. M., Albouy, G., & Doyon, J. (2013). Neural correlates of the age-related changes in motor sequence learning and motor adaptation in older adults. *Frontiers in Human Neuroscience*, *7*, 1–13. https://doi.org/10.3389/fnhum.2013.00142

Klapp, S. T. (1995). Motor response programming during simple and choice reaction time: The role of practice. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(5), 1015–1027. https://doi.org/10.1037/0096-1523.21.5.1015

Klapp, S. T. (2003). Reaction time analysis of two types of motor preparation for speech articulation: Action as a sequence of chunks. *Journal of Motor Behavior*, *35*(2), 135–150. https://doi.org/10.1080/00222890309602129

Knuijt, S., Kalf, J., Van Engelen, B., Geurts, A., & Swart, B. de. (2017). Reference values of maximum performance tests of speech production. *International Journal of Speech-Language Pathology*, *21*(1), 56–64. https://doi.org/10.1080/17549507.2017.1380227

Krampe, R. T., Kliegl, R., & Mayr, U. (2005). Timing, sequencing, and executive control in repetitive movement production. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(3), 379–397. https://doi.org/10.1037/0096-1523.31.3.379

Krause, J. C., & Braida, L. D. (2004). Properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, *115*(1), 362–378. https://doi.org/10.1121/1.416659

Krueger, M., Schulte, M., Zokoll, M. A., Wagener, K. C., Meis, M., Brand, T., & Holube, I. (2017). Relationship between listening effort and speech intelligibility in noise. *American Journal of Audiology*, *26*(3S), 378–392. https://doi.org/10.1044/2017_AJA-16-0136

Kryter, K. D. (1962a). Methods for the calculation and use of the Articulation Index. *The Journal of the Acoustical Society of America*, *34*(11), 1689–1697. https://doi.org/10.1121/1.1909094

Kryter, K. D. (1962b). Validation of the Articulation Index. *The Journal of the Acoustical Society of America*, *34*(11), 1698–1702. https://doi.org/10.1121/1.1909096

Ladefoged, P., & Disner, S. F. (2012). *Vowels and consonants*. John Wiley & Sons.

Lam, J., & Tjaden K. (2013). Intelligibility of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1429–1440. https://doi.org/10.1044/1092-4388(2013/12-0335)

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, *14*(4), 677–709. https://doi.org/10.1044/jshr.1404.677

Lee, D.-Y., & Baese-Berk, M. M. (2020). The maintenance of clear speech in naturalistic conversations. *The Journal of the Acoustical Society of America*, *147*(5), 3702–3711. https://doi.org/10.1121/10.0001315

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

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–75. https://doi.org/10.1017/S0140525X99001776

Lima, S. D., Hale, S., & Myerson, J. (1991). How general is general slowing? Evidence from the lexical domain. *Psychology and Aging*, *6*(3), 416–425. https://doi.org/10.1037/0882-7974.6.3.416

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In H. W. J. & M. A. (Eds.), *Speech production and speech modelling* (pp. 403–439). Springer.

Lively, S. E., Pisoni, D. B., Van Summers, W., & Bernacki, R. H. (1993). Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. *Journal of the Acoustical Society of America*, *93*(5), 2962–2973. https://doi.org/10.1121/1.405815

Logan, G. D. (1985). Executive control of thought and action. *Acta Psychologica*, *60*(2-3), 193–210. https://doi.org/10.1016/0001-6918(85)90055-1

Lombard, É. (1911). Le signe de l'elevation de la voix. *Annales Des Maladies de L'Oreille Et Du Larynx*, *37*, 101–109.

Lu, Y., & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *The Journal of the Acoustical Society of America*, *124*(5), 3261–3275. https://doi.org/10.1121/1.2990705

Lu, Y., & Cooke, M. (2009). The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise. *Speech Communication*, *51*(12), 1253–1262. https://doi.org/10.1016/j.specom.2009.07.002

Maas, E. (2017). Speech and nonspeech: What are we talking about? *International Journal of Speech-Language Pathology*, *19*(4), 345–359. https://doi.org/10.1080/17549507.2016.1221995

Maas, E., & Mailend, M. (2012). Speech planning happens before speech execution: Online reaction time methods in the study of apraxia of speech. *Journal of Speech, Language, and Hearing Research*, *55*(October), 1523–1535. https://doi.org/10.1044/1092-4388(2012/11-0311)

Maas, E., Robin, D. A., Wright, D. L., & Ballard, K. J. (2008). Motor programming in apraxia of speech. *Brain and Language*, *106*(2), 107–118. https://doi.org/10.1016/j.bandl.2008.03.004

Machač, P., & Skarnitzl, R. (2009). *Principles of phonetic segmentation*. Epocha Publishing House.

MacPherson, M. K. (2019). Cognitive load affects speech motor performance differently in older and younger adults. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1258–1277. https://doi.org/10.1044/2018_JSLHR-S-17-0222

Mailend, M. L., Maas, E., Beeson, P. M., Story, B. H., & Forster, K. I. (2019). Speech motor planning in the context of phonetically similar words: Evidence from apraxia of speech and aphasia. *Neuropsychologia*, *127*(February), 171–184. https://doi.org/10.1016/j.neuropsychologia.2019.02.018

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America*, *125*(6), 3962–3973. https://doi.org/10.1121/1.2990715

Marcoux, K., & Ernestus, M. (2019). Pitch in native and non-native Lombard speech. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th international congress of phonetic sciences* (pp. 2605–2609). Canberra, Australia: Australasian Speech Science; Technology Association Inc.

Mayr, U., & Kliegl, R. (2000). Complex semantic processing in old age: Does it stay or does it go? *Psychology and Aging, 15*(1), 29–43. https://doi.org/10.1037//0882-7974.15.1.29

McCloy, D. R. (2016). *phonR: tools for phoneticians and phonologists. [R package]. Version 1.0-7.* https://rdrr.io/cran/phonR/

McFarland, D. H., Baum, S. R., & Chabot, C. (1996). Speech compensation to structural modifications of the oral cavity. *The Journal of the Acoustical Society of America, 100*(2), 1093–1104. https://doi.org/10.1121/1.416286

McMillan, C. T., & Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition, 117*(3), 243–260. https://doi.org/10.1016/j.cognition.2010.08.019

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*(1), 8–14. https://doi.org/10.1177/0963721411429458.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100. https://doi.org/10.1006/COGP.1999.0734

Moers, C., Meyer, A., & Janse, E. (2017). Effects of word frequency and transitional probability on word reading durations of younger and older speakers. *Language and Speech, 60*(2), 289–317. https://doi.org/10.1177/0023830916649215

Monaco, E., Pellet Cheneval, P., & Laganaro, M. (2017). Facilitation and interference of phoneme repetition and phoneme similarity in speech production. *Language, Cognition and Neuroscience, 32*(5), 650–660. https://doi.org/10.1080/23273798.2016.1257730

Moyer, A. (2013). *Foreign accent: The phenomenon of non-native speech*. Cambridge University Press.

Musz, E., & Thompson-Schill, S. L. (2017). Tracking competition and cognitive control during language comprehension with multi-voxel pattern analysis. *Brain and Language,* 165, 21–32. https://doi.org/10.1016/j.bandl.2016.11.002

Murman, D. L. (2015). The impact of age on cognition. *Seminars in Hearing, 36*(3), 111–121. https://doi.org/10.1055/ s-0035-1555115

Netelenbos, N., Gibb, R. L., Li, F., & Gonzalez, C. L. (2018). Articulation speaks to executive function: An investigation in 4-to 6-year-olds. *Frontiers in Psychology, 9*, 172. https://doi.org/10.3389/fpsyg.2018.00172

Niermeyer, M. A., Suchy, Y., & Ziemnik, R. E. (2017). Motor sequencing in older adulthood: relationships with executive functioning and effects of complexity. *The Clinical Neuropsychologist, 31*(3), 598–618. https://doi.org/10.1080/13854046.2016.1257071

Nijland, L., Terband, H., & Maassen, B. (2015). Cognitive functions in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 58*(3), 550–565. https://doi.org/10.1044/2015_JSLHR-S-14-0084

Nooteboom, S. G., & Quené, H. (2017). Self-monitoring for speech errors: Two-stage detection and repair with and without auditory feedback. *Journal of Memory and Language, 95*, 19–35. https://doi.org/10.1016/j.jml.2017.01.007

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology, 63*(1), 1–33. https://doi.org/10.1016/j.cogpsych.2011.05.001

Nozari, N., & Novick, J. (2017). Monitoring and control in language production. *Current Directions in Psychological Science, 26*(5), 403–410. https://doi.org/10.1177/0963721417702419

Novak, A., Dlouha, O., Capkova, B., & Vohradnik, M. (1991). Voice fatigue after theater performance in actors. *Folia Phoniatrica Et Logopaedica, 43*(2), 74–78.

Peter, B., Lancaster, H., Vose, C., Middleton, K., & Stoel-Gammon, C. (2018). Sequential processing deficit as a shared persisting biomarker in dyslexia and childhood apraxia of speech. *Clinical Linguistics & Phonetics, 32*(4), 316–346. https://doi.org/10.1080/02699206.2017.1375560

Piai, V., & Roelofs, A. (2013). Working memory capacity and dual-task interference in picture naming. *Acta Psychologica, 142*(3), 332–342. https://doi.org/10.1016/j.actpsy.2013.01.006

Pick, H. L., Siegel, G. M., Fox, P. W., Garber, S. R., & Kearney, J. K. (1989). Inhibiting the Lombard effect. *The Journal of the Acoustical Society of America*, *85*(2), 894–900. https://doi.org/10.1121/1.397561

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech, Language, and Hearing Research*, *29*(4), 434–446. https://doi.org/10.1044/jshr.2904.434

Pittman, A. L., & Wiley, T. L. (2001). Recognition of speech produced in noise. *Journal of Speech, Language, and Hearing Research 44*(3), 487–496. https://doi.org/10.1044/1092-4388(2001/038)

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*(1), 25–42. https://doi.org/10.1146/annurev.neuro.13.1.25

Purcell, D. W., & Munhall, K. G. (2006). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, *120*(2), 966–977. https://doi.org/10.1121/1.2217714

Reilly, K. J., & Spencer, K. A. (2013a). Sequence complexity effects on speech production in healthy speakers and speakers with hypokinetic or ataxic dysarthria. *PLoS ONE*, *8*(10), 1–14. https://doi.org/10.1371/journal.pone.0077450

Reilly, K. J., & Spencer, K. A. (2013b). Speech serial control in healthy speakers and speakers with hypokinetic or ataxic dysarthria: Effects of sequence length and practice. *Frontiers in Human Neuroscience*, *7*(OCT), 1–17. https://doi.org/10.3389/fnhum.2013.00665

Rodriguez-Fornells, A., de Diego Balaguer, R., & Münte, T. F. (2006). Executive control in bilingual language processing. *Language Learning*, *56*(1), 133–190. https://doi.org/10.1111/j.1467-9922.2006.00359

Roelofs, A. (2008). Attention, gaze shifting, and dual-task interference from phonological encoding in spoken word planning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1580–1598. https://doi.org/10.1037/a0012476

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*(2), 207–231. https://doi.org/10.1037/0096-3445.124.2.207

Sadagopan, N., & Smith, A. (2013). Age differences in speech motor performance on a novel speech task. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1552–1566. https://doi.org/10.1044/1092-4388(2013/12-0293)

Salami, A., Eriksson, J., & Nyberg, L. (2012). Opposing effects of aging on large-scale brain systems for memory encoding and cognitive control. *Journal of Neuroscience*, *32*(31), 10749–10757. https://doi.org/10.1523/JNEUROSCI.0278-12.2012

Saletta, M., Goffman, L., Ward, C., & Oleson, J. (2018). Influence of language load on speech motor skill in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *61*(3), 675–689. https://doi.org/10.1044/2017_jslhr-l-17-0066

Shafto, M. A., Stamatakis, E. A., Tam, P. P., & Tyler, L. K. (2010). Word retrieval failures in old age: The relationship between structure and function. *Journal of Cognitive Neuroscience*, *22*(7), 1530–1540. https://doi.org/10.1162/jocn.2009.21321

Shao, Z., Roelofs, A., & Meyer, A. S. (2012). Sources of individual differences in the speed of naming objects and actions: The contribution of executive control. *Quarterly Journal of Experimental Psychology*, *65*(10), 1927–1944. https://doi.org/10.1080/17470218.2012.670252

Shen, C., & Janse, E. (2019). Articulatory control in speech production. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 2533–2537). Canberra, Australia: Australasian Speech Science; Technology Association Inc.

Shen, C., & Janse, E. (2020). Maximum speech performance and executive control in young adult speakers. *Journal of Speech, Language, and Hearing Research*, *63*(11), 3611–3627. https://doi.org/10.1044/2020_JSLHR-19-00257

Shriberg, L. D., Lohmeier, H. L., Strand, E. A., & Jakielski, K. J. (2012). Encoding, memory, and transcoding deficits in childhood apraxia of speech. *Clinical Linguistics and Phonetics*, *26*(5), 445–482. https://doi.org/10.3109/02699206.2012.655841

Sikora, K., Roelofs, A., Hermans, D., & Knoors, H. (2016). Executive control in spoken noun-phrase production: Contributions of updating, inhibiting, and shifting. *The Quarterly Journal of Experimental Psychology*, *69*(9), 1719–1740. https://doi.org/10.1080/17470218.2015.1093007

Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America*, *118*(3), 1677–1688. https://doi.org/10.1121/1.2000788

Smith, A., & Goffman, L. (1998). Stability and patterning of speech movement sequences in children and adults. *Journal of Speech, Language, and Hearing Research*, *41*(1), 18–30. https://doi.org/10.1044/jslhr.4101.18

Smith, A., Sadagopan, N., Walsh, B., & Weber-Fox, C. (2010). Increasing phonological complexity reveals heightened instability in inter-articulatory coordination in adults who stutter. *Journal of Fluency Disorders*, *35*(1), 1–18. https://doi.org/10.1016/j.jfludis.2009.12.001

Solomon, N. P. (2008). Vocal fatigue and its relation to vocal hyperfunction. *International Journal of Speech-Language Pathology*, *10*(4), 254–266. https://doi.org/10.1080/14417040701730990

Sparto, P. J., Fuhrman, S. I., Redfern, M. S., Perera, S., Jennings, J. R., & Furman, J. M. (2014). Postural adjustment errors during lateral step initiation in older and younger adults. *Experimental Brain Research*, *232*(12), 3977–3989. https://doi.org/10.1007/s00221-014-4081-z

Staiger, A., Schölderle, T., Brendel, B., & Ziegler, W. (2017). Dissociating oral motor capabilities: Evidence from patients with movement disorders. *Neuropsychologia*, *95*, 40–53. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA. 2016.12.010

Steeneken, H. J. M., & Houtgast, T. (1980). A physical method for measuring speech-transmission quality. *The Journal of the Acoustical Society of America*, *67*(1), 318–326. https://doi.org/10.1121/1.384464

Tang, Y., & Cooke, M. (2012). Optimised spectral weightings for noise-dependent speech intelligibility enhancement. *Proceedings of the Annual Conference of the International Speech Communication Association*, 955–958.

Tang, Y., & Cooke, M. (2016). Glimpse-based metrics for predicting speech intelligibility in additive noise conditions. *Proceedings of the Annual Conference of the International Speech Communication Association*, 2488–2492.

Tang, Y., Cooke, M., & Valentini-Botinhao, C. (2016). Evaluating the predictions of objective intelligibility metrics for modified and synthetic speech. *Computer Speech and Language*, *35*, 73–92. https://doi.org/10.1016/j.csl.2015.06.002

Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017). Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America*, *142*(2), 493–503. https://doi.org/10.1121/1.4995998

Terband, H., & Maassen, B. (2010). Speech motor development in childhood apraxia of speech: generating testable hypotheses by neurocomputational modeling. *Folia Phoniatrica Et Logopaedica*, *62*(3), 134–142. https://doi.org/10.1159/000287212

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952–981. https://doi.org/10.1080/01690960903498424

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America*, *88*(1), 97–100. https://doi.org/10.1121/1.399849

Tremblay, P., Deschamps, I., Bédard, P., Tessier, M. H., Carrier, M., & Thibeault, M. (2018). Aging of speech production, from articulatory accuracy to motor timing. *Psychology and Aging*, *33*(7), 1022–1034. https://doi.org/10.1037/pag0000306

Tremblay, P., Deschamps, I., Dick, A. S. (2019). Neuromotor organization of speech production. In de Zubicaray G. I. and Schiller N. O. (Eds.), *The Oxford handbook of neurolinguistics* (pp. 370–401). https://doi.org/10.1093/oxfordhb/9780190672027.013.15

Tremblay, P., Sato, M., & Deschamps, I. (2017). Age differences in the motor control of speech: An fMRI study of healthy aging. *Human Brain Mapping*, *38*(5), 2751–2771. https://doi.org/10.1002/hbm.23558

Tsao, Y.-C., & Weismer, G. (1997). Interspeaker variation in habitual speaking rate: Evidence for a neuromuscular component. *Journal of Speech Language and Hearing Research*, *40*(4), 858–866. https://doi.org/10.1044/1092-4388(2006/083)

Tuomainen, O., & Hazan, V. (2016). Articulation rate in adverse listening conditions in younger and older adults. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH 2016*, 2105–2109. https://doi.org/10.21437/Interspeech.2016-843

Tuomainen O., Taschenberger L., Rosen S., & Hazan V. (2021) Speech modifications in interactive speech: effects of age, sex and noise type. *Philosophical Transactions of the Royal Society B, 377(1841)*, 20200398. http://doi.org/10.1098/rstb.2020.0398

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154. https://doi.org/10.1016/0749-596X(89)90040-5

Turnbull, R. (2019). Listener-oriented phonetic reduction and theory of mind. *Language, Cognition and Neuroscience*, *34*(6), 747–768. https://doi.org/10.1080/23273798.2019.1579349

Uchanski, R. M. (2008). Clear speech. In D.B. Pisoni and R.E. Remez (Eds.), *The handbook of speech perception* (pp. 207–235). Wiley. https://doi.org/10.1002/9780470757024.ch9

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., & Durlach, N. I. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech, Language, and Hearing Research*, *39*(3), 494–509. https://doi.org/10.1044/jshr.3903.494

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence*, *33*(1), 67–81. https://doi.org/10.1016/j.intell.2004.08.003

Valentini-Botinhao, C., Maia, R., Yamagishi, J., King, S., & Zen, H. (2012). Cepstral analysis based on the glimpse proportion measure for improving the intelligibility of HMM-based synthetic speech in noise. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2*, 3997–4000. https://doi.org/10.1109/ICASSP.2012.6288794

Van Brenk, F., & Lowit, A. (2012). The relationship between acoustic indices of speech motor control variability and other measures of speech performance in dysarthria. *Journal of Medical Speech Language Pathology*, *20*(4), 24–29.

Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America*, *84*(3), 917–928. https://doi.org/10.1121/1.396660

Villegas, J., Perkins, J., & Wilson, I. (2021). Effects of task and language nativeness on the Lombard effect and on its onset and offset timing. *The Journal of the Acoustical Society of America*, *149*(3), 1855–1865. https://doi.org/10.1121/10.0003772

Walsh, B., & Smith, A. (2002). Articulatory movements in adolescents. *Journal of Speech, Language, and Hearing Research*, *45*(6), 1119–1133. https://doi.org/10.1044/1092-4388(2002/090)

Wang, Y. T., Kent, R. D., Duffy, J. R., Thomas, J. E., & Weismer, G. (2004). Alternating motion rate as an index of speech motor disorder in traumatic brain injury. *Clinical Linguistics and Phonetics*, *18*(1), 57–84. https://doi.org/10.1080/02699200310001596160

Wilshire, C. E. (1999). The "tongue twister" paradigm as a technique for studying phonological encoding. *Language and Speech*, *42*(1), 57–82. https://doi.org/10.1177/00238309990420010301

Yang, C.-C., Chung, Y.-M., Chi, L.-Y., Chen, H.-H., & Wang, Y.-T. (2011). Analysis of verbal diadochokinesis in normal speech using the diadochokinetic rate analysis program. *Journal of Dental Sciences*, *6*(4), 221–226. https://doi.org/10.1016/J.JDS.2011.09.007

Yaruss, J. S., & Logan, K. J. (2002). Evaluating rate, accuracy, and fluency of young children's diadochokinetic productions: a preliminary investigation. *Journal of Fluency Disorders*, *27*(1), 65–86. https://doi.org/10.1007/s10732-014-9237-2

Ziegler, W. (2002). Task-related factors in oral motor control: Speech and oral diadochokinesis in dysarthria and apraxia of speech. *Brain and Language*, *80*(3), 556–575. https://doi.org/10.1006

Zollinger, S. A., & Brumm, H. (2011). The Lombard effect. *Current Biology*, *21*(16), R614–R615. https://doi.org/10.1016/j.cub.2011.06.003

Appendices

Appendix A – RaMax Corpus

Overview

The Radboud Maximum Speech Performance Corpus (RaMax) contains speech data from 78 native Dutch speakers producing speech Stimuli in two maximum performance speech tasks, i.e., a DDK task and a tongue-twister task. The two maximum performance speech tasks were administered as (part of) the experiment in **Chapters 2** and **3** and in Shen and Janse (2019) and Shen and Janse (2020).

Availability

The RaMax Corpus is published together with this thesis and is available for research purposes upon request via Zenodo (a general-purpose open-access repository). The corpus is licensed under the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).

Data collection

Speakers

A total number of 78 participants (age: *M* = 23 years, *SD* = 3; 61 women) were recruited online through the Radboud Research Participation System. Participants' demographic information including gender and age, together with their DDK performance are provided in a separate text file named as 'speaker information'. Participants were all native Dutch speakers with normal or corrected-to-normal vision and had no reported history of speech, hearing, or reading disabilities nor past diagnosis of speech pathology or brain injury.

Speaker demographic information is documented in the file 'Speaker_information.txt' and can be accessed via DOI: https://doi.org/10.5281/zenodo.5645385.

Stimuli

The participants were instructed to produce the following Stimuli at their maximum speed as accurately as possible. The stimuli in the two tasks were presented on a computer screen using PowerPoint slides to the participants.

The stimuli for the DDK task include three monosyllabic ('pa', 'ta', 'ka'), two disyllabic ('pata', 'taka'), and four trisyllabic ('pataka', 'katapa', 'kapotte', 'pakketten') items respectively. Two out of the four trisyllabic items are real Dutch words (i.e., 'kapotte' – broken and 'pakketten' – packages) and two were non-words. These stimuli were presented to all participants in a fixed order (i.e., non-words: 'pataka', 'katapa' followed by real words: 'kapotte', 'pakketten').

The stimuli for the tongue-twister task contain four Dutch sentences (with English translation in parentheses) presented in a fixed order as shown below:

1. De poes kotst in de postzak (The cat puked in the mail bag)
2. Frits vindt visfrietjes vreselijk vies (Frits finds fish-fries terribly gross)
3. Ik bak een plak bakbloedworst (I fry a slice of blood-sausage)
4. Papa pakt de blauwe platte bakpan (Daddy grabs the blue flat frying pan)

Prior to the above-listed tongue-twister simuli, two additional tongue-twister sentences were presented as practice stimuli:

- Slimme Sjaantje sloeg de slome slager (Smart Sjaantje hit the slow butcher)
- Bakker Bas bakt de bolle broodjes bruin (Baker Bas bakes the round buns brown)

Audio recordings

One audio recording was made per participant using a Sennheiser ME 64 cardioid capsule microphone (10 - 20,000 Hz) on an adjustable table stand. The speech was recorded through a preamplifier (Audi Ton) onto a steady-state 2 wave/mp3 recorder (Roland R-05) with a sampling rate of 44.1 kHz. The recordings were made in a sound-attenuating recording booth at Radboud University Centre for Languages Studies.

Speech data can be accessed via DOI: https://doi.org/10.5281/zenodo.5651099.

Audio processing, file names, and TextGrids

The long audio recording per speaker was segmented into task-length recordings (one audio file for the DDK and one audio file for the tongue twister task) using Praat version 5.3.78 (Boersma & Weenink, 2017) on a Windows 10 Enterprise 64-bit operating system. The audio files are provided as mono channel, 16-bit, 44.1 kHz, uncompressed WAV files. The file name follows the template 'SpeakerNumber_TaskName.wav'. Praat TextGrid files are provided together with the DDK and tongue twister task audio files, indicating which utterance is produced when during the accompanying audio file (TextGrid file names following the template 'SpeakerNumber_TaskName.TextGrid').

TextGrid files can be accessed via DOI: https://doi.org/10.5281/zenodo.5651099.

Appendix B – RaLoCo Corpus

Overview

Radboud Lombard Corpus (RaLoCo) contains speech data from 78 native Dutch speakers' Dutch sentence-reading material (the same 78 speakers as in the RaMax corpus). Speakers read out sentences in two conditions: a habitual condition, in which they were instructed to read out 48 sentences fluently; and a clear-Lombard condition where instructed to read out the same 48 sentences as clearly as possible while they were hearing loud speech-shaped noise (at 78 dB SPL) via headphones. Speakers were presented with 48 unique sentences in one of eight random orders for the production of the habitual speech condition, followed by the same 48 sentences in a different random order for the production of the clear-Lombard speech condition. During the reading task, the habitual style always preceded the Lombard style for all participants to avoid potential spill-over effects from Lombard to habitual speech production. The four long lists (differing only in order) were rotated over participants such that trial effects could be isolated from sentence (i.e., item) effects. Participants' sentence production was live monitored by the experimenter. Once an error or disfluency was detected, participants were asked to re-produce the sentence again.

Additionally, two types of rating data were also included. Specifically, intelligibility ratings from a computer metric (the high-energy glimpse proportion, or HEGP) were included for the entirety of the speech dataset (i.e., 7488 ratings in total with one rating per utterance). Moreover, listening effort ratings from 231 human subjects were available for a subset (i.e., for 48 out of the total of 78 speakers, or a total of 4608 unique sentences: 48 speakers x 48 sentences x 2 speech styles) of the speech dataset.

Availability

The RaLoCo Corpus is published together with this thesis and is available for research purposes upon request via Zenodo (a general-purpose open-access repository). The corpus is licensed under the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).

Speech data

Participants

For the speech data, a total number of 78 participants (age: M = 23 years, SD = 3; 61 women) while for the human rating data, a total number of 231 participants were recruited online through the Radboud Research Participation System. Participants were all native Dutch speakers with normal or corrected-to-normal vision and had no reported history of speech, hearing, no reading disabilities nor past diagnosis of speech pathology or brain injury.

Speaker demographic information is documented in the file 'Speaker_information.txt' and can be accessed via DOI: https://doi.org/10.5281/zenodo.5645385.

Stimuli

The stimuli for the speech task were 48 syntactically-similar sentences. They were presented to the speakers via PowerPoint slides on a computer screen to the participants. A total number of four different sentence-lists with randomised sentence-order were rotated among the participants.

24 out of the 48 sentences contain a keyword in direct object position. A keyword has cV(c)c structure with 'c' being obstruents and 'V' being one of the three corner vowels in Dutch phonology (/i:/, /a:/, /u:/).

An example sentence is: 'Martin had de taak vandaag binnen een uur afgemaakt'. The literal English translation of this sentence is 'Martin had the task today within an hour finished'. A list of sentence materials (in Dutch orthography) is also provided in the corpus to allow automatic text-to-speech alignment of the audio files.

Sentence orthographic transcriptions and IDs are documented in the file 'Sentence_ID.txt' and can be accessed via DOI: https://doi.org/10.5281/zenodo.5645385.

Audio recordings

The recordings were made in a sound-attenuating booth at Radboud University Centre for Languages Studies. Speech was recorded using a Sennheiser ME 64 cardioid capsule microphone (10 - 20,000 Hz) on an adjustable table stand through a preamplifier (Audi Ton) onto a steady-state recorder 2 wave/mp3 recorder Roland R-05.

Speakers were instructed to read out the sentences accurately for the habitual condition (without wearing headphones), and to read out the sentences as clearly as possible while hearing speech-shaped noise (at 82 dB SPL) played through a pair of Sennheiser HD 215-II closed headphones (12 - 22,000 Hz) for the clear-Lombard condition.

Speech data can be accessed via DOI: https://doi.org/10.5281/zenodo.4040685.

Audio processing and file name

The recordings in the speech database are segmented sentence-length, mono channel, 16-bit, 44.1 kHz, uncompressed WAV files (i.e., 7488 files in total). All processing was made using Praat version 5.3.78 (Boersma & Weenink, 2017) on a Windows 10 Enterprise

64-bit operating system. The file name follows the template 'SpeakerNumber_Sentence-Number_RecordingCondition.wav', e.g., '1_K1_Nat.wav'. The letter 'K' indicates key sentence and 'Nat' represents habitual reading style.

Rating data

HEGP ratings

The total of 7488 sentences were rated in terms of intelligibility using the high-energy glimpse proportion metric, or HEGP (Tang & Cooke, 2016). The same speech-shaped noise used for clear-Lombard speech elicitation was used as the added noise-masker for the HEGP calculation at six different signal-to-noise ratio (SNR) levels (i.e., -10 dB, -5 dB, 0 dB, 5 dB, 10 dB, 15 dB, and 20 dB). HEGP scores were calculated by first computing all raw 'glimpses' defined as time-frequency regions where the energy of the target speech exceeds that of the masker, then selecting the subset of 'high-energy' glimpses, defined as those whose energy exceeds the mean speech-plus-masker energy, measured independently for each frequency region. The HEGP ratings lie between 0 and 1, with higher numbers indicating a higher glimpse proportion escaping energetic masking, thus suggesting higher predicted intelligibility. The HEGP ratings are documented in the file 'HEGP_ratings.txt' and can be accessed via DOI: https://doi.org/10.5281/zenodo.5645385.

Human ratings

The selected 4608 sentences were rated in terms of listening effort by a total of 231 human listeners. The listeners completed an online listening experiment to rate the listening effort they had to spend in order to understand the target sentences. Each listener rated a total of 48 unique sentences (such that no sentence was repeated twice), and within the 48 sentences, 24 were produced in the habitual and 24 were produced in the clear-Lombard speaking style. The sentences were mixed with the same speech-shaped noise used to elicit the clear-Lombard speech at -6 dB SNR (this SNR was chosen because the differences in intelligibility between habitual and Lombard speech were largest around this SNR level). Every set of 48 sentences was composed of sentences produced by six different speakers, so that each listener rated four unique habitual and four unique clear-Lombard sentences from one speaker. The sentences were presented in a semi-random order, and raters could only listen to a sentence once prior to providing their rating.

The human ratings ranged between 1 and 7, with 1 indicating 'no effort at all to understand' and 7 indicating 'extremely effortful to understand'. The listeners were explicitly instructed to also consider whether they could understand part of the sentence and to only provide a rating of 7 if they had hardly understood anything, and to choose a lower rating if they had understood at least some words. The raw

human listening effort ratings are documented in the file 'Lombard_Corpus_Assesments.csv' and can be accessed via DOI: https://doi.org/10.5281/zenodo.5645385.

Appendix C – Description of Research Data Management

Collection process

Participants were asked to produce speech (e.g., by reading sentences). Fragments of this speech were presented to a speech intelligibility metric or listeners in standard experiments to evaluate the speech. In addition, participants were involved in simple experiments that screened their hearing, cognitive capacity (such as working memory), and/or speech motor control (by having to pronounce words or nonsense words really fast). Examples of these simple experiments are an operation span test (test of working memory) and a maximum performance speech task. An operation span test requires participants to store and regularly update memory representations while performing another cognitively demanding task. A maximum speech performance speech task requires participants to produce syllable sequences quickly and accurately on repeat.

Participants were asked, via questionnaires, about their age (not their date of birth), language background, their use of (potentially) multiple languages, and whether they have a history of hearing or speech problems. The planned research was captured under several standard protocols described by the Ethics committee for the Humanities. Each test session lasted maximally 90 minutes and contained several breaks. Response data was acquired in the experimental labs of the faculty and stored on the data server facilities (i.e., a work-group folder) provided by the faculty's Humanities Lab.

Informed consent

The informed consent form and information documents were approved by the ethics committee for the Humanities of Radboud University.

The information document describes what the participant is expected to do during the study. The document also states that there is no anticipated risk or discomfort for the participant, that there is no chance of coincidental findings, and what the payment is for taking part in this study (the regular fee of 10 euros per hour in the form of a gift voucher). The information document also stresses the confidentiality of the collected data, that the participants' data is only used within the research group, that their data will be anonymised, and that they can indicate on the consent form what they consent to (regarding access to any audio recordings, e.g., whether they agree to the researchers playing their audio files at a conference for illustration purposes). The information document also states that their participation is voluntary and that they can withdraw from the study, or have their data withdrawn, until 24 hours after having participated. The information document also lists the name and contact

details of the principal investigator, should they have any questions, and the name and contact details of the secretary of the ethics committee (in case of any complaints). The informed consent form is to be signed both by the participant and the researcher, and states that participation is voluntary, that the participant knows and understands how the data will be processed and stored, and that the researcher has informed the participant correctly. Whenever applicable (for instance, in case participants were told they were talking to someone with a hearing impairment and this was not actually true), there is a debriefing document explaining that this was make-believe and why this was necessary for the purpose of the study.

Privacy

Data collection involved collection of critical data. Critical data are names on the informed consent forms and audio recordings from which participants may be recognisable from their voice or from what they say. In addition, questionnaires are critical data as they might be retraceable to specific individuals in case certain participants have a very unique language background in combination with a certain age, and a hearing or speech impairment. Informed consent forms were stored separately from the collected data. Prior to data collection, a key was developed (only available to the researcher) to generate participant codes to link these informed consent forms to data from specific participants. These anonymous participant codes were used (rather than participant names) on the study questionnaires, in participant lists or logbooks on which participant is assigned to which experimental study and/or experimental condition, and in the experimental programmes (as participant identifier). Participant names were collected (and stored) as part of the informed consent procedure. As participants were paid with gift vouchers, we did not need to collect participant addresses, phone numbers, or bank account numbers. Whenever paper sheets containing signatures and names were collected as proof of participants' voucher receipt, these sheets were kept by the Radboud Humanities lab manager (in compliance with the university's safety regulations for sensitive and critical data). Audio recordings were collected for the purpose of the study, as the overall aim of the research project is to investigate (by way of acoustic and perceptual analysis of elicited speech) how individual speakers clarify their speech. Participants indicated on their informed consent form whether they agree to their speech recordings being used for listening studies (in which we present speech fragments of their audio recordings to listeners), and for illustration purposes at academic meetings.

The files holding these personal data (scans of the informed consent forms, the key linking participant names to participant codes, questionnaires, and the pseudo-anonymous audio recordings) were registered in compliance with the university's safety regulations for sensitive and critical data (and are kept for a minimum of 10 years).

Because the informed consent forms contain the participant's possible consent to data sharing, the informed consent forms (and the key linking participant names to participant codes) are stored as long as the audio recordings are stored for possible re-use.

All data belonging to the same participant were marked with an anonymous participant code. At the start of each study, the researcher generated a key consisting of a sequence of a letter and several digits. This key was retraceable to the individual in question (in order to comply with the Dutch law on protection of personal data: Wet Bescherming Persoonsdata).

The questionnaire itself did not ask people any critical information (that is, information that could lead to their identification). As an example, we asked individuals for their age, as we think it may be important, but not their date of birth (the latter being critical information, but not the former).

Long-term storage

All relevant research data were stored in the form of corpora on Zenodo (a general-purpose open-access repository). For more details of the corpora published together with this thesis, see Appendix A and Appendix B.

Please indicate whether you will store your data for the long term. If not, explain why.		
Type of data	**Long-term storage?**	**If no, why?**
All data	Yes, to comply with Radboud University's research data management policy	
All anonymised data and/or pseudo-anonymised audio recording data for which participants have granted access to external researchers for research purposes (plus the accompanying informed consent forms and key linking anonymous participant code to participant name)	Yes, for possible re-use of data	

Please indicate where you will store your data long term and what the minimum and maximum retention period will be.		
Type of data	**Repository**	**Retention period**
Data acquisition collection	Central University server operated and maintained by ICS	Minimum period for all data types: 10 years Maximum period (only applies to anonymised/pseudo-anonymised data for which participants have granted access to external researchers for re-use, and the accompanying informed consent forms and the key linking the participant names to participant codes): as long as this research data may be relevant to other researchers
Research Documentation Collection	Central University server operated and maintained by ICS	Minimum retention period: 10 years
Data sharing collection	Central University server operated and maintained by ICS	Minimum retention period: 10 years Maximum period (only applies to anonymised/pseudo-anonymised data for which participants have granted access to external researchers for re-use and the accompanying informed consent forms and the key linking the participant names to participant codes): as long as this data may be relevant to other researchers

Giving access to data

This project is funded by the Horizon 2020 research programme of the European Commission. There are no real requirements on sharing of data, but the funder generally encourages sharing and re-use of research data. We allow interested consortium partners access to the anonymised data, and to the pseudo-anonymised audio recordings available for re-use (whenever permission is granted by the participants).

The various corpora can be accessed via:

https://doi.org/10.5281/zenodo.5651099.
https://doi.org/10.5281/zenodo.4040685.
https://doi.org/10.5281/zenodo.5645385.

Are there any privacy or security issues that concern the sharing of data after research? If so, please describe them and indicate how you will address them.
As participants may be recognisable/retraceable from their voices, despite the anonymised participant codes, their speech data can only be shared if participants themselves have agreed to sharing (this is a question on the informed consent form, where participants can indicate whom their speech recordings can be shared with). Anonymised data (such as experimental data files that only contain text) may be shared with external researchers (if access has been granted to them by the supervisors of this project).

Please indicate which access level you want to use, who controls the access to your data and if you are going to place an embargo period on the access of your data.

Type of data	Access level	Access control	Embargo
Raw	No access		
Processed critical/ sensitive data: informed consent forms and questionnaires	No access		
Processed critical/ sensitive data: all other (pseudo-anonymised or anonymised) data	Restricted access upon request	Supervisors (i.e., PIs Ernestus and Janse)	Not applicable
Analysed anonymised data	Restricted access upon request	Supervisors (i.e., PIs M. Ernestus E. and Janse)	Not applicable

Who is the target audience for your data?
Data collected in this research project is relevant for researchers working on spoken language communication.

Nederlandse Samenvatting

Van de ongeveer zevenenhalf miljard mensen op de wereld zijn er geen twee die op dezelfde manier spreken. Zelfs sprekers van dezelfde taal, leeftijd, geslacht en dialectvariant kunnen verschillen in de manier waarop ze spreken. Zo kunnen sprekers verschillen in hoe snel ze spreken, in hoe duidelijk ze spreken, of in hoe vaak ze struikelen over hun eigen woorden.

Sprekers kunnen hun spreekstijl ook aanpassen aan de situatie of de omgeving waarin ze communiceren. Zo moeten sprekers vaak 'harder praten' om te compenseren voor omgevingslawaai (bv. in rumoerige restaurants of cafés) of duidelijker articuleren om rekening te houden met eventuele moeilijkheden die hun gesprekspartners ondervinden (bv. als hun gesprekspartners slechthorend zijn of communiceren in een taal die niet hun moedertaal is) om ervoor te zorgen dat hun boodschap goed wordt begrepen. Niet alle sprekers zijn echter even goed in staat om hun spraak te 'verrijken' voor luisteraars.

Dit proefschrift gaat in op de vraag hoe sprekers hun spraakproductie 'controleren', zodat ze hun spraak kunnen aanpassen om aan verschillende (communicatieve) vereisten te voldoen. Meer specifiek richt dit proefschrift zich voornamelijk op drie verschillende aspecten of bronnen van variabiliteit in spraakproductie tussen sprekers: 1) verschillen in de maximale spreeksnelheid en nauwkeurigheid die sprekers kunnen bereiken; 2) verschillen in spraakvoorbereidingsprocessen die geassocieerd kunnen worden met ouder wordende volwassenen; en 3) verschillen in spraakverrijkingsgedrag bij het spreken tegen achtergrondlawaai. In de hoofdstukken 2 en 3 wordt het verband onderzocht tussen de cognitieve controle van sprekers en hun maximale spraakprestaties (die representatief worden geacht voor de hun spraakmotorische controle). Hoofdstuk 4 richt zich op leeftijdsverschillen in de tijd die sprekers nodig hebben om spraak voor te bereiden en te initiëren. Hoofdstuk 5 onderzoekt de manier waarop sprekers hun spraak verrijken vanuit twee invalshoeken: de consistentie waarmee zij hun spraakverrijking vasthouden; en het mogelijke verband tussen hun spraakverrijkingsgedrag en mate van hun spraakmotorische controle.

De resultaten uit dit proefschrift laten zien dat de (maximale) spraakprestaties van individuele sprekers samenhangen met cognitieve vaardigheden en spreektaakeisen. Ten tweede hebben de resultaten weinig bewijs geleverd voor leeftijdsgerelateerde achteruitgang in spraakplanning en -initiatieprocessen. Ten derde is aangetoond dat de mate van (verrijkte) spraakproductie van sprekers verre van statisch is over de tijd. De spraakcorpora die met dit proefschrift worden gepubliceerd openen mogelijkheden om meer vragen te beantwoorden over verschillen tussen sprekers in

uiteenlopende spreekcondities, en de effecten daarvan op luisteraars. Toekomstig onderzoek zou ook de associatie tussen individuele verschillen in cognitieve en/of spraakmotorische controle en andere aspecten van spraakproductie, bijvoorbeeld (de imitatie van) accent(en), kunnen onderzoeken.

English Summary

In the world of around seven-and-a-half-billion people, no two speakers are alike. Even speakers of the same language, age, gender, and dialectal variant may differ in the way they speak. For instance, speakers may differ in how fast they speak, in how clearly they speak, or in how often they stumble over their own speech.

Speakers can also vary their speech depending on the situation or environment they communicate in. For instance, speakers often need to 'speak up' to counter ambient noise (e.g., in noisy restaurants or pubs) or to speak more clearly to account for potential difficulties that their interlocutors may have (e.g., their interlocutors being hearing impaired or non-native listeners) to make sure their messages are properly understood. However, not all speakers are equally capable of enriching their speech for listeners.

This thesis addresses the question of how speakers 'control' their speech production, so that they can adapt their speech to meet various (communicative) needs. More specifically, this thesis mainly focuses on three different aspects or sources of between-speaker variability in speech production: 1) differences in the maximum speech rate and accuracy speakers can achieve; 2) differences in speech preparation processes that may be associated with increasing adult age; and 3) differences in speech enrichment behaviour when speaking in noise. Chapters 2 and 3 investigate the link between speakers' cognitive control and maximum speech performance (indexing speakers' speech motor control). Chapter 4 focuses on age differences in the time speakers need to prepare and initiate speech. Chapter 5 investigates speakers' speech enrichment from two angles: the consistency with which they maintain their speech enrichment; and the potential link between their speech enrichment behaviour and indices of their speech motor control.

The results from this doctoral thesis highlight that individual speakers' (maximum) speech performance is associated with cognitive abilities and speech task requirements. Secondly, the results have shown little evidence of age-related decline in speech planning and initiation processes. Thirdly, speakers' (enriched) speech production has been shown to be far from static over time. The speech corpora that are published with this thesis open up opportunities for answering more questions about between-speaker differences in diverse speaking conditions, and their effects on listeners. Future research could also explore the association between individual differences in cognitive and/or speech motor control and other aspects of speech production, for instance, (the imitation of) accent(s).

中文摘要

在这个大约有75亿人口的世界上，我们找不到两个完全相同的说话人。即使是年龄相仿、性别相同、以及使用相同语言和方言变体的说话人，他们的说话方式也可能不同。例如，说话人可能在说话的速度、说话的清晰度、或说话时犯口误的频率上有所不同。说话人还可以根据他们交流的情况或环境来改变他们的说话方式。说话人往往需要“大声说话”来对抗环境噪音（例如，在环境嘈杂的餐厅或酒吧），或者说得更慢也更清楚，考虑到对话人或听话人可能有的困难（例如，对话人有听力障碍或是非母语听众），以确保他们的话语信息被最大限度地正确理解。然而，并不是所有的说话人都可以同样有效地为对话人或听话人丰富他们的言语。

本博士论文讨论的问题主要是：说话人如何“控制”他们的言语产出，以便能够调整他们的言语产出来满足各种（交流）需求。更具体地说，本论文主要研究说话人个体言语产出差异来源的三个不同方面：一、在言语产出过程中说话人能达到的最快语速和最高准确度的差异；二、可能与成人年龄增长有关的言语计划和言语启动过程所需时间的差异；三、在噪音环境中说话人的言语产出及言语丰富性的个体差异。论文的第二章和第三章主要研究了说话人的认知控制和最优言语产出任务（索引说话人的言语运动控制）之间的联系。第四章主要研究了说话人在言语产出的准备和起始过程中所需时间上的年龄差异。第五章从两个角度调查了说话人言语丰富性的个体差异：说话人在单位时间内保持言语丰富性的能力；以及他们的言语丰富性与言语运动控制指数之间的潜在联系。

这篇博士论文的研究结果强调，说话人个体的（最优）言语表现与说话人的认知能力和言语任务要求有关。第二，几乎没有有效实验证据证明言语计划和言语启动过程的快慢与成人年龄增长相关的衰退有关。第三，实验结果证明了说话人言语产出的丰富性在单位时间内并非一成不变的特性。与本论文一起发表的语料库为回答关于不同说话条件下说话人之间的差异及其对对话人或听话人的影响的问题提供了机会。未来的研究还可以探索认知和/或言语运动控制的个体差异与言语产出的其他方面之间的联系，例如，说话人模仿或使用各种口音的能力。

Acknowledgements

Having imagined writing the acknowledgements of my thesis in many different (wildish) settings over the years: sitting in front of an antique wooden table in a cabin next to a tranquil lake or lying on a curvy beach bench next to the turquois Mediterranean Sea. Here I am, finally, just in my pandemic-induced home office, on my eight-year-old laptop, writing these acknowledgements.

First and foremost, I would like to express my sincere gratitude to my supervisors: Esther Janse and Mirjam Ernestus. Esther, thank you for picking me as the candidate for your project in ENRICH. I still remember our first conversation during which we commented on each other's English accent at the bar in Gilwell Park. Back then, I still had traces of a 'northerner'/'Sheffielder' accent, but over the years, I would almost add a 'hè' at the end of my English sentence if my attention lapses. Thank you, Esther, for being such a supportive supervisor in all possible ways, and for knowing just what is needed at the right moment to keep me going until the very end. I could not have asked for a better PhD supervisor. Mirjam, thank you for all your valuable and strategic advice on my research project in general as well as your timely and insightful comments on all my written drafts. Your expertise in the field and your guidance during the past years are very much appreciated.

Thank you to my funding agency: ENRICH European Training Network (ETN) for setting up such an incredible research network that encourages interdisciplinarity, collaboration, and mobility. I feel so incredibly lucky to be able to carry out my PhD research within ENRICH ETN and am very grateful to many people in the ENRICH team. I would like to thank the senior investigators: Anita, Clara, Deniz, George, Inma, Jan, María Luisa, Paul, Patti, Volker, and Simon for all your valuable input and suggestions. Thank you, Mónica, for always finding us the best deals during our various network meetings and conferences. Thanks also to my fellow Enrichies from around the world: Anna, Amy, Avashna, Carol, Dip, Elif, Gerard, Jewely, Max, Olina, Sneha, and Shifas, for creating some of the most fun but also inspirational memories in London, in Edinburgh and the Scottish Highlands, in Crete, in Vitoria, in Nijmegen, in Den Haag, in Oldenburg, in Aachen, and in Toulouse! Special thanks to Martin and Valerie for being my supervisors within the network and for being amazing hosts for my secondment at your respective institute. Martin, thank you for giving me the opportunity to collaborate with you on a paper and for being ultra-fast at responding to my emails with all kinds of questions and requests!

Thank you to the manuscript committee: Amalia, Ben, Hans Rutger, Hugo, and Marina, for taking the time to read and evaluate my thesis, for finding improvements, and for approving it! I look forward to seeing (some of) you in person in Nijmegen in June!

Thank you to my two lovely paranymphs: Katherine and Lotte. Katherine, we have shared so much throughout this PhD journey: from friendship, to research, to presentation, to accommodation! We are so similar yet so different, I guess that's what makes our conversations always so interesting and thought-provoking (at least for me). Lotte, I really like how enthusiastic and experienced you are about so many different things, despite being so young! We can chat about (proper loose leafy) teas, travelling, research, business ideas, things to watch, and fun places and activities to do non-stop! Thank you both for being my paranymphs and for sharing this special time with me.

Thank you to my colleagues and friends at the Erasmusgebouw and the MPI: Aurélia, Aurora, Chantal, Claire, Emily, Elly, Federica, Ferdy, Figen, Gert-Jan, Hannah, Julia, Kätja, Marlou, Merel, Mónica, Saskia, Tashi, Theresa, Thijs, Xiaoru, Wei, Xing, and Yu, among many others. I really enjoyed those spontaneous chats and cakes on the 8th floor of the Erasmusgebouw, those light-hearted lunch gatherings at the Refter, and those more organised pizza evenings at the MPI – the good old pre-pandemic times! Annika and Lisa: you were the best office mates that made 9.18a the best office I've ever worked in! Dear (former) members of the Speech Production and Comprehension research group: Hanno, Lou, Louis, Martijn, Mirjam B., Robert, Sophie, and Tim, thank you for all your valuable comments and questions that helped refining my research project.

Special thanks to Margret and Bob at the CLS lab and my (pilot) participants for helping me with my data collection; to Louis, Susanne, Laurel, and Phillip for stats-related support and courses; to Joe for all your tips on RMarkdown, Illustrator, aesthetics in academic work, and advice on job hunting and application; and to my most tech-savvy friend, Jinbiao, for helping me with all kinds of coding-related problems, pulling me out of despair whenever I fail to debug one (or more) nasty line(s) of code.

I am also very grateful for being a member of not one but two graduate schools during my PhD at Radboud: Graduate School for the Humanities and the International Max Planck Research School for Language Sciences. Thank you to the coordinators: Peter, Nicolet, and Kevin for organising so many activities to truly enrich my PhD life. Ranging from useful and practical workshops/courses and informative talks/presentations to fun social gatherings!

My time at Nijmegen has also been made fun with my friends outside the 'language circle'. Thank you, Kristina, for all the internationally-oriented delicious food and chats and for bringing me into an 'unknown' world of historians. Thank you, Maarten, for being my 'Dutch buddy' in Nijmegen and for sharing your, often quite wise, life hacks with me. Thanks also to the members of the Campus Choir for the often cheerful rehearsals (special thanks to Celine for looking after my two naughty fluff balls and for our Japanese-language exchange and casual-walking sessions). I would also like to thank my Chinese friends in Nijmegen for sharing food, experiences, memories, and tips and tricks for living in the Netherlands as a Chinese expat: Cong Z., Chao G., Huan W., Hongling, Lei W., Muqing, Xiaoyan, Xinyu and Bob (you're half Chinese in my eye because you can make Lanzhou beef noodles like no other Dutch can ;)).

谢谢我在中国的家人们: 多年来, 你们源源不断的关心和鼓励让我这个常年在外的他乡异客不孤单, 让我知道只要我回去, 有你们的地方就有家。谢谢我的妈妈在我看上去似乎毫无学习天赋和动力的那段时间依旧选择相信我和我所做的一切幼稚以及不那么幼稚的选择, 也谢谢你对我所有无条件的包容、支持、还有无私的爱。还有谢谢我在中国以及世界各地的小伙伴们: 娇娇, 安妮, 小天, 小树, 娜等等。你们让我觉得我五湖四海都有人, 都有家!

Mijn dank gaat ook uit naar mijn familie en vrienden in Nederland. René en Sylvia, Lisanne en Tom: bedankt dat jullie mijn familie in Nederland willen zijn, dat jullie mij accepteren zoals ik ben, en mij met open armen hebben verwelkomd. Dank aan de familie Verbeek voor het heerlijke Nederlandse en Franse eten en de leuke dagjes uit; en aan de familie Koenders voor het gezelschap en leuke gesprekken (met lekkere hutspot en Indonesisch eten). Door de fijne familiebijeenkomsten hebben jullie mij als een echte Nederlandse laten voelen!

Tom V. en Caro, Luuk en Sanne, Nick, Harold, Sjobbe en Jieun, Roderik R. en Laurée, Jori en Fleur: dank jullie voor de feestjes met hapjes en drankjes. Je volharding in het spreken van Nederlands met mij heeft me zeker geholpen bij mijn integratie in de Nederlandse samenleving en de Nederlandse taal ;).

Last but not least, Roderik, my best friend, teammate, travel companion, Jiujutsu uke, workout buddy, private chauffeur (though sometimes rather unwillingly), in-house star chef, and most importantly, love of my life: you are everything I could ever ask for in a life partner. Thank you for all the joyful moments and precious memories, for your tolerance and patience with me when I was out of any, and for your constant encouragement and support when I was slacking. I would have not done any of what I have done without you by my side. I look forward to the future with you and our xiao hu niu ^_^.

Curriculum Vitae

Chen Verbeek-Shen (Lanzhou, China, 1990) obtained her bachelor's degree in Sociology and Social Work from Hangzhou Normal University, China, in 2012. After working as an IELTS teacher at New Oriental Education & Technology Group in Hangzhou, China for one year, Chen moved to England to study Forensic Speech Science at University of York, receiving her MSc in September 2014. Her curiosity in linguistics and speech science motivated her to keep learning and researching aspects of language and linguistics for another 2.5 years at the University of Sheffield in England with funding from the Arts and Humanities Research Council. In 2017, Chen moved to Nijmegen, the Netherlands to work as an early stage researcher and PhD candidate in a European Training Network (ENRICH) funded by EU's H2020 research and innovation programme. Since March 2021, Chen has started working as a lecturer at Avans University of Applied Sciences in 's-Hertogenbosch, the Netherlands.

Publications

Shen, C., & Janse, E. (2020). Maximum speech performance and executive control in young adult speakers. *Journal of Speech, Language, and Hearing Research*, *63*(11), 3611–3627. https://doi.org/10.1044/2020_JSLHR-19-00257

Shen, C., & Janse, E. (2019). Articulatory control in speech production. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 2533–2537). Canberra, Australia: Australasian Speech Science; Technology Association Inc.

Shen C., Cooke M., Janse E. (2019) Individual articulatory control in speech enrichment. In *Proceedings of the 23rd International Congress on Acoustics* (pp. 5761–5765), Aachen, Germany.

Shen C., Xu Y. (2016) Prosodic focus with post-focus compression in Lan-Yin Mandarin. In *Proceedings of Speech Prosody 8* (pp. 340–344), Boston, United States.

Shen C., Watt D. (2015) Accent categorisation by lay listeners: which type of 'native ear' works better? *York Papers In Linguistics (YPL2)*, 14, 106–131.