Humboldt-Universität zu Berlin

Dissertation

The Influence of NATURE and NURTURE on Speaker-Specific Parameters in Twins' Speech: Acoustics, Articulation and Perception

zur Erlangung des akademischen Grades
Doctor philosophiae
(Dr. phil.)

Philosophische Fakultät II

Melanie Weirich

Dekan: Prof. Dr. Helga Schwalm

1. Gutachter: Prof. Dr. B. Pompino-Marschall
2. Gutachter: Prof. Dr. J. Harrington

Datum der Einreichung der Dissertation: 13. Juli 2011

Datum der Promotion: 11. November 2011

Zusammenfassung

Die Dissertation thematisiert sprecherspezifische Variabilität bei ein- und zweieiigen Zwillingen hinsichtlich Artikulation, Akustik und Perzeption. Die zentrale Fragestellung ist, ob sprecherspezifische Charakteristika auf *physiologisch-biologischen* Differenzen der Sprecher beruhen (BIOLOGIE), oder sich auf *gelernte, umweltabhängige* Unterschiede zurückführen lassen (UMWELT).

Artikulatorische und akustische Daten wurden von 4 eineiigen Zwillingspaaren (EZ, 100% genetische Übereinstimmung) und 3 zweieiigen Zwillingspaaren (ZZ, 50% genetische Übereinstimmung) analysiert. Zusätzlich wurde ein Perzeptionstest zur auditiven Ähnlichkeit der Zwillinge durchgeführt. Auf einen großen Einfluss des Faktors BIOLOGIE lässt sich schließen, wenn sich EZ ähnlicher sind als ZZ. Sind sich aber ZZ genauso ähnlich wie EZ, zeigt sich die Wichtigkeit der gleichen Lernumgebung (UMWELT).

Die Ergebnisse weisen auf einen großen Einfluss des Faktors UMWELT und stützen die Hypothese, dass sprachliche Ziele gelernt sind und sich am auditiven Feedback orientieren. Darüber hinaus wurden drei Faktoren gefunden, die den Einfluss der BIOLOGIE intensivieren: a) Lautklasse, b) Wortakzent und c) Koartikulation. Plosive und Sibilanten sind aufgrund des stärker ausgeprägten linguo-palatalen Kontaktes mehr durch die individuelle Physiologie beeinflusst als Vokale. Außerdem wurde ein größerer Effekt des Faktors BIOLOGIE in unbetonten als in betonten Silben gefunden. Zusätzlich stellten sich koartikulatorische Prozesse als wichtig heraus: dynamische Parameter – artikulatorische Gesten und akustische Transitionen – sind stärker durch die Physiologie beeinflusst als statische Parameter – artikulatorische Ziele und stabile akustische Regionen.

Sowohl der Faktor BIOLOGIE als auch der Faktor UMWELT sind einflussreiche Größen hinsichtlich sprecherspezifischer Variabilität. Welcher der beiden Faktoren die übergeordnete Rolle übernimmt, hängt von den spezifischen Charakteristika des untersuchten Parameters ab.

Schlagworte: Sprachproduktion, Sprachperzeption, sprecherspezifisch, Variabilität, Zwillinge, Umwelt, Biologie

Abstract

This dissertation examines inter-speaker variability in monozygotic (MZ) and dizygotic (DZ) twin pairs in regard to articulation, acoustics and perception. The aim of the study is to evaluate whether speaker-specific variability reflects *physiological differences* between speakers (NATURE) or bases on *learned variation due to social environmental influences* (NURTURE).

Articulatory and acoustic data was analyzed from 4 MZ twin pairs (100% identical genes) and 3 DZ twin pairs (50 % identical genes). Additionally, a perception experiment was carried out to explore the perceived auditory similarity. The effect of NATURE should have a larger impact than the effect of NURTURE, if a parameter differs more in DZ than in MZ twin pairs. If MZ and DZ twins show the same amount of inter-speaker variability, NURTURE seems to be crucial.

Results point to the importance of NURTURE and shared social environment. Nevertheless, three factors were found that intensify the effect of NATURE: a) phoneme class, b) lexical stress, and c) degree of coarticulation. Somatosensory feedback plays a larger role for consonants than for vowels, and thus individual physiology was found to shape articulation more in sibilants and stops than in vowels. Additionally, a stronger impact of NATURE was found in parameters that are auditorily less salient: unstressed syllables were more similar in MZ than in DZ twins, while for stressed syllables this was not the case. Moreover, coarticulation turned out to be essential: dynamic parameters – articulatory gestures and acoustic transitions – were more influenced by physiological constraints (NATURE) than static parameters – articulatory targets and stable acoustic regions.

Thus, both NATURE and NURTURE are crucial influencing factors in speaker-specific variability. However, the relative importance of the two factors is highly dependent on the specific characteristics of the investigated parameter.

Keywords: speech production, speech perception, speaker-specific, variability, twins, nature-nurture

ACKNOWLEDGMENTS

I am indebted to many people I met while working on this thesis. First of all, I would like to thank the participating twins. This project would never have been able to take place without their patience, kindness and understanding.

This work was carried out at the *Zentrum für Allgemeine Sprachwissenschaft* (ZAS) in Berlin, where all articulatory and acoustic recordings were made. It was supported by the "Bundesministerium für Bildung und Forschung" (BMBF, Grant Nr. 01UG0711). Many thanks to Jörg Dreyer for all technical support and his general assistance during the EMA-recordings.

I would like to thank Bernd Pompino-Marschall for the discussion of preliminary results and thoughtful comments on previous versions of this work, and Jonathan Harrington for agreeing to be the second thesis advisor.

I am greatly indebted to Susanne Fuchs and Stefanie Jannedy, who dedicated many hours to listening to my ideas and problems and who were a consistent source of guidance and support over the entire course of this project. Many thanks for all the time they spent with me, my thesis and my problems, for their insightful comments and their continuous positive feedback.

I would furthermore like to express my gratitude to a number of colleagues who offered their support. Jana Brunner was a great help with all kinds of MATLAB problems and provided copious scripting advice. Many thanks to Leonardo Lancia, who was a huge support in the preparation of the perception test and who offered his kind assistance in the statistical analyses. I would like to thank Ralf Winkler and Daniel Pape for providing helpful, time saving praat scripts that made my life much easier. I am also grateful to Martine Toda for thoughtful suggestions on the exploration of the sibilant data. Thanks to Marzena Zygis for helpful comments on the vowel chapter.

Thanks to the heart of PB1 – the student assistants Claudia Blankenstein, Micaela Mertins, Anna Theis, Vivien Hein and Anna Sapronova – without whom the work at the ZAS would not be the same and the pasta would not taste as good. Thanks to Matthias Ziervogel for his patient help in managing the references, to Susanne Schröder for her motivating coffee, and to Caterina Petrone and Stef for storming into my office and distracting me.

I am grateful to my mother and sister for always supporting me through their indestructible belief in me. Thank you, Micha, for being in my life and for being you.

Table of Contents

1 VARIABILITY IN SPEECH PRODUCTION

1.1 Possible influencing factors

Intra- and inter-speaker variability is one of the hallmarks of communication. Speaker-specific acoustic and articulatory variability are essential topics and crucial parameters in research on speech production, perception, and speaker recognition. One aim of this study is to explore just how intra- and inter-speaker variability is influenced by factors such as NATURE and NURTURE. Ladefoged & Broadbent (1957, p. 98) state that "the idiosyncratic features of a person's speech" may "be a part of an individual's learned speech behavior" or might be "due to anatomical and physiological considerations." Many fields of research explore the causes and consequences of the variability of certain human properties and discuss the impact of NATURE and NURTURE. However, what influence NATURE on the one hand and NURTURE on the other have on the acoustics and articulation in speech production and how these determinants interact in terms of intra- and inter-speaker variability is less clear. The factors NATURE and NURTURE can be described and specified as biological determinants (i.e. genetics, physiology, biomechanics) and non-biological determinants (i.e. social environment, learning, linguistic factors).

Lindblom (1984) encourages the "search for biological precursors of phonological and phonetic structure" (p. 75) and suggests applying models of evolutionary biology to phonetic problems. Biology and hence physiological and biomechanical factors play an important role in terms of motor control and articulatory targets in speech production. The question as to the nature of the representations of speech in the speaker's brain is a key topic in language and speech research. The search for possible correlates of the speech production task in the physical space has been addressed in several theories, such as the Motor Theory (Liberman et al. 1967, Liberman & Mattingly 1985), the Direct Realist Theory (Fowler 1986, 1991), the Acoustic Invariance Theory (Stevens & Blumstein 1978), the Auditory Enhancement Theory

(Diehl & Kluender 1989) or the Adaptive Variability Theory (Lindblom 1988, 1990) (for an overview see e.g. Hawkins 1999, Perrier 2005). One of the main questions discussed is whether the representations of speech are articulatory positions/targets or spectral properties/acoustic targets or even both. It has been shown that speech perception is multimodal, since the articulatory movements and the acoustic signal are taken into account when both modalities are available (e.g. McGurk & MacDonald 1976). Nevertheless, the question arises as to whether one modality is more essential than the other.

If articulatory targets are assumed in speech production, these targets might be influenced by the individual physiology of the speaker (e.g. his palatal shape and tongue size) even though the process of reaching the respective stored articulatory target is learned. In speech perception theories like the *Motor Theory* (Liberman et al. 1967, Liberman & Mattingly 1985) and the *Direct Realist Theory* (Fowler 1986, 1991) the objects of speech perception are regarded as being articulatory. The neuromotor system, the muscle commands, and the articulatory gestures are considered to be the representations in speech production and the objects of speech perception, since no invariance is seen in the acoustic domain. These articulatory gestures or targets are influenced by physical, biological and individual physiological restrictions and, hence, the factor NATURE.

In theories like the *Acoustic Invariance Theory* (Stevens & Blumstein 1978), the *Auditory Enhancement Theory* (Diehl & Kluender 1989, Diehl & Kingston 1991) and the *Adaptive Variability Theory* (Lindblom 1988, 1990), the object of speech perception is assumed to be acoustic or auditory. While the *Acoustic Invariance Theory* proposes that invariant properties can be found in the acoustic signal, the *Auditory Enhancement Theory* and the *Adaptive Variability Theory* assume acoustic variability. The *Auditory Enhancement Theory* emphasizes that the combination of different acoustic properties can result in the same perceptual properties which in turm form distincitve features that build the units of speech percpetion. The *Adaptive Variability Theory* rejects the existence of any physical invariance and places the focus on the listener: "The invariance of linguistic categories is ultimately to be defined only at the level of listener comprehension" (Lindblom 1988, p. 160). Here, perceptual representations of speech are seen as being crucial. Similarly, Ladefoged (1984) emphasizes that the purpose of language and speech is that speakers and listeners can communicate. He sees language as a self-organizing *social* institution and points out that the speaker has to produce "a sufficiently

distinct sequence of sounds for the listener to be able to get the message in a sufficiently short length of time" (p. 83). Thus, *articulatory effort* and *auditory distinctiveness* are crucial factors that form language.

In addition to language transmitting information, Ladefoged points out that speech also conveys sociolinguistic information and idiosyncratic characteristics of the speaker. These phonetic details cannot be ascribed to universal principles but mirror "local history and personal desire" (p. 85). Thus, the impact of individual language experiences and social-environmental influences become relevant. Moreover, the factors *learning* and NURTURE come to the fore when we speak of auditory goals. Several studies have investigated the link between perception and production and found auditory feedback but also auditory acuity (i.e. the sensitivity of the auditory apparatus) to be crucial in speech production and inter-speaker variability (Diehl & Kingston 1991, Newman et al. 2001, Jones & Munhall 2003, Perkell, Guenther et al. 2004, Perkell, Matthies et al. 2004, Ghosh et al. 2010).

For the following investigation it is important to emphasize that multimodal representations are assumed in the speech production process. Both articulatory targets and acoustic or auditory goals are seen as beeing crucial. However, differences in the relative importance of both representations might exist depending on the phoneme class (and in particular the production of the sounds and the amount of linguo-palatal contact). With respect to these two domains of representation (articulatory vs. acoustic/auditory) the two factors NATURE and NURTURE can have different degrees of influence on inter-speaker variability in speech production. Thus, 1) when do physiological constraints prevail over the impact of our social environment, and when does NATURE have an impact on the character of articulatory targets? And, in contrast, 2) in which cases are we free to choose different articulatory strategies, and when does NURTURE and hence social environment and learned auditory goals, seem to be a determining factor?

In the present chapter I will first briefly introduce the factors NATURE (i.e. physiology, biomechanics and vocal tract properties) and NURTURE (i.e. social environment and learning from observing), and then I will have a closer look at the particular implications of these factors concerning intra- and inter-speaker variability in speech. Parallel, the contrasting roles of somatosensory feedback and auditory feedback will be discussed. Several studies that

a) could shed some light on the influence of somatosensory and/or auditory feedback on speech production and b) could give some useful hints regarding the potential impact of NATURE and/or NURTURE on inter-speaker variability will be presented. Furthermore, the splitting of the speech signal into *targets* (static positions) and *transitions* (dynamic traces between targets) will be described and the possible varying influence of NATURE on these two characteristics of speech will be considered. Thus, in Chapter 1 the emphasis is placed on the potential impact of the two factors NATURE and NURTURE on speech, while in Chapter 2 the focus lies on the speech group under investigation – twins – and their possible contribution to the discussion of the abovementioned influencing factors.

1.1.1 NATURE

Organic variation concerns the anatomical structure and physiology of a speaker's vocal apparatus (e.g. the size of the larynx, the shape of the vocal tract, the size of the teeth and the bite, the stiffness of the tongue muscles, etc.) and its mechanical properties. The importance of genetic factors as an influence on vocal quality has been recognized in many studies. Physical characteristics (including the vocal apparatus) are genetically determined (Sataloff 1995, Flach et al. 1968). For instance, Fitch & Giedd (1999) examined morphometric data, including midsagittal vocal tract length, shape, and proportions, by means of magnetic resonance imaging in 129 normally developed speakers (aged between 2 and 25 years). Speaker-specific differences were found, and a significant positive correlation between vocal tract length and body size (height and weight) became evident. Moreover, differences between male and female vocal tract morphology were found in terms of overall length and relative proportions. Anatomical sources of speaker variation also include the vocal folds. For example, the length and mass of the vocal folds have an influence on the fundamental frequency (F0) – longer vocal folds lead to a lower F0. Furthermore, the physiological characteristics of the vocal folds affect the voice quality – for instance, genetically determined distortions in the adduction of the vocal folds can lead to a breathier voice. Extensive evidence exists for the relationship between increasing age and decreasing fundamental frequency (Decoster & Debruyne 2000, Linville & Rens 2001, Helfrich 1979). By contrast, the influences of age on formant patterns are less clear. Lindblom & Sundberg (1971) suggest that the lengthening of the vocal tract with increasing age (which is due to a lowering of the

respiratory system and the digestive tract: see Luchsinger & Arnold 1965) is responsible for a downwards shift of the formant frequency ranges with age. Other studies have found no such lowering (Meurer et al. 2004, Labov 1994). However, the vocal tract length influences the vowel formant frequency ranges, and the vocal tract geometry affects small phonetic details like gestural coordination patterns.

Beck (1999) classifies the sources of interpersonal organic variation into three types. First, life cycle changes (from birth to puberty, from puberty to maturity, from maturity to senescence) influence the vocal apparatus structure of each individual. Second, endogenous and environmental factors can influence the growth and final shape and size of the physical characteristics of a person. While endogenous factors fall under genetic control and determine the maximal growth potential, environmental factors including low socio-economic status (and possibly as a consequence poor nutrition) as well as emotional disturbance, large family size, being a younger or older sibling, etc. may inhibit the growth potential. And third, intra- and inter-speaker variation can arise from trauma or disease. None of the environmental sources of organic variation affect any of the speakers of the investigated group in this study: *twins*. Here, the influence of life cycle changes is the same for both speakers of the twin pair, both are the same age at the same time, and since they have grown up together in the same family environmental factors are common to both. That anatomical and physiological characteristics are more similar in monozygotic twins (MZ) than in dizygotic twins (DZ) has also been shown in several studies (e.g. Lundström 1948, Langer et al. 1999). Lundström's early medical dissertation on anatomical variation in twins showed that identical twin pairs reveal less variation in the size and position of the jaw and the teeth than non-identical twin pairs. Langer et al. (1999) found in his ultrasound study that the thyroid volume is almost the same in identical twins and is similar in non-identical twins.

The monozygotic and dizygotic twins that served as speakers for the present study had no trauma or diseases, thus it can be assumed that the organic variation is near to zero for the monozygotic twins and slight but existent for the dizygotic twins due to endogenous factors and possible differences in the genetic material. A closer inspection of this assumption (a more similar physiology for MZ than for DZ twins) for the participating subjects will be made in Chapter 3 (Sections 3.1.1 & 3.1.2).

1.1.2 NURTURE

Regarding the acquisition of language Chambers (2003) stated that "when children acquire their mother tongues, they evidently acquire the local variants and the norms of their usage too" (p. 174). Behavioral sources of talker variation in speech may be language specific and/or dialect specific. Both factors shape the phonemic inventory, the prosody and the phonetic implementation. Linker (1982), for example, investigated the lip positions of vowels of native speakers of Cantonese and French and found that the languages differ in the amount of lip protrusion they use to make the same acoustic distinction between vowels. Thus, children learn to produce the acoustic goal by watching and listening. Ladefoged (1984) points out that such language-specific behavior (such as the different articulatory strategies that can be used to achieve the acoustic output of /u/) is associated with group identity and a sense of belonging and cannot be given a physical explanation based on anatomical differences between the two speaker groups.

However, in addition, idiosyncratic variation can appear within a group of talkers speaking the same language and dialect (independently of anatomical differences). This learned variation or phonetic variation (as Laver 1980 calls it) deals with differences in the way in which an individual uses his or her vocal apparatus. The search for idiosyncratic features of a given speaker is a major theme within the field of forensic speaker identification. In the forensic literature, *acoustic variability* is a crucial research topic. For example, Rose (2002) discusses the question of whether every speaker has a unique voice, and if yes, whether it can be *measured*. Furthermore, the non-linear relation between articulation and acoustics but also between acoustics and perception (Stevens 1972) has to be kept in mind. Small differences in articulation can result in large differences in acoustics. Differences in acoustics do not necessarily lead to differences in perception (note also that the perception of a phoneme contrast is dependent on phonology, i.e. the phoneme inventory of the respective language a person speaks). Moreover, differences in articulation do not necessarily result in differences in the acoustic output. Individual *articulatory variability* can occur in terms of motor equivalence, since some speech sounds can be produced by different alternatives in articulation, demonstrating "the capacity of a motor system to achieve the same end product with considerable variation in the individual components that contribute to that output" (Hughes

& Abbs 1976, p. 199). For instance, the same acoustic output necessary for the vowel /u/, i.e. low second formant frequencies, can be achieved by rounding the lips, lowering the larynx or moving the tongue backwards (Perkell et al. 1993, Savariaux et al. 1995).

Learned variation is assumed to be a crucial factor in the study, since the subjects are twins and grew up together sharing their social environment and auditory goals during speech acquisition. This holds true for monozygotic and dizygotic twin pairs and is discussed in more detail in the twin chapter (cf. Chapter 2). An examination of this assumption (same amount of shared social environment for MZ and DZ twins) regarding the participating subjects in the current study is presented in Section 3.1.

1.1.3 Targets and transitions

A typical feature of speech is its continuous nature, in particular, the context-dependent variability in speech. The movements of different articulators interact with each other and overlap in time in the speech production process, or as Farnetani (1999) puts it "the vocal tract configuration at any point in time is influenced by more than one segment" (p. 371). This phenomenon is known under the term coarticulation. An interesting aspect of coarticulation is that it may or may not be audible. It can be observed and described in the articulatory domain, i.e. movements or articulatory gestures that overlap in time, or in the acoustic domain, i.e. spectral consequences. Kent (1983), Fowler (1980), Farnetani (1999) and Hardcastle & Hewlett (1999) for example, give further insights into theories and experimental data on coarticulation. The aspect that is important for the current study is that coarticulation can serve as a possible source of inter-speaker variation, showing the impact of the factors NATURE and NURTURE. As a consequence of the continuity of the speech signal, the speech production process can be described in terms of *targets* and *transitions*. Nolan et al. (2006) assume that the speech signal contains a) linguistically determined *targets*, which are constrained by the shared language system, and b) organically determined and speaker-specific individually learned *transitions*, which link the adjacent linguistic targets. In their overview of the origins of coarticulation, Kühnert & Nolan (1999, p. 25) state: "it may be that coarticulatory strategies are essentially idiosyncratic with each individual free to develop a personal solution to the integration of successive segments." Interestingly, Nolan & Oh

(1996) found coarticulation parameters (in the acoustics of /l/ and /r/) to be speaker-specific and idiosyncratic but could not find this speaker-specificity in the speech of identical twins. This fact points to a strong influence of NATURE and biology on coarticulation parameters and will be an issue discussed in this investigation (see e.g. Chapter 5 on sibilant targets vs. sibilant-vowel transitions). Nolan & Oh explain their results with the assumption that speakers share *segmental targets* but hide differences in their vocal tracts through different *transitions* (i.e. coarticulation parameters) between the targets. Rose (2002, p. 189) argues similarly and states that a variety "does not specify how you get from articulatory target A to target B" and that "individuals may be free to find their own articulatory solution."

1.2 The role of NATURE and somatosensory feedback in speech production

In this section, I will discuss and evaluate potentially influential issues related to the factor NATURE. First, the role of vocal tract properties will be treated: in which way does a speaker's physiology affect not only intra- but also inter-speaker variability? Here, the interaction of the tongue with the (individually shaped) palate – a natural confinement of the vocal tract – is examined. In addition, the role of somatosensory feedback and also somatosensory acuity will be illuminated by presenting some insightful results of studies with perturbation experiments.

1.2.1 The influence of vocal tract properties on intra- and inter-speaker variability

Speech has a biological foundation as speech production and perception are strongly constrained by an individual's physiology. The vibration of the vocal folds generated by the air stream coming from the lungs provides the source of the speech signal. The cross-sectional area of the vocal tract serves as a filter for the acoustic wave, and the particular shape of the vocal tract during articulation influences its filter characteristics. Furthermore, biological speaker-specific restrictions result in individual variations in the vocal tract form. Therefore, speakers differ in their formant values (Fant 1960).

In articulation the functioning of the tongue muscles influences the tongue movements during speech and hence restricts the possible resulting articulatory gestures. Regarding the economy of speech gestures, Lindblom (1983, p. 217) assumes that "languages tend to evolve sound patterns that can be seen as adaptations to biological constraints of speech production." Lindblom's *H&H Theory* (1990) explains the lack of invariance in speech production by the fact that speakers adapt their speech behavior to communicative and situational demands. They can choose their speech behavior along a continuum from hyperspeech (output oriented) to hypospeech (system oriented). The output-oriented nature of speech aims at achieving plasticity in favor of a successful communication process. The system- or production-oriented nature of speech, on the other hand, is based on a low-cost form of behavior and the principle of economy. Speech production is influenced by the demands of the output and the constraints of the system, thus a characteristic of speech is its adaptive behavior. However, first and foremost, this adaptive speech behavior and therefore, "speaking and listening are shaped by biologically general processes" (Lindblom 1990, p. 403). The speech signals we are able to produce and perceive are limited by our physiology (Fant 1960, Lindblom 1983, for an overview see e.g. Fuchs et al. 2007).

In addition, other theories like *Articulatory Phonology* (Browman & Goldstein 1986, 1989, 1990, 1992), where the basic phonological unit is the articulatory gesture, have pointed out that phonology and physiology cannot be seen in isolation. Here, phonological units are described in terms of dynamically specified units of articulation, i.e. gestures. Speech is then organized in a constellation of different gestures which may overlap in time. Since a speech gesture is characterized by a coordinative structure of articulators (and thus the muscles that move them), biology and physiology play an important role in this theory.

Stevens proposes in his *Quantal Theory of Speech Perception* (1972, 1989) that those speech sounds are preferred in the world's languages in which articulatory variability has only little impact on the acoustic output and thus on the perceived phonological category. A further assumption that could be drawn from this is that phonology is shaped by the interaction between properties of articulation and of acoustics.

A number of studies have taken into account the fact that speech production is based on physiological constraints and have investigated the interplay between biology and intra- and

inter-speaker variability in speech production. With respect to the role of NATURE in the speech production process, one important issue is the **influence of vocal tract properties** on the **phonetic realisation of various phonemes and differences in intra-speaker variability** between different sounds.

A connection between biology and differences in articulatory variability was found by Shiller et al. (2002). Physiological parameters constrain the capability of our speech motor system. In their study an asymmetric relationship between jaw stiffness (i.e. the resistance to displacement) and kinematic variability could be shown: higher stiffness values were observed for jaw protrusion and retraction, which goes hand in hand with reduced kinematic variability. In addition, for high jaw positions, stiffness was greater and kinematic variability smaller. The authors conclude that kinematic variability in speech is influenced by the restrictions of humans' biomechanical apparatus, and thus phonology and in particular differences in token-to-token variability are shaped by NATURE.

For vowels, Perkell & Nelson (1985) found in their study higher articulatory precision in the positioning of dorsal tongue points near the place of maximal constriction for /i/ and /a/ in a direction perpendicular to the vocal tract midline compared to the direction parallel to the midline. They interpret this as supporting evidence of physiological "saturation effects." Furthermore, studies by Mooshammer et al. (2004) and Brunner et al. (2005) have shown that there is less variability when the amount of linguo-palatal contact is large, suggesting high biomechanical restrictions in the production of high vowels. This issue will be discussed in more detail in Chapter 4 on vowels.

For consonants, Fuchs et al. (2006) also studied the role of the palate in speech motor control and investigated tongue tip kinematics and tongue-palate contacts to compare the production strategies of alveolar stops and fricatives. They studied the productions of 4 German speakers by means of electromagnetic midsagittal articulography (EMMA) and electropalatography (EPG). Results support the assumed two different production strategies for stops and fricatives. For reason of stability and simplicity the articulatory target for stops is planned beyond the actual contact location (i.e. a *virtual target*, see Löfqvist & Gracco 1997). For fricatives, a more precise tongue (and jaw) position is needed and the target lies at the lateral margins of the palate. The existence of these two strategies, the difference in control and fine-

tuning of the tongue muscles between the categories, and the significant role of the palate as a physiological restriction in speech production could be supported by their findings. Their results revealed differences in articulation parameters between the phoneme categories stops and fricatives, for example differences in deceleration peaks, in movement amplitude, or in velocity peaks and durations of the closing gesture.

Velar stops are described as being strongly influenced by anatomical and physiological properties; especially interesting here is the looping movement of the tongue that occurs when a velar stop is surrounded by vowels (Houde 1967, Hoole et al. 1998, Mooshammer et al. 1995, Perrier et al. 2003, Brunner et al. 2011). Among others, Perrier et al. (2003) investigated in their modeling study the looping trajectories of the tongue. They conducted simulations of V1CV2 sequences with C being a velar consonant by using a 2-D tongue model. Results show that the looping trajectory during the sequence is influenced by the quality of the vowels and also by the consonantal target.

However, an additional influential factor turned out to be the *speaker*. The authors explain **inter-speaker variability** in looping trajectories that were found elsewhere (Mooshammer et al. 1995) partly by speaker-specific differences in physiology: "while the general orientation of the loop is the same for each speaker, the amplitude of the sliding movement during the closure depends on speaker-specific properties, at a control and at a physical level" (Perrier et al. 2003, p. 1594). Thus, the authors assume that biomechanical factors – for example the way the tongue muscles produce the velar closure, and the interaction with the palate during the consonantal closure – are crucial in explaining the trajectories, and no general optimization principle that plans the entire trajectory is necessary to explain the looping movement as proposed by Löfqvist & Gracco (2002). This will be discussed in more detail in Chapter 6.

That anatomical properties, in particular the shape of the palate, may play a role in inter-speaker variability in vowels is also shown in a study by Brunner et al. (2009). Articulatory and acoustic variability was investigated in 32 speakers by means of EPG and acoustic recordings. Results showed less articulatory variability in tongue height in speakers with flat palates. The authors assume that speakers with a flat palate are more constrained in their articulatory variability, since small variation in the tongue position has a larger impact on the area function and henceforth on the acoustics than in speakers with a dome-shaped palate. Speakers with a

dome-shaped palate did not show a congruent pattern in articulatory variability, leaving the authors to conclude that they "have a greater range of possible levels of variation since the articulatory variability they can allow for without changing the acoustic output considerably is higher." (p. 3941)

Another study investigating inter-speaker variability in vowels deals with the possible relationship between speaker-specific vocal tract geometries and their articulatory vowel space (Winkler et al. 2006, Fuchs et al. 2008). The authors investigated the articulatory distances between the corner vowels in 9 French speakers by means of magnetic resonance imaging (MRI) and looked for a relationship to the speakers' pharynx length. Results indicate that speakers with a longer pharynx also produce larger displacements between low back and high front vowels.

Hence, it has been shown that vocal tract properties influence the production strategies of several sounds, thus NATURE seems to affect the phoneme inventory and the particular variability that is allowed for each phoneme. In addition, speaker-specific characteristics can at least partly be explained by differences related to vocal tract properties, such as the shape of the palate. One crucial issue regarding the influence of vocal tract properties on speech is the interaction of the tongue with the borders of the vocal tract such as the palate. This interaction and the role of somatosensory feedback in speech will be discussed in the following section.

1.2.2 The relevance of somatosensory feedback in speech production

The importance of the biological predispositions for speech and in particular the **relationship between tongue and palate** is shown in a paper by Stone (1995). The paper emphasizes the dependency of the tongue on the palate to shape it in a way that is necessary for the production of certain sounds (e.g. sibilants). Not only the creating of specific tongue shapes but also the accommodation of aerodynamic changes and coordination patterns with the jaw are only possible due to the interplay of tongue and palate. For example, for the production of consonants the tongue does not control all its movements only by using its muscles, but "…it uses the resistance afforded by the palate to fine-tune its shape" (Stone 1995, p. 147; see

also Stone & Lundberg 1994). Tactile feedback through tongue-palate contact is also crucial in coordinating movements with the jaw. Aerodynamic patterns and a manipulation of the airflow that is achieved by a certain tongue-palate contact are significant factors during the production of fricatives. Vowels show a location-to-shape relationship, thus tongue shape and position are strongly correlated and predictable: a higher tongue has a steeper slope (e.g. /i/) than a lower tongue (e.g. /a/) (cf. "saturation effect" above). For high vowels the palate might also play a role in providing an upper boundary and tactile feedback for fine motor tongue positioning. Thus, biological constraints shape the phonetic nature of speech, and **tactile (or somatosensory) feedback** and physiological predispositions are meaningful factors in the (individual) production of speech.

Evidence for the importance of somatosensory feedback in speech production comes from studies conducting **perturbation experiments**. Honda et al. (2002) and Honda & Murano (2003) investigated compensatory responses of articulators to unexpected perturbations of the palate shape. The thickness of an artificial palate was decreased and increased dynamically through inflation during speech with normal and masked auditory feedback. Acoustic, perceptual and articulatory (by means of EMA) measurements were conducted of the fricative /ʃ/ and the affricate /tʃ/ in CV syllables. Rapid compensation strategies of the tongue were observed around the second repetition of the syllable even when auditory feedback was masked. However, when auditory feedback was masked, some speech errors occurred in the following syllables. Thus, the authors assume that tactile feedback is used for rapid compensation (actively and immediately) and auditory feedback helps to complete the fine articulatory adjustment (with a longer time delay).

Another interesting and comprehensive perturbation study was conducted by Brunner (2009). In her dissertation project she discusses the nature of phonemic targets in speech production. What is the aim of a speaker: a certain acoustic output or a certain articulatory movement? To answer this question the vocal tract shapes of seven speakers were modified by palatal prostheses which were worn by the speakers over a period of two weeks. Several acoustic and articulatory recordings (by means of EMA) were made prior to, during and after the perturbation. In addition, in some cases auditory feedback was masked to investigate whether speakers compensate while only tactile feedback is available. Thus, the study tries to shed some light on the questions of whether speakers adapt to acoustic or articulatory targets and

what roles auditory feedback and somatosensory feedback play. As expected, the results point to the importance of both articulatory and acoustic targets. Articulatory representations are used when no auditory feedback is available. Especially vowels but also fricatives could be classified in a reasonable way. The author compares her study with the perturbation study of Jones & Munhall (2003), where only the length of the teeth was prolonged and no somatosensory feedback on the palate was perturbed. Here the speakers needed the auditory feedback to compensate their speech. Since the speakers in Brunner's study could feel the prosthesis, they could use tactile feedback to estimate a new articulatory position. However, the author emphasizes that the "articulatory representation was overrun by the acoustics" (p. 121) in several cases; for example, she found the use of motor equivalent strategies in the production of /u/ (where the articulation changes but the acoustic output stays constant).

Studies from the research group around Nasir & Ostry (2008, 2009) deal with the fact that speech production relies on sensory input; and the different roles of auditory and somatosensory feedback on speech motor learning are discussed. Their studies indicate the crucial influence of somatosensory feedback (rather than auditory) in **speech motor learning**. In a study with profoundly deaf adults with their cochlear implants (CI) turned off and normal hearing control subjects, Nasir & Ostry (2008) use a robotic device to alter somatosensory feedback. The load of the perturbation device was dependent on the velocity of the jaw movement. The perturbation displaces the jaw in a protrusion direction and speakers had to read aloud different utterances before, during and after the perturbation. The subjects were tested over a sequence of 300 utterances. Note that somatosensory but not auditory feedback was affected by the perturbation. Results showed that subjects learned during the perturbation to adapt to the altered somatosensory feedback. To estimate the adaptation the authors measured jaw trajectories and compared them between the groups and the different time points (before the perturbation, during training, at the end, and after the perturbation). Interestingly the implant users with their CIs turned off corrected their speech movements in the same way as the normal hearing controls did. Moreover, all of the implant users adapted, but only two-thirds of the control group did, pointing to the possibility that the profoundly deaf adults may even have a heightened sensitivity to somatosensory input as a consequence of their hearing loss. The authors assume from their results that auditory input is not necessary for speech learning but that it is dependent on somatosensory feedback. They

also support their results with the ability of postlingually deaf individuals to speak intelligibly in the absence of auditory feedback (at least for some time after the hearing loss) because they use stored motor programs.

In a follow-up study Nasir & Ostry (2009) found that speech motor learning can even affect a speaker's auditory map. They again conducted a perturbation experiment where the jaw movement path was displaced and somatosensory feedback was influenced. A mechanical load was applied to the jaw that resulted in jaw protrusion and the load varied depending on the velocity of the jaw. Subjects were asked to read several test items ('bad', 'mad', 'had' and 'sad') that involve large jaw movements and result in large perturbation loads, thus promoting adaptation. In addition, subjects participated in a perception test before and after the perturbation where they had to identify a stimulus drawn from an eight-step computer-generated continuum as either 'head' or 'had'. Results showed 1) that speakers corrected for the mechanical perturbation with practice, thus revealing motor learning, and 2) that speakers additionally showed changes in their perceptual categorization of phonemes after the experiment: stimuli were more often classified as 'head'. Thus, the authors assume a relation between auditory plasticity and speech motor learning. It should be noted that the possibility exists that the perceptual boundaries might also change due to sensory adaptation or auditory memory effects. However, since no perceptual shift was found in a control group with no perturbation and in subjects that did not adapt to the perturbation, the authors conclude that a link between motor learning and perceptual change exists.

These studies reveal the important role of somatosensory feedback in maintaining programmed speech movements when auditory feedback is perturbed.

Niemi et al. (2006) provide further evidence for the importance of somatosensory feedback. In their acoustic study they investigated the spectra of the sibilant /s/ when the somatosensory feedback was reduced by using local anesthesia of the right lingual nerve. Five participants produced /s/ in a variety of phonetic contexts, and the spectral characteristics (among others Center of Gravity and kurtosis) of the sibilants were analyzed. Results were twofold: 1) the reduced sensation affected the spectral output, indicating the relevance of somatosensory feedback in producing the sibilant, and 2) the output varied between different speakers, indicating speaker-specific compensatory mechanisms. It should be noted that the

results have to be interpreted carefully since it is not exactly clear what happens during the anesthesia. There might be differences in the degree of anesthesia and the spacial expansion between the different speakers that cannot be controlled for, and this in turn might influence the different acoustic outputs as well as the individual compensation strategies.

The study by Ghosh et al. (2010) also examines the role of somatosensory feedback in sibilant productions. They measured auditory acuity (by means of a discrimination task between /s/ and /ʃ/) and **somatosensory acuity** of their speakers. Somatosensory acuity was measured by pressing small plastic domes with grooves of different spacings against the participants' tongue tip: subjects were then asked to identify the orientation of the grooves. Furthermore, the relation between the measured acuities and the production of the sibilant contrast was investigated. Results show that a combination of somatosensory and auditory acuity best predicts the produced contrast, and based on this, the authors assume that sibilants have auditory and somatosensory goals. More information on studies addressing variability in production strategies of sibilants is given in the relevant chapter (Chapter 5).

To sum up, NATURE or biological predispositions have been found to play a role in the speech production process. Intra- and inter-speaker variability is influenced by physiological properties. A piece of evidence for the crucial impact of the factor NATURE is the significant role of somatosensory feedback in speech production and speech motor learning. The relevance of somatosensory feedback has been shown by several studies as described above. For one thing, vocal tract properties (like the physiology of the jaw and tongue muscles, the shape of the palate or the interplay of tongue and palate) affect the phonetic realisation of various phonemes and the differences in articulatory token-to-token variability. Furthermore, and even more relevant for the present study, differences in individual physiology have been found to influence speaker-specific articulatory behavior (palatal shape influences the amount of intra-speaker variability, vocal tract geometry affects articulatory distances in vowel space).

Thus, one hypothesis that will be investigated in this study is that speaker-specific variability is at least partly influenced by individual differences in a speaker's physiology. To investigate and evaluate the possible impact of NATURE on inter-speaker variability, the subject group

under investigation consists of monozygotic and dizygotic twins, who differ in the extent of their shared physiology. The subject group is described in more detail in Chapter 2.

So far we have not paid much attention to the fact that a speaker's physiology does not only play a role in the production part of the speech process but also in perception: only those sounds can be part of a phoneme inventory that can be *perceived*, and in particular that can be reliably distinguished by a native listener from other phonemes of that language. Just how much phoneme dependent acoustic variability is allowed in the language is dependent on the phoneme inventory. To this end, a relationship between the amount of intra-speaker variability and the size of the phoneme inventory exists. The aspect of *learning* comes to the fore which is part of the discussion in the following section. Here, the second factor that may influence the degree of inter-speaker variability will be presented in more detail: NURTURE – i.e. learning from observing, the influence of social environment and the role of auditory feedback.

1.3 The role of NURTURE and auditory feedback in speech production

In this section the various impact factors related with the concept of NURTURE will be discussed. First, some light will be shed on the influence of social-environmental factors and the processes of observation, adaptation and learning in both first and second language acquisition. In regard to language change and socio-linguistic variation, exemplar based models play a role as theoretical conceptualizations that integrate a perception and a production component. In addition, several selected studies examining the interacting relationship between speech perception and speech production will be presented. A particular focus will be laid on perturbation studies that revealed the importance of auditory feedback and auditory acuity. Finally, studies from speech pathology will be discussed that can give further support for the relevance of auditory feedback and the process of learning.

1.3.1 The influence of social environment, observation and adaptation

Speech acquisition has to proceed at least in part independently of individual differences in the physiology of the vocal apparatus, as it is in general possible for a child to learn and speak any existing language, provided that it is young enough and does not have any speech, language or hearing impairment. Theories of learning in psychology such as the *Social Learning Theory* of Albert Bandura (1977) emphasize that people in general learn by **observing and mimicking.** In terms of language acquisition this implies that children learn the syntactic and prosodic structures, phonological patterns and lexical entries of a language through imitation of the people surrounding and talking to them (i.e. especially in the beginning, mothers). Moreover, dialectal pronunciation and sociolinguistic parameters of the parents are also observed and absorbed by the child. Thus, **social-environmental factors** (NURTURE) play an important role in speech production. In regard to the influence of NURTURE on learning, a very important recent discovery in neuroscience is *mirror neurons* (di Pellegrino et al. 1992, Rizzolatti et al. 2001). *Mirror neurons* (observed in primates and birds) are neurons that are activated both when an animal acts and when it observes someone else doing the same action; hence, they link action observation with action execution. These mirror neurons are also assumed to be present in humans, and some researchers believe them to be very important in imitation and language acquisition (Rizzolatti & Craighero 2004, Rizzolatti & Arbib 1998). However, additional research must be done to support these assumptions.

An influence of NURTURE on speech has also been found in studies dealing with **second language acquisition and bilingualism**. The loss of a first or second language (L1, L2) or a portion of that language in bilingual speakers is known under the term *language attrition*. The native language that one speaks can affect the performance in a second language. This interference phenomenon from the first language to the second language system has long been a focus of research (e.g. Köpke & Schmid 2007). However, in addition, studies have also shown that the second language can influence the native language, i.e. *first language attrition* (FLA) (see e.g. Cook 2003 for an overview). Depending on how old the speaker is when he learns the second language and how long he has been living in the new language environment, the proficiency of L1 can be affected.

In ***exemplar theories*** and *usage-based models* (Bybee 2001, Pierrehumbert 2001, 2002, Johnson 1997), perception and social-environmental influences are seen to be crucial, since it is assumed that more recently encountered utterances are stored with higher activation levels than older utterances. Hence, sociolinguistic variation may partly be explained by a change in NURTURE. Moreover, speaker-specific patterns of pronunciation are handled in exemplar-based models by defining language sound patterns by extension rather than by rule. However, there are relatively few studies that test these assumptions (i.e. that people store exemplars of speech), and nothing is known about where and how the information that is constantly updated is stored (Johnson 2007).

From studies within the fields of **language change and sociolinguistic**s it is known that all languages change subtly over time. This includes pronunciation but also syntactic structures or lexical patterns. The actual starting point of language change, which is also described in the literature under the concept of the *actuation problem*, is difficult to assess since a change can be observed only when the leveling process has already started and the change has diverged and been adopted by a larger group of people (Labov 1980). The leveling process covers the adoption and diffusion of the innovation by way of social networks and speech communities. Studies have shown that younger (female) members of a speech community are often the leaders of a change (Warren 2005). An interesting investigation of an individual's language change over a lifetime was conducted by Harrington (2000, 2005, 2007) and Harrington et al. (2000) regarding the Queen of England's pronunciation between the years 1952-1980. They investigated her annual Christmas message over the years and analyzed the first two formants of 11 vowels. The results show that the Queen's pronunciation changed over the years: the pronunciation was influenced by the standard southern British accent of the 1980s and became less RP like. This accent is typically associated with speakers who are younger but also lower in the social hierarchy of Britain. Hence, the study shows a good example of the influence of the adoption of subtle changes in speech, which takes place without the speaker even beeing conscious of it. Harrington (2005) extends the earlier research by analyzing changes that might be due to vocal tract maturation. He investigated formant changes in schwa vowels. The schwa is characterized by a quasi-neutral phonetic quality, it may be virtually targetless (van Bergem 1994) and there is no evidence for any diachronic change. By this means the study tries to separate changes that might be due to vocal tract maturation (in

schwas) from those due to phonetic changes (in other vowels). Significant decreases in F1, F2 and F4 and a significant increase in F3 were found in the Queen's schwas between 1950 and 1990. The author notes that these specific and variable formant changes can be neither a result of perceptual compensations nor of physiological effects of vocal tract maturation like an increase in vocal tract length. Thus, even though it cannot be ruled out that the observed tensing of the Queen's [ɪ:] vowel might also partly be influenced by long-term physiological and/or perceptual changes over a period of 40 years, a phonetic change has taken place, too.

1.3.2 The relevance of learning and auditory feedback in speech production

The relevance of the **factor *learning*** becomes evident by looking at the interplay between language's phoneme inventories and the respective allowed amount of phonetic variation in these phonemes. In cross-linguistic studies it has been shown that languages differ in the amount of variability within a certain phoneme, depending on the size of the phoneme inventory of these languages (Lavoie 2002, Manuel 1990, Jongman et al. 1985). A language's phoneme inventory constrains the variation allowed in the realization of a phoneme. While Manuel (1990) has shown that the phoneme inventory constrains the coarticulation of vowels, Lavoie (2002) found the same for the variation in manner of articulation. Lavoie investigated /k/ in English and Spanish and found it to be more variable in terms of finding frication noise accompanying the realization of the stop in English than in Spanish. Lavoie explains this occurrence with the fact that English has no contrastive voiceless velar fricative, but Spanish does. The study of Jongman et al. (1985) revealed that the place of articulation in consonants is also restricted by the phoneme inventory. For instance, English and Dutch display more variation in place of articulation for stops than Malayalam, because English and Dutch do not need to distinguish between a dental and an alveolar stop. These studies reveal the fact that the amount of allowed phoneme variation is restricted by the respective phoneme inventory. Thus, speakers of a language have to follow the restrictions that a language has concerning variability and learn to adapt their productions to the allowed variability.

Auditory references play an important role in the process of learning. In the acquisition of new sounds in a second language it has been demonstrated that speakers make use of

auditory perceptual categories as a reference for articulation (Flege 1995). The ability to distinguish different speech sounds helps to master the language and advances fluency. Thus, training in speech perception can facilitate and expedite the learning of sound production (Rvachew 1994, Bradlow et al. 1997). Bradlow et al. investigated the effect of perceptual identification training of /r/-/l/ on the production of these contrasting sounds in adult Japanese speakers learning English as a second language. The Japanese participants were recorded before and after the training program and English listeners rated these productions in a two-alternative minimal-pair identification task. All speakers showed improvements in the perception and identification of the /r/-/l/ contrast, and also in the production of these sounds, as measured by the more accurate identification rate of English listeners following perceptual learning.

The **interacting relationship between speech perception and speech production** is described in Guenther's neurolinguistic model of speech production (Guenther 1995, Guenther et al. 1998, Guenther et al. 2006): he assumes a feedback-based learning process that results in an internal model of the required speech movements. Acoustic results are compared to stored auditory goals and by this means a speaker learns and stabilizes the necessary speech motor commands. Auditory feedback and thus correct auditory representations play an important role in learning new speech sounds. Lipski et al. (2011) trained native German speakers on Italian geminates with and without auditory feedback. They found that auditory feedback is necessary for learning non-native speech sounds and precise coordination of articulation even when somatosensory feedback is salient (in the production of a bilabial plosive).

Several studies have investigated the relationship between speech perception and speech production in normal speech (Newman 2003, Perkell, Guenther et al. 2004, Perkell, Matthies et al. 2004, Perkell et al. 2008). Newman's (2003) study reveals the importance of auditory feedback and speech perception in speech production by comparing acoustic parameters of listeners' perceptual prototypes with their average productions. The relevant acoustic parameters found for the **production-perception correlations** turned out to be VOT for stops, and spectral peak values for fricatives. The study shows the relevance of perception in speech production: individual differences in production are correlated to differences in perception.

In Guenther's DIVA model of speech motor planning (Guenther 1995, Guenther et al. 1998, Guenther et al. 2006) phonemic goals correspond to multidimensional regions in auditory *and* somatosensory domains, where the latter may be more important for sibilants than for vowels, given their more extensive contact between the tongue and other oral structures. The particular size and spacing of these goal regions may depend on the individual **speaker's perceptual/auditory *acuity***. Auditory acuity is defined as clearness, sharpness or distinctness of perception or the sensitivity of the auditory apparatus of a speaker. It is estimated by a same-different discrimination task, in which the *just noticeable difference* between two stimuli of a speaker is determined. The authors of the model assume that speakers with higher acuity should form smaller goal regions that are spaced further apart. The role of a speaker's perceptual acuity in speech production has been shown for vowels (Perkell, Guenther et al. 2004, Perkell et al. 2008), but also for sibilants (Perkell, Matthies et al. 2004, Ghosh et al. 2010). Perkell, Guenther et al. (2004) found that speakers with more accurate discriminations of vowel contrasts were also more distinct in producing this contrast. The results are interpreted in favor of a model of speech production in which "articulatory movements for vowels are planned primarily in auditory space" (Perkell, Guenther et al. 2004, p. 2338). Perkell, Matthies et al. (2004) investigated the influence of auditory goals on sibilants. The amount of contact between tongue tip and alveolar ridge during the production of /s/ and /ʃ/, the acoustic spectra of the fricatives, and the speaker's perceptual acuity were examined for each of the subjects. The results again point to a strong relationship between the ability to perceive differences between sibilant continua and the distinctiveness in producing the sound contrasts. However, it should be noted that Ghosh et al. (2010), among others, emphasize that besides auditory goals, somatosensory goals also play a crucial role in the production of sibilants. This might be due to the difference in the amount of tongue-palate contact between sibilants and vowels and should be kept in mind for the further analysis. This topic will be discussed in more detail in the chapter on inter-speaker variability in sibilants (Chapter 5).

Even though no **perturbation experiment** will be done in this study, this type of experiment is relevant for the present analysis since it can give insights into the role of auditory feedback. Speakers compensate for a mismatch between target and acoustic result through a correction of articulation as soon as 100 to 150 ms after the perturbation (Lipski et al. 2011). The study

of Jones & Munhall (2003) investigates the **contribution of auditory feedback to the process of adapting** while speaking with a dental prosthesis that extends the length of the maxillary incisor teeth. Subjects had to say /tas/ with normal auditory feedback available and with masking noise. A perception experiment revealed that speakers used auditory information to compensate for the vocal tract modifications: productions made with auditory feedback available were judged by the listeners to be more "normal" than productions made without auditory feedback. The study shows that auditory feedback can be used for online corrections but also for longer-term calibration since the compensatory articulations were learned and still available after auditory feedback was removed again. An interesting finding was that the listeners' evaluations of the different productions of /s/ were more sensitive than the acoustic measurements (Center of Gravitiy). The acoustic analysis could not find any significant differences in learning depending on auditory feedback, pointing to a reliance on auditory *and* somatosensory goals for the production of /s/.

Recently, investigations have been conducted to test the role of **online** auditory or somatosensory **feedback perturbation** in speech production and speech motor planning. In these studies the auditory feedback is perturbed in real time and as a result the subjects spontaneously change their speech production to compensate for the perturbation while listening to their own altered productions. Shiller et al. (2009) found compensatory responses (i.e. an adaptation of motor plans for /s/-productions) as a reaction to auditory perturbations of the sibilant. The spectrum of the sibilant was modified in real time by shifting the first spectral moment (centroid) down so that it was closer to the fricative /ʃ/. Subjects reacted and adjusted their production to counteract the effect of the perturbation. As a result, an increase in the /s/-centroid frequency was found (compared to the non-perturbed condition). Furthermore, adaptive strategies in response to auditory feedback perturbation have been found for F0 (Jones & Munhall 2000) and vowel quality (Villacorta et al. 2007). The study of Perkell et al. (2007) is situated in the field of *speech sensorimotor adaptation*, which the authors describe as "an alternation of the performance of a motor task that results from the modification of sensory feedback" (Perkell et al. 2007, p. 2306). Three experiments were conducted: 1) to investigate auditory sensorimotor adaptation of the first formant in different vowels, 2) to examine the relation between speaker acuity and amount of compensation to auditory perturbation, and 3) to simulate the subject's performance in experiments 1 and 2 by

means of the DIVA model (Guenther et al. 1998). A first important finding was that subjects compensate their speech in response to perturbations of the first formant in the acoustic feedback by producing vowels with first formants shifted *opposite* to the perturbation. The compensations persisted for a period of time when auditory feedback was masked, revealing true adaptation. Furthermore, a generalization of the compensation took place, since other vowels were affected, too. Second, inter-speaker variation in the extent of adaptation was found and could be explained by differences in auditory acuity: subjects with greater acuity showed greater compensatory responses.

In addition to studies with static speech sounds like monophthongs, temporal manipulations of consonants have also been investigated. Mitsuya et al. (2009) examined in their study the role of online auditory feedback in the voice onset time (VOT) of the alveolar stops /t/ and /d/. While saying "dip" or "tip" the subjects listened to their own voice saying the respective other word. They found that speakers changed their speech and compensated for the VOT perturbation, i.e. they lengthened VOT for /t/ (made it more t-like) when they heard /d/ with a short VOT. Recently, studies have also investigated time-varying sounds by means of formant trajectories (Cai et al. 2010). Cai et al. perturbed the auditory feedback (i.e. second formant frequency trajectory) of 20 native speakers of Mandarin while they were producing the triphthong /iau/ and measured their patterns of auditory-motor adaptation. Results again show that speakers change their formant trajectories in the direction opposite to that of the perturbation.

The abovementioned studies regarding 1) the link between production and perception in speech and 2) the compensation for and adaptation to online auditory perturbations point to the general importance of auditory feedback in speech production. The ability of a speaker to react to an acoustic stimulus by using compensatory responses that persist even when feedback is temporarily blocked or removed is a crucial factor in speech production. This can also be seen in studies addressing issues in **speech pathology** and in particular with subjects who have hearing impairments and with cochlear implant users. It is well known that the speech produced by the deaf is generally of low intelligibility with all kinds of phonetic details concerning the quality and duration of vowels and consonants being affected. Various studies have been conducted and research has shown that phonetic parameters like acoustic distance between vowels, acoustic clustering within vowel categories and vowel duration are affected

by degradation of auditory feedback (Vick et al. 2001, Perkell et al. 2001, 2007, Svirsky & Tobey 1991, Lane et al. 1995, 2005). The investigation by Ménard et al. (2007) studies the impact of auditory feedback on vowel production in postlingually deaf adults with cochlear implants (CI) and in a normal-hearing control group. Three recordings containing 9 American English vowels were made a) prior to implantation of the CI, b) one month after implantation and c) one year after implantation, and two feedback conditions after the implantation (implant processor turned on and off) were examined. An acoustic vowel space for each speaker, recording and condition was created by calculating Euclidean Distances between the mean formant frequencies for all vowel pairs. In agreement with previous studies, CI users had in all cases lower vowel contrast values than the control group. As expected the vowel contrasts were larger with hearing on than off and improved from one month to one year after the implantation. The authors assume that the CI users could retune their auditory feedback system to some extent within one year, resulting in increased vowel contrast distances.

To sum up, NURTURE has been shown to be an influencing factor in the speech production process. Factors like social environment and learning by observing play a crucial role in speech acquisition (of both first and second languages). A variety of studies have demonstrated the important role of auditory feedback in speech production. Processes like online monitoring, situational adaptation and memory experiences reveal that stored auditory goals are used in speech. Acoustic inter-speaker variability in speech can result from differences in social environment (or the perceptual input). Moreover, differences between speakers' productions can be explained by differences in speakers' perceptual abilities and auditory acuities.

Thus, with respect to the present study and the investigation of twins' speech it should be kept in mind that NURTURE is a possible influencing factor in inter-speaker variability in speech and that learned auditory goals are considered to play an important role. However, in the preceding section the power of NATURE and the role of somatosensory feedback were also made clear. The question that arises is which factor is more important in influencing inter-speaker variability: NATURE or NURTURE? Moreover, do additional factors exist that interact and possibly intensify the power of one of the two?

1.4 NATURE vs. NURTURE?

There is agreement in research that speech has a biological grounding. NATURE is a shaping force by virtue of the fact that physiological constraints have an influence on the shape of phonological systems. Studies on vowels and consonants (fricatives and especially velar stops) have shown that anatomical restrictions like the vocal tract geometry and the palatal shape influence token-to-token variability. Moreover, individual differences in physiology cause inter-speaker variability. Tactile or somatosensory feedback is necessary for articulation (for sibilants more than for vowels), and especially perturbation experiments have shown the important role of somatosensory feedback over auditory feedback, since the adaptation to the perturbation takes place even without auditory feedback. In addition, somatosensory feedback is necessary to store motor programs and influences speech motor learning. Studies have shown that inter-speaker variability can be explained in terms of differences in somatosensory acuity.

Yet, the roles of social environment and adaptation are well known and are gaining recognition in theories from psychology, like the Social Learning Theory. Moreover, an influence of NURTURE also finds support in theories of speech perception, for example in exemplar-based models (Bybee 2001, Pierrehumbert 2001, 2002, Johnson 1997), and helps to explain language change. In addition, many studies have revealed the importance of auditory feedback in speech production. Studies on second language acquisition have shown that auditory perceptual categories are used as a reference for articulation and the acquisition of new sounds. From studies in speech pathology we know that auditory impairment leads to reduced intelligibility and lower phoneme contrasts. However, in the speech production of normally developed populations, the link between perception and production has also been found and the importance of stored auditory goals has been shown as well. A speaker's perceptual or auditory acuity can explain inter-speaker variability in vowels but also in sibilants (although to a lesser extent, and here somatosensory acuity has to be taken into account too). Perturbation experiments have revealed that speakers use adaptive strategies in response to auditory feedback perturbation and that differences between speakers in the extent of adaptation to an online feedback perturbation can result from differences in auditory acuity.

To summarize, it does not seem reasonable to neglect one of the two factors in favor of the other. Both NATURE and NURTURE have explanatory power and both are crucial factors in human speech production and perception. Thus, no choice will be made between them. Both factors contribute to the speech process and can account for intra- and inter-speaker variability as has been shown in the abovementioned studies. Both are crucial components of a self-monitored continuously updated system. NATURE can influence fine-phonetic details (e.g. the shape of the palate affects articulatory variability). In addition, within the confines of physiology (NATURE) speakers have choices they make or do not make (depending on NURTURE). However, it is unclear whether one of the two factors plays a greater role than the other: the relative importance of the factors might change depending on which phoneme category is considered. For example, it could be hypothesized that for vowels auditory goals are most important, but for consonants (and here especially sibilants) somatosensory goals and physiological restrictions are crucial since the two sound groups differ in their amount of linguo-palatal contact. In the following analysis a distinction between vowels on the one hand (Chapter 4) and sibilants on the other (Chapter 5) will be made and possible differences in the influence of NATURE between these two sound classes will be discussed.

A second hypothesis could be that acoustic *transitions* and articulatory *gestures* are more influenced by individual anatomical and physiological characteristics of a speaker than acoustic and articulatory *targets*. Targets are oriented towards auditory goals and are shared between speakers of the same speech group, but transitions result as a by-product when speakers move from one target to the next and mirror individual differences in physiology. Thus, a distinction will be made between articulatory and acoustic targets (in vowels and sibilants) on the one hand, and acoustic transitions (in sibilant-vowel sequences, see Chapter 5) and articulatory gestures (in /aka/ sequences, see Chapter 6) on the other. The analyses aim at finding a possible difference in the influence of the factor NATURE on these two speech characteristics.

Thus, two main research questions should be kept in mind for the analyses of the twin data in the following chapters:

1) Is there a difference in the influence of NATURE on inter-speaker variability depending on the phoneme category (and thus the amount of linguo-palatal contact)?

2) Is there a difference in the influence of NATURE on inter-speaker variability depending on the particular characteristics of the analyzed parameter: target (static) vs. transition (dynamic)?

In this study I investigate inter-speaker variability in twins' speech, where the lively debate on NATURE vs. NURTURE comes into play as a result of the subject group I have chosen. The characteristics of this subject group and their significance for the present study are described in the following chapter.

2 THE ROLE OF TWIN STUDIES IN INVESTIGATING THE FACTORS NATURE AND NURTURE

2.1 Twin studies

Twin studies are a common type of investigation in the field of psychology. The origins of twin studies go back to the late 19th century and are ascribed to Sir Francis Galton (Galton 1876), even though it is not clear whether he was aware of the difference between monozygotic twins (who are genetically identical) and dizygotic twins (who share around 50% of their genes on average). Nowadays the systematic comparison of the within-pair similarity of monozygotic (hereafter **MZ**) twins with that of dizygotic (hereafter **DZ**) twins is a standard procedure in the field of behavioral genetic research. The aim is to investigate individual differences and to explain the variation in terms of two possible influencing factors: (1) *NATURE* (genes and physiology) and (2) *NURTURE* (environment). The latter factor refers to social-environmental factors that contribute to the resemblance between individuals who grow up in the same family. The Equal Environments Assumption (EEA) assumes that MZ and DZ twins share the same amount of environmentally based similarity. This crucial assumption has been investigated intensively and studies of mislabeled twin pairs (i.e. DZ twins that grew up as MZ twins and MZ twins that grew up as DZ twins) have shown the validity of this assumption (Scarr & Carter-Saltzman 1979). Additionally, the study by Koeppen-Schomerus et al. (2003) regarding language and cognitive measures of 2- and 3-year-old twins and non-twin siblings shows that the estimated amounts of shared environment are more than twice as large for twins (DZ and MZ) as compared to non-twin siblings. The only difference in terms of the two factors *NATURE* and *NURTURE* between MZ and DZ twins who are still living together and share their social environment is the difference in their genetic or physiological similarity. Thus, the assumption can be made that if MZ twins are more similar in an investigated parameter than DZ twins, it points to the

importance of the genetic influence. Regarding personality traits, research suggests that around 40% of the variance in personality is due to genetic variance and around 60% is due to variance in the person's specific (non-shared) environment (Wolf et al. 2003). The extensive German Observational Study of Adult Twins (GOSAT) investigates the influence of shared environment on individual differences in personality traits. Three hundred sex-matched twin pairs aged between 18 and 70 were observed, tested and interviewed by 60 different judges for each twin (Spinath et al. 1999). They found that differences in personality can be explained by "additive genetic influences" (42%), "shared environment" (18%), and "specific environment" (35%). The assumption could be confirmed that behavioral tendencies are grounded in genes, but specific reactions in specific situations are not. An interesting finding relevant to the current study is that the factor *shared environment* turned out to be more important than was expected in earlier studies. The factor seems to play an important role and should be kept in mind when investigating differences in monozygotic and dizygotic twins.

In order to study the influence of biological parameters (physiology, biomechanics) as well as of non-biological parameters (learning, environmental factors), this study investigates inter-speaker variability in the speech of monozygotic twins and dizygotic twins. In detail, this means that if high inter-speaker variability in a certain speech parameter within an MZ twin pair is found, the influence of genetics and physiology on this parameter would seem to be rather small. Results can also be discussed in terms of articulatory and auditory targets. It should be noted that articulatory targets can of course also be learned and do not have to be different only because of different preconditions in physiology. However, if an influence of biomechanics and vocal tract physiology is assumed in speech production, the detailed articulatory target positions and precise movements might be influenced by NATURE, and thus articulatory (and hence also acoustic) inter-speaker variability in MZ twin pairs should be very low (independent of the time they spend together). If, on the other hand, auditory goals are assumed and the role of NURTURE is seen as the important factor, the MZ twins should differ in their acoustic output when they are living apart from each other. In addition, if a DZ pair that spends most of their time together is very similar in their acoustic outputs of a certain speech parameter, the role of NURTURE and auditory goals prevails over the impact of shared physiology and NATURE.

In the present chapter the emphasis is placed on the possible contribution of investigating *twins' speech* in exploring the effects of NATURE and NURTURE on inter-speaker variability. Twin studies in the fields of speech acquisition and speech pathology (Section 2.2.1) as well as addressing normal speech are described. Here, a subdivision is made into studies doing perception tests and exploring the perceived auditory similarity of twins (cf. Section 2.2.2.1) and acoustic studies looking for spectral parameters in the speech signal that can differentiate twins (cf. Section 2.2.2.2). The chapter ends with a description of a pilot study that was done to help find suitable speech material and detailed research questions for the present study (2.3).

2.2 Twin studies in speech research

2.2.1 Speech acquisition and speech pathology

Within the debate about the innateness of language and linguistic knowledge (Chomsky 1975) twin studies have especially been conducted in the fields of speech acquisition and speech pathology. Locke & Mather (1989) investigated in their twin study the genetic factors in the ontogeny of speech with 13 MZ and 13 DZ twin pairs. They analyzed more extensively data from Mather & Black (1984), who tested monozygotic and dizygotic twin pairs aged 3-5 years on their **verbal abilities** using the *Templin-Darley Screening Test of Articulation* (Templin & Darley 1969). The production errors were classified into sound substitutions, omissions and distortions by two different examiners. From their paper it is not clear whether the investigators recorded the children or analyzed the speech errors directly while listening to them. In some cases it might be difficult to decide whether an error is a substitution or a distortion, especially when it cannot be checked by listening to it several times or even better by analyzing the acoustic signal or even the underlying articulatory pattern. Pouplier & Gouldstein (2005) investigated the perception and categorization of phonologically ill-formed errors and their actual articulatory mechanisms. They showed that the errors are perceived in most cases as substitutions even though other mechanisms are at work, like the simultaneous production of two gestures. Thus, the authors point out the difficulty of an impressionistic categorization of speech errors and the thereby resulting asymmetry in error distributions.

Locke & Mather compared the within-pair similarity between MZ and DZ twins and found that MZ twins are significantly more similar regarding the quantity of speech errors than DZ twins: 82% of the speech errors are shared errors for MZ twins vs. approximately 60% for DZ twins (and also for unrelated pairs that match in sex, age, dialect and socio-economic class). However, no significant difference between MZ and DZ twins was found for the quality or kind of error (i.e. if it is a substitution, omission or distortion). Since it has been shown in Pouplier & Goldstein's study how difficult it is to categorize a speech error perceptually and it seems that no acoustic analysis was done to check the classification of errors, these results should be treated carefully. The misidentification of speech errors due to the difficulty of impressionistic categorization could be a possible explanation for the somewhat surprising result that MZ twins are more likely to misproduce the same phonological target than DZ twins but not in the same way.

Ooki (2005) found a genetic effect on the occurrence of **stuttering** and tics in Japanese twins. Ooki investigated 1896 male and female MZ and DZ twin pairs from 3 to 15 years by means of a questionnaire concerning the prevalence of tics and stuttering. The concordance rates for both parameters were higher for MZ twins than for DZ twins regardless of sex combination.

Simberg et al. (2009) investigated in a huge twin sample of 125 monozygotic and 108 dizygotic Finnish twin pairs the influence of genetic and environmental factors on **dysphonia**. Dysphonia is an impairment of the ability to produce voice and can have functional or organic causes. It is a phonation disorder and is typically caused by an interruption of the periodically vibrating vocal folds. The twins completed a questionnaire concerning vocal symptoms. Participants had to report how frequently they suffer from symptoms like "my voice gets strained," "I feel pain in my throat," and "I have difficulty being heard" on a scale from daily to never. In addition, the twins were grouped into participants working in voice-demanding jobs (e.g. teacher, salesman, lawyer), and less voice-demanding occupations (engineer, factory worker, researcher). Results showed that differences in symptoms of dysphonia were explained by non-shared environmental effects (65%) and to a lesser degree by genetic effects (35%). Thus, even though a genetic effect in dysphonia could be found, it turned out that environmental factors play an even more important role. Hence the importance of controlling non-shared environmental factors in a study containing twins becomes obvious and has to be kept in mind. The relevance of voice

quality properties like micro-perturbations in frequency (measured as *jitter*) will come to the fore again in Chapter 7, when the acoustic correlates of perceived auditory similarity are analyzed and discussed.

2.2.2 Normal speech

Comparing the within-pair similarity of DZ and MZ twins regarding speaker-specific characteristics of *normal* speech is rather new and a less frequent research topic (see Loakes 2006, p. 41). Still, some studies regarding perceptual and acoustic differences within twin pairs have been conducted, and several of the relevant studies will be described and examined more closely in the following section. First, perception experiments addressing perceived similarity of twins' speech are described, and then studies dealing with acoustic analysis are reviewed.

2.2.2.1 Perception experiments

Since a perception experiment investigating the perceived auditory similarity of MZ and DZ twins is carried out in the present investigation (see Chapter 7), some relevant studies concerning this topic are described below.

The similarity of twins' voices has been observed even by twins themselves, who have difficulties identifying their own voices when presented with recordings of their own voice and the voice of their twin (Gedda et al. 1960, Cornut 1971). Perception experiments have revealed the striking similarity between twins' voices, but have also shown that monozygotic and dizygotic twin pairs can be differentiated above chance. In a study by Whiteside & Rixon (2000) the two voices from one Irish English speaking male monozygotic twin pair had to be identified by listeners familiar with their voices. The stimuli were based on pure monosyllables (produced by one speaker) and 'fused' monosyllables (produced by different speakers of the same twin pair). Listeners successfully identified the speaker for the pure syllables in about 70% of the cases but had difficulties as expected with the fused syllables. Given that there were only 2 speakers from which the listeners had to choose, the misidentification rate for the pure syllables of 30% was actually quite high. The method of using fused syllables is somewhat difficult to interpret since no clear conclusions can be drawn from the results. The

only information that can be gained by this is the expected fact that listeners have difficulties assigning a fused stimulus to one of the twins. However, the study could show that familiar listeners can distinguish twins above chance even with short monosyllabic words, but that the twins' voices are nevertheless perceptually very similar since the probability of correct identification was only 70%. Note, though, that an investigation with a subject sample of only one twin pair can only be seen as a case study and no generalizations should be made from this.

Decoster et al. (2001) also investigated the question of whether twin' voices sound similar and whether they can be identified as twins. Thirty MZ twin pairs (20 female and 10 male pairs) aged between 18 and 31 took part in their study. The investigation was twofold: 1) a perception test and 2) an acoustic analysis were conduced. Ten listeners heard three stimuli of which two were a twin pair and the third was a non-related speaker from one of the other twin pairs. Nothing detailed is said about the third voice, so we do not know if this third stimulus is matched for age or how large the degree of (subjectively) perceived auditory similarity is. The listeners were asked to decide which of the two stimuli were from twins. Two experimental conditions regarding the length of the presented stimuli were used: a) two sentences of read speech of standard Dutch and b) a 2.5-second midsection of a sustained /a/. As expected, the results showed that read speech allowed twins to be correctly selected in more cases than /a/. For female voices 82% of the read speech samples and 63% of the sustained /a/ samples were correctly labeled. Interestingly male voices had lower scores: only 74% (and 52% respectively) of the twin pairs were matched correctly. For the acoustic analysis the authors examined the fundamental frequency (F0) of the sentences and the sustained /a/ and assessed the correlations of mean F0 between the two speakers of each pair (intra-twin pair). For both conditions a significant degree of correlation between the speakers of the male and female twin pairs could be found, but with a higher significance for male voices, even though males were more difficult to identify as twins in the perception test. The authors assume that F0 is a useful characteristic in the perceptual identification of twins but also note that male twins that were more similar in their mean F0 were actually more difficult to identify. The results point to an influence of mean F0 on perceived similarity but also make clear that other, not analyzed acoustic parameters must have an effect.

Another perception study that is relevant to the current investigation and deals with the perception of personal identity in speech is from Johnson & Azara (2000). In their introduction they describe different accounts for inter-speaker variability. The 'radical invariance' view explains all individual differences between speakers of the same dialect with anatomical factors (Nordström & Lindblom 1975). However, we find idiosyncratic variation that is not determined by anatomy or dialect but turns out to be due to speakers' learned individual speaking strategies (e.g. Johnson et al. 1993). For this reason, Johnson & Azara suggest analyzing the perceived similarity of twins that share anatomy and dialect. They assume that perceived talker variability within this speaker group is due to idiosyncratic variation. In addition, they investigated the question of whether the sensitivity to the perceived differences between speakers is higher when the listeners know that twins served as speakers. To this end, they carried out different perception tests with varying instructions for the listeners (regarding the information about whether speakers were twins or not), but it turned out that the results were stable over the changes in experimental conditions. They conducted 3 perception tests (each with 10 listeners) with the speech of 6 female twin pairs ranging in age from 20 to 67 and speaking various dialects of American English. Five of the pairs were MZ and one pair was DZ. Participants listened to stimuli-pairs and decided whether the stimuli came from different speakers or the same speakers. Different words served as stimuli, but in the result section no further information is given about the possible influence of different stimuli on the perceived similarity. In general, the results of the perception tests indicate that listeners can distinguish isolated words spoken by unrelated speakers but also (to a lesser degree) by twins. An interesting finding is the pair-specific perceived similarity: two unrelated speakers were more similar (within a perceptual map that was constructed by a multidimensional scaling analysis) than two speakers of the same twin pair. As possible influencing factors on perceived similarity the authors name age and dialect but also point to other acoustic correlates, like aspects of phonation such as breathiness. However, these factors were not analyzed in this study. Another finding was that the dizygotic twins were not less confusable than the monozygotic twins, but since the group of dizygotic twins was represented by only one pair, this result should not be generalized due to the effect of speaker- (or pair-) specific parameters. Interestingly the factor *age* turned out to play no role, since the oldest twin pair and the youngest twin pair were most similar to each other. A limitation of the study is the heterogeneous speaker group, which makes the

interpretation of the results somewhat difficult, especially since little to no information is given about the time the twins spent/lived together or the attitude they have towards being a twin. Both factors could have an influence on speech characteristics and may explain the pair-specific similarity. However, the study shows that learned idiosyncratic variation independent of anatomical or linguistic variation is evident, since listeners were able to identify twins above chance by listening to just one word.

All of these studies have served as helpful examples of perception tests with twins' speech for the perception test that is conducted in this work as described in Chapter 7. In the present study, emphasis has been placed on sampling a more homogeneous group of speakers with controlled environmental factors like time spent together or attitude towards being a twin. In addition, more than one dizygotic twin pair represents this group. Moreover, an acoustic analysis is conducted that looks for acoustic correlates which could explain the results regarding differences in perceived similarity.

2.2.2.2 Acoustic studies

The most frequently acoustically investigated speech parameter in twins' speech is **fundamental frequency (F0)** and results point to a great influence of physiology on this parameter, since MZ twins reveal higher correlations than DZ twins (Przybyla et al. 1992, Debruyne et al. 2002). The comprehensive study of Przybyla et al. (1992) includes 122 twins (61 pairs), who were analyzed regarding their speaking fundamental frequency. However, of the large number of twin pairs who took part in this study only 9 pairs were DZ. Nevertheless, the study provides some evidence regarding the impact of genetics on the voice parameter fundamental frequency, with MZ and DZ twins showing higher correlations than unrelated pairs, and MZ pairs showing fewer intra-pair differences than DZ pairs. In addition to these findings the authors emphasize the correlation between age and weight and fundamental frequency, and the importance of controlling for this factor when investigating the impact of genetics.

A lesser influence of identical genes and physiology and a greater impact of environmental factors were found by Debruyne et al. (2002) for what they call **variation of speaking fundamental frequency**, as MZ and DZ twins revealed the same amount of similarity. Their

study comprises 60 twin pairs (30 MZ and 30 DZ) and investigates the speaking fundamental frequency (SFF) and the intra-speaker variation of SFF from a read text. The investigators controlled for environmental influences like smoking and drinking habits or the possible influence of medication, but no information is given about the amount of time the DZ or MZ twins had recently spent together. It cannot be factored out that a mutual influence and adaptation of the siblings might play a role even in a source-based parameter like pitch range, since the speaker's *used* pitch range does not necessarily have to be the *organic* one, i.e. "the maximum range of which the speaker is physically capable, given the biologically determined factors of his or her laryngeal anatomy and physiology" (Laver 1994, p. 457). The authors conclude from their findings of a more similar SFF in MZ than in DZ twins that this indicates a genetic influence on the SFF. Note, however, that since we do not know from the study whether the MZ and DZ twins show the same amount of shared social environment, this factor cannot be eliminated.

In the multi-parameter study of van Lierde et al. (2005) **voice quality characteristics** in 45 MZ twin pairs aged from 8 to 61 years were investigated. The authors controlled for possible influential factors like voice disorders, neuromotor dysfunction, chronic or upper airway problems, intensive smoking behavior, heavy vocal abuse, or hearing problems. Subjective (auditory evaluation) and objective (aerodynamic, voice range, acoustic) measurements were taken. The authors analyzed, for example, perceptual voice characteristics, the maximum phonation time, vocal performances, and the overall vocal quality by means of the *Dysphonia Severity Index*.[1] Correlation coefficients were calculated to investigate the relationship between the twins among the variables of voice production and to evaluate the twin inter-correlations. Their hypothesis that the investigated voice quality parameters are strongly influenced by genetics and biology and thus should be very similar in MZ twins could be confirmed for nearly all parameters. Interestingly, for the two parameters *shimmer* (micro-perturbations in amplitude) and *jitter* (micro-perturbations in frequency) only, no significant

[1] The Dysphonia Severity Index (DSI) is a quantitative measure of perceived vocal quality and is based on the weighted combination of the following voice measurements: highest frequency (F0-high in Hertz), lowest intensity (I-low in decibels), maximum phonation time (MPT in seconds), and jitter (in percent). The DSI ranges from +5 (in healthy voices) to -5 (in severely dysphonic voices).

twin inter-correlation coefficient could be obtained. Van Lierde et al. suggest that these two parameters may be influenced by other factors that have not been controlled for in their study, such as environment, state of health, anxiety or tension. They point to the complex interaction between genotype and environment and suggest further research. Note that the parameters jitter and shimmer are also investigated in the current study and will be discussed in Chapter 7 in connection with acoustic correlates of perceived similarity.

Results from the clinical study of Fuchs et al. (2000) support the abovementioned influence of genetics on voice quality. Fuchs et al. verified their assumption that the vocal performance and several acoustic features are more similar in monozygotic twins than in non-related persons independent of the age of the twin pairs. They investigated, for example, F0 range, minimum and maximum F0, mean F0 and voice intensity in 31 twin pairs (aged between 18 and 75) and in sex- and age-matched non-related pairs.

Ryalls et al. (2004) compared **voice onset time** (VOT) in two MZ twin pairs. The pairs differed in the amount of time they spent together: the one younger pair (21 years) was still living together, while the older pair (70 years) had lived in two different parts of the USA since they were 25. The authors measured VOT word-initially in 6 stop consonants (3 voiced stops: /b d g/ and 3 voiceless stops: /p t k/) within a CVC sequence. They found the younger twins to be more similar in their VOT than the older twins. Based on their results as well as earlier literature they conclude that *source* characteristics like voice properties are influenced in a stronger way by genetic constraints than *filter* characteristics like VOT or formant frequencies. These results may certainly provide some hints as to the influence of genetics on speech, however the small subject sample of only 1 pair per investigated group (living together and living apart) makes it difficult to draw any general conclusions. Additionally, the difference in age between the two groups (21 and 75) is a factor that is not accounted for in the study. The authors explain the differences in VOT for the older twin pair by their different environments, but since no other information on the subjects is given in their paper (nothing is known about the state of health of the subjects, smoking habits or medication) the effect of external factors other than living with the twin on the measured VOT cannot be excluded. It would be interesting to see whether two pairs that match in age, health and smoking habits, perhaps also living in the same area, and only differ ing in the amount of shared time would result in the same findings. Furthermore, the classification of

VOT as a *filter* characteristic is somewhat misleading. VOT is defined as the time between the burst, that is the release of the stop, and the start of voicing (in the paper this time point is defined as the highest point in the first cycle associated with periodic vocal fold vibration). However, in addition to the time factor (which the authors relate to the *filter* aspects of speech), the physiology of the vocal folds and the strength of the aspiration, hence the *source,* are definitely involved in creating VOT. Particularly the voicing that can take place before the actual burst (i.e. voicing during closure, VDC) and which the authors measured as negative VOT is considered to be associated with biological constraints of the vocal folds.

Within the area of forensic speaker identification, Nolan & Oh (1996) tried to find speaker variation that is independent of anatomy and linguistic background. To this end, they acoustically analyzed /l/ and /r/ phonemes of 3 identical female twin pairs between 21 and 23 years. All pairs were living together and studying at the same university, thus the authors could control for the influence of environmental factors. Actually, the study aimed at finding differences in **coarticulation patterns** that had been demonstrated in an earlier study for non-siblings (Nolan 1983). However, the authors could not find major differences in the coarticulatory behavior of /l/ and /r/ for the twins. Nevertheless, the study reveals that "monozygotic twins are not necessarily phonetically identical and that they can make use of the leeway allowed them by the phonological system of their language" (Nolan & Oh 1996, p. 48f) since the results show differences in vowel formants following /l/ and /r/ and thus the use of different pronunciation alternatives for these phonemes within the pairs. The authors emphasize that differences in voices can be classified into *organic* and *learned* factors, and that *learned* factors can result from *copying* from people around us, but also from *choosing* in order to mark individuality. This points to the importance of the attitude a subject has towards being a twin, since negative attitudes may trigger an exaggeration of individual (speech) behavior in order to dissociate oneself from a twin. The authors suggest an interesting hypothesis based on their results: the lack of coarticulatory differentiation in twins (in contrast to non-siblings) might be explained by the fact that non-related speakers "satisfy shared segmental **targets** but betray the effects of non-isomorphic vocal tracts in the **transitions** between the targets. […] [T]he traces of those alternative articulations in passing from one target to the next" (Nolan & Oh 1996, p. 47) – and thus the influence of NATURE – can then be found in the transitions of non-related speakers but not in those of monozygotic twins with a (nearly) identical vocal

apparatus. This assumption should be kept in mind, since the present study is going to investigate differences in targets as well as in transitions (cf. Chapter 5).

Another study concerning coarticulatory patterns was conducted by Whiteside & Rixon (2003). Altogether three studies of Whiteside & Rixon focus on the speech patterns of one male monozygotic twin pair (2000, 2001, 2003). The earliest study (2000), which investigated perceived speaker similarity by means of a perception test with the MZ voices, was described above. The other two studies contain acoustic analyses of the twins' voices. Note that these analyses have to be seen as case studies, since the investigations and results are based on only one twin pair. Nevertheless, some interesting findings can be reported. In the study from 2001, Whiteside & Rixon did a comprehensive analysis of 21 acoustic parameters (durational and frequency related) from 160 words and found nine parameters which show significant differences, among them vowel duration parameters, formant patterns (F1 at onset and midpoint, F2 at midpoint of the vowel), and voice onset time. In the later study (2003), Whiteside & Rixon compare the speech of the MZ twin pair with that of their age- and sex-matched sibling who took part in the study two years later. They emphasize the match for demographic factors and weight and height measures, and the shared environmental factors like the family and school (= NURTURE), but there is no information about the attitude the 3 siblings have towards each other and how much time they spent together. The study investigates read speech (various CVC monosyllabic words) and concentrates on coarticulation patterns in terms of F2 vowel onsets and F2 vowel targets in CV sequences. Therefore, locus equations are plotted, which can be seen as phonetic descriptors of the place of articulation (Sussman et al. 1991, 1992) and describe the linear relationship of F2 vowel onset and F2 vowel target. The steepness and slope of these regression lines are compared over the phonetic contexts (i.e. a bilabial, alveolar, velar and glottal consonant /b, d, g, h/) and the three speakers. The method section is very precise and the acoustic and statistical analysis is comprehensive. In addition to variation in locus equations that results from the different phonetic contexts, Whiteside & Rixon emphasize speaker-specific variation and find the twins to be more similar to each other in their coarticulation patterns as expressed in locus equations compared to the age-matched brother of the twins: regarding the different consonants, the order of the steepness of the slope values was glottal > bilabial > velar > alveolar for the twins, which is in line with 18/20 speakers of a study by Sussman et al. (1992).

The order of the slopes of the brother matches with the 2 remaining speakers of Sussman's study: glottal > velar > bilabial > alveolar. The authors explain the more similar patterns for the twins with their greater physical similarity in their vocal tracts compared to their brother. Regarding the current study it should be kept in mind that interestingly F2 vowel onsets (not targets) and coarticulation patterns expressed in the relation between F2 vowel onset and target reveal the highest similarity between the twins in contrast to their brother. As in Nolan & Oh's study, this again indicates the importance of coarticulation and transitions when considering the influence of biology and shared physiology (NATURE).

Further insight into speaker-specific characteristics of twins' speech is given by the comprehensive study by Loakes (2006). Her dissertation project is situated in the field of **forensic speaker identification** and therefore concentrates on finding the best combinations of parameters for discriminating the speech of similar sounding twins. Overall, the author claims that "speaker variation is governed by a speaker's physical dimensions, and shows that the concept of learned variation, or 'choice', also plays a major role" (Loakes 2006, p. 1). It is obvious that it cannot be assumed that all speakers have a completely unique voice, since very often an overlap in the phonetic output can be found. However, she also emphasizes that speaker-specific parameters in speech are evident, since even in identical twin pairs differences in the phonetic output can be found. Loakes investigated speaker-specific characteristics of Australian English in two corpora: a) conversational (spontaneous) speech of four male twin pairs (3 MZ and 1 DZ) aged between 18 and 20, recorded twice over time, and b) telephone recorded speech of five male twin pairs (4 MZ and 1 DZ) aged between 27 and 32. The study contains a **complex acoustic and auditory analysis of several vowels and consonants in varying phonetic contexts**. For statistical analysis ANOVAS were conducted and the data was reanalyzed and checked by means of a likelihood ratio approach. General results show that twins are more difficult to distinguish than unrelated speakers, but the data nevertheless reveal acoustic and auditory differences that can be used in forensic terms. Individual differences depend on the speakers and the parameters under investigation. Regarding the production of vowels, the study reveals that F3 (third formant) turned out to be the most speaker-specific formant and that lax vowels are more speaker-specific than tense vowels (Loakes 2006). In an earlier paper Loakes (2004) states that F2 and F3 of /ɪ/ were found to be the most speaker-specific characteristics in twins' speech (4 MZ and 1 DZ pair)

and are reliable parameters when comparing same- and different-speaker pairs. Regarding the production of consonants, a consistent frication of /k/ and /p/ seems to be a salient auditory and acoustic parameter that can be used even for discriminating twins. The study contributes to the field of inter-speaker variability in terms of determining speaker-specific parameters that can be found even in the speech of twins. Therefore, the speech corpus used for the current study has been developed with her results in mind (cf. Chapter 4 on vowels).

The study of Künzel (2010) is another forensic study that deals with the differentiation of twins (26 female and 9 male German monozygotic twin pairs). The author emphasizes that distinguishing the speech of monozygotic twins is an extremely challenging task since the smallest possible amount of inter-speaker variation due to organic and learned factors is expected. The twins not only share anatomical parameters related to speech production but they are also exposed to the same conditions for socialization by being brought up in the same social environment. Künzel refers to a study by Rosenberg (1973) in which a twin pair was confused in 96% of the cases in a same/different listening experiment while the automatic system was able to distinguish the twins without error. This advantage of an **automatic forensic speaker identification system** (which also takes coarticulatory features such as transitions between neighboring sounds into account) may be explained by signal parameters that can be detected by the system but are not audible to the listener. In the study an automatic system for forensic speaker recognition (BATVOX 3.1) was used to investigate inter-speaker, intra-twin and intra-speaker similarity coefficients or likelihood ratios. Results indicate that their approach succeeded in distinguishing even similar sounding voices of monozygotic twins, but the results vary in terms of individual pairs of speakers as has also been found elsewhere (Johnson & Azara 2000). In addition, the performance of the system was better for male than for female twins. One explanation the author gives is that female voices are generally more difficult to analyze in terms of spectrum-related parameters, since a higher fundamental frequency results in a less dense spacing of the harmonics and eventually in less sound- and speaker-related information in the spectrum (Peterson & Barney 1952). In addition, an analysis of the fundamental frequency of their speakers revealed that the female twins showed very similar mean F0 values (the amount of intra-speaker and intra-twin pair differences was nearly identical) while the male twins differed. However, the study lacks a

detailed explanation of the analysis algorithm of the automatic system BATVOX, and therefore the conclusions should be treated carefully.

To summarize, there are still only few phonetic studies on inter-speaker variability in twins. It has been found that MZ twins are more similar than DZ twins in their acoustic output (e.g. mean F0, voice quality parameters, coarticulatory patterns), but even MZ twins can be distinguished by auditory analysis above chance (in perception tests with familiar listeners) and in acoustic analysis (in formant patterns or by using automatic speaker recognition systems). While these studies can give some insights into the impact of NATURE and NURTURE on inter-speaker variability in speech, they are based only on acoustic investigations. There is a great lack of articulatory studies in this field. I am not aware of any study investigating articulation patterns in the speech of twins. To investigate this missing link and to fill this research gap, the current study includes an articulatory analysis in addition to perceptual and acoustic analyses. One aim of the present investigation is to look for differences in articulation strategies which might explain differences in the acoustic output or which may not even be registered by an acoustic analysis due to motor equivalent strategies. This study also expands this goal from solely seeking speaker-specific parameters to explaining them in terms of genetic/physiological factors (NATURE) and environmental/behavioral factors (NURTURE) by comparing the amount of inter-speaker variability between MZ twins and DZ twins, who differ in their degree of physiological similarity. In addition, the results will be discussed with regard to the role of somatosensory and auditory feedback as well as articulatory and auditory goals in speech.

In the following section I will describe the pilot twin study that was carried out to find suitable speech material for the present investigation and to determine further possible research questions

2.3 Pilot twin study

A pilot study was carried out to establish acoustic differences in the speech parameters of identical (MZ) and non-identical (DZ) twins. This study was intended to help locate phonemes that show acoustic differences within twin pairs and therefore promise to also show differences in articulation, although the relation between acoustics and articulation is not linear. To optimize the probability of finding differences within twin pairs, the speech material should show in general high inter-speaker variability but low intra-speaker variability.

Four identical twin pairs and 1 non-identical twin pair were recorded reading a relatively large corpus (a subset of the subjects described in Chapter 3). In addition they were asked about their attitude towards being a twin and the amount of time they spend together. (cf. Table A.1 in the appendix). The phonemes under investigation were part of a target word (in some cases nonsense words following the phonotactics of German were used where no corresponding words existed). Each word was embedded in the carrier sentence "Ich habe … gesagt" (I said ….). Subjects were seated in a sound-attenuated room and asked to read the different sentences that appeared on a screen in front of them. Each sentence was repeated 5 times in a randomized order. To avoid readers rendering a repetitive pattern, filler sentences were also recorded. In total there were 1300 test sentences for analysis (10 speakers x 26 target words x 5 repetitions). Table 1 gives an overview of the recorded and analyzed speech material.

Altogether, 26 sounds were investigated for each of the 10 speakers. The target sounds were marked off on a sound by sound basis in PRAAT (version 5.1, Boersma & Weenink 2009). The oscillogram and the spectrogram were used to define on- and offsets of vowels and stops. The on- and offsets of the vowels were determined on the basis of changes in formant structures (stable part of F2), and the stops were labeled from the end of the second formant of the preceding vowel to the start of voicing of the following vowel (cf. Figure 1). Four formants (F1-F4) were measured with the help of a PRAAT script[2] in the middle of each

[2] The following settings were used for the measurement of the formants: maximum number of formants = 5, maximum frequency for female = 5500 Hz, for male = 5000 Hz, positive time step = 0.01, window length = 0.0025 s, pre-emphasis from 50 Hz.

vowel. For plosives two parameters were analyzed: Voice Onset Time (VOT) for each plosive, and Voicing During Closure (VDC) for the voiced stops. To compensate for differences in speech rate, VOT and VDC were calculated as a percentage of the word or phoneme respectively: VOT (%) = length of VOT/length of word, VDC (%) = length of VDC/length of stop closure.

Table 1: Speech material and analyzed parameters of the pilot test.

	Phoneme	**Target Word**	**Analysis**
Vowels	/a:/, /a/	M**a**ße, M**a**sse	Formants: F1-F4
(long – short)	/i:/, /ɪ/	M**ie**te, M**i**tte	
	/e:/, /ɛ/	D**e**ko, D**e**cke	
	/u:/, /ʊ/	K**u**hle, K**u**lle	
	/y:/, /ʏ/	H**ü**te, H**ü**tte	
	/o:/	W**o**ge	
Plosives	/p/, /b/	**P**asse, **B**asse	Voice Onset Time (VOT) & Voicing During Closure (VDC)
(voiceless – voiced)	/t/, /d/	**T**asse, **D**asse	
pretonic	/k/, /g/	**K**asse, **G**asse	
posttonic	/p/, /b/	Ma**pp**e, Ma**bb**e	
	/t/, /d/	Ma**tt**e, Ma**dd**e	
	/k/, /g/	Ma**ck**e, Ma**gg**e	
Fricatives	/s/	Ta**ss**e	Long Term Average Spectra
posttonic	/ʃ/	Ta**sch**e	
pretonic	/ʃ/	**Sch**ule	

Figure 1 gives an example of a segmentation of the voiced stop /b/ in word-initial position. The closure time of the stop is separated into a voiced part and a voiceless part. The time between the burst and the beginning of voicing for the following vowel is marked as VOT (Pompino-Marschall 2003). Thus, VDC (%) = length of voiced/length of voiced + voiceless, and VOT (%) = length of VOT/length of <basse>.

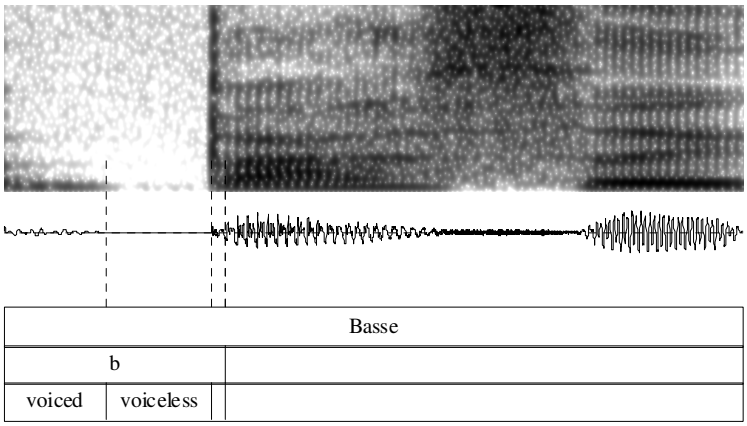


Figure 1: Spectrogram and oscillogram of <basse> with a labeled /b/ and its voiced and voiceless parts.

T-tests were calculated for each twin pair and also for each unrelated sex-matched speaker pair for the parameters F1-F4, VOT and VDC. Long-term average spectra of the fricatives were analyzed for each speaker and compared within the pairs.

The general results for vowels concerning all twin pairs were as follows: the vowels /ʊ/ and /a/ turned out to be the most speaker-specific within all twin pairs. The number of significant differences within all pairs in F1-F4 for each vowel was counted and a chi-square test showed a significant influence of place of articulation ($X^2 = 4{,}879$, df = 1, $p < .005$): the central and back vowels [a, aː, ʊ, uː, oː] revealed more significant differences than the front vowels [ɪ, iː, ɛ, eː, ʏ, yː]. The formant with the most differences among F1 through F4 was F3, followed by F1 and F4. F2 showed the least variation within all twin pairs. Comparing speakers within and across twin pairs, a clear discrepancy in the amount of differences in the formant patterns could be found: for the unrelated pairs the first three formants showed a relatively stable

probability of over 50% of showing differences and F4 indicated a somewhat smaller value (45%), whereas the twins revealed an average probability of only 28%.

The results comparing the MZ twin pairs with the DZ twin pair showed that the DZ pair reveals a higher probability of showing differences only in F3. The differences between inter-speaker variability within DZ and MZ pairs were largest in F1. Interestingly, here the MZ pairs showed a higher probability of showing differences. In general, these results point to the importance of a shared environment over physiological identity and support the hypothesis of auditory goals as targets in speech production regarding the acoustics of vowels.

Table 2: Significant differences in Voicing During Closure (VDC) and Voice Onset Time (VOT) within the twin pairs ($p < .05$).

Twin	MZm1	MZm2	MZf1	MZf2	DZf
VDC in a stressed syllable	Quality of acoustic recordings not sufficient	Basse Dasse Gasse Tasse Kasse	–	Dasse Gasse	Dasse Gasse
in an unstressed syllable		–	–	–	–
Amount of differences in %	x	41%	0%	17%	17%
VOT in a stressed syllable		Passe Tasse Kasse	Passe Tasse	Basse Dasse Tasse Kasse	Basse
in an unstressed syllable	Mappe Matte Macke	Mappe Matte Macke	–	–	Macke
Amount of differences in %	25%	50%	17%	33%	17%

The results regarding inter-speaker variability in plosives are shown in Table 2. Significant differences ($p < .05$) between speakers are reported for each twin pair. The five twin pairs are coded with the following information: MZ or DZ, m (male) or f (female), and a running number (1 or 2). Altogether, 12 sounds were investigated for each of the parameters VDC and VOT. To quantify the amount of differences and to compare them between the twins, the probability of showing differences in each parameter was calculated for each pair (number of differences/12). In general, more differences were found regarding VOT than VDC for the MZ pairs, although not for the DZ pair. This could indicate a stronger influence of physiology on source-based parameters like VDC than on VOT. Furthermore, an impact of the factor *stress* on inter-speaker variability in plosives could be found: it was more likely to find differences in VOT (voice onset time) and VDC (voicing during closure) within all twin pairs, but especially in MZ pairs in *stressed* rather than in *unstressed* syllables. Since unstressed syllables are perceptually less prominent (McAllister 1991, van Bergem 1993) they might be more variable, and thus they do not show significant differences. The stressed syllables are less variable and the learned auditory goals are more crucial here. Another explanation could be that physiology has a greater influence on speech production in *unstressed* than in *stressed* syllables. Of course, the results of the pilot study have to be treated very carefully, since the DZ twins were represented by only one pair.

To summarize, for vowels a clear difference between unrelated speakers and twins was found regarding the amount of inter-speaker variability in F1-F4. Thus, shared NATURE and/or NURTURE must have an effect. However, no clear distinction could be made between MZ and DZ twins. These results point to a superior role of NURTURE over NATURE concerning the acoustics of the analyzed stressed vowels. Note that, if a lack of difference between MZ and DZ twins came about because of a high degree of variability between speakers in producing stressed vowels in general, no difference between twins and unrelated speakers would have been found. Nevertheless, it could be hypothesized that the DZ twins, who differ in their physiology, also differ in their articulatory behavior leading the same acoustic outputs. This will be investigated further in the present study (Chapter 4). For stops, a greater impact of NATURE on the source-based parameter VDC than on VOT was found. In addtion, NATURE seems to influence unstressed syllables more than stressed syllables.

Thus, based on the results of the pilot study a stronger impact of NATURE on consonants than on vowels is hypothesized for the following investigation. Furthermore, since the factor *stress* turns out to be a significant interacting parameter, it is suggested that the impact of NATURE can be intensified by the factor *stress:* namely, a difference in the amount of inter-speaker variability between MZ and DZ is assumed in unstressed syllables but not in stressed syllables. This issue will be examined in one of the research questions in the investigation of vowels in Chapter 4.

2.4 Summary and outline of the study

Intra- and inter-speaker variability is a characteristic of speech. In Chapter 1 it has been shown that NATURE and NURTURE have explanatory power in finding reasons for token-to-token variability and speaker-specific characteristics in speech. However, from the discussion above it can also be seen that differences might exist in the relevant impact of NATURE and NURTURE on articulatory and acoustic variability. Additionally, it has been illustrated that auditory and somatosensory feedback are crucial in the speech production process. The representations of speech are considered to be multimodal, i.e. both acoustic and articulatory. However, a hierarchy might exist in the influence of these two modalities with respect to the phoneme category (and thus the production strategy). Sounds like *sibilants*, with a great deal of linguo-palatal contact – and hence a greater impact of somatosensory than of auditory feedback – might be more influenced by physiological constraints (NATURE) than *vowels*.

With respect to the present study it should be kept in mind that the investigated speaker groups (i.e. monozygotic twins and dizygotic twins) differ with respect to NATURE but not NURTURE. Hence, the different degrees of these two impact factors can be investigated and described by analyzing the amount of articulatory and acoustic inter-speaker variability within the different types of twins. In other words, if MZ twins are more similar than DZ twins in a certain parameter, NATURE would seem to be an important influencing factor.

From the results of the pilot study it is obvious that the factor *stress* can also affect the power of the influencing factor NATURE. Thus, a hypothesis discussed in the following investigation is that stressed syllables, which correspond to learned auditory goals, are less

influenced by physiology than unstressed syllables, which are more influenced by the coarticulation process.

In addition, coarticulation patterns or *transitions* might be more affected by speaker-specific physiological differences (NATURE) than *targets*. While speakers share the relevant linguistic targets, they reveal the speaker-specific patterns of their vocal tract through differences in the transitions between these targets (i.e. how a speaker gets from target A to target B). These individual coarticulation strategies might be reflected both in the articulation and in the acoustic domain.

Thus, the issue to be discussed in the present study is the impact of NATURE and NURTURE on inter-speaker variability. However, the impact of NATURE might be intensified by several additional parameters. Two of the main research questions discussed in the present study have been stated above (question 1 and 2). In addition a third research question will be analyzed (question 3):

1) Is there a difference in the influence of NATURE on inter-speaker variability depending on the **phoneme category** (and thus the amount of linguo-palatal contact)?

2) Is there a difference in the influence of NATURE on inter-speaker variability depending on the particular characteristics of the analyzed parameter: **target (static) vs. transition (dynamic)**?

3) Is there a difference in the influence of NATURE on inter-speaker variability depending on **stress** (vowel in a stressed syllable vs. an unstressed syllable)?

The structure of the remainder of this thesis is as follows:

Chapter 3 gives an overview of the methodology of this study. It includes a description of the speech corpus and of the subjects who participated in the following analyses. The experimental setups of the acoustic and articulatory recordings are explained. Furthermore, the acoustic and articulatory analyses concerning the following investigations are described.

Chapters 4 to 7 present the results of the conducted analyses concerning articulatory and acoustic inter-speaker variability in twins' speech. With respect to the analyses that are carried out in this study, the distinction between articulatory and acoustic *targets,* the character of articulatory *gestures* and acoustic *transitions*, and the articulatory and acoustic realization of *phoneme contrasts* should be kept in mind. It is assumed that the investigated factors NATURE and NURTURE differ in their relative influence on inter-speaker variability between *targets, transitions, gestures* and *phoneme contrasts*.

The vowel chapter (Chapter 4) includes an articulatory and acoustic analysis of the *vowel targets* /a, i:, u:/. Articulatory target positions of the tongue (in horizontal and vertical dimensions) and the shape of the tongue are investigated and compared within the twin pairs. The acoustic analysis comprises an investigation of the acoustic targets in terms of the formants F1 to F4. In addition, inter-speaker variability in the acoustic realization of the *vowel space* between /a/, /i:/ and /u:/ as defined by F1 and F2 is described. Furthermore, the factor *stress* (/i:, i/ in a pretonic vs. posttonic syllable) and the factor *consonant context* (/i:/ following a velar stop vs. a liquid) are taken into account as possible influencing factors on the impact of NATURE on inter-speaker variability.

In the sibilant chapter (Chapter 5) the results regarding inter-speaker variability in articulatory and acoustic *targets* of /s/ and /ʃ/ are presented. Furthermore, the articulatory and acoustic realization of the *phoneme contrast* between /s/ and /ʃ/ is investigated within the twin pairs. A third focus lies on coarticulatory patterns, i.e. the analysis and comparison of vowel-sibilant *transitions* (in terms of F2 and F3 transitions) between the speakers.

In the articulatory analysis of Chapter 6 the looping movement of the tongue during VCV sequences (with C being a velar consonant) is investigated. Inter-speaker variability in the size and shape of the articulatory *gesture* is investigated within the same speaker, in MZ pairs, in DZ pairs and in unrelated speakers.

Chapter 7 presents the results of a perception experiment examining the auditory similarity of the female MZ and DZ twins. Additionally, an acoustic analysis of the stimuli used explores the reasons for the differences in perceived auditory similarity between the twin pairs.

3 METHODOLOGY

3.1 Subjects

Three male and four female twin pairs between 20 and 34 years participated in this study. The genetic similarity (zygosity) of these twin pairs was determined by a genetic laboratory through a genotypic comparison based upon 16 different genetic markers. Monozygotic twin pairs are 100% genetically identical. If a twin pair differs in any of the 16 DNA markers, they must be dizygotic. When a reasonable number of markers (here 16) reveals no differences, it can be concluded that the twin pair is monozygotic (Spinath 2005). Hence, our subjects can be divided into two groups concerning genetic identity and thus the factor *biology* or NATURE: (1) four monozygotic twin pairs (genetically identical) and (2) three dizygotic twin pairs (genetically non-identical). Group (1) consists of two male and two female twin pairs, whereas group (2) consists of one male and two female twin pairs.[3] All twin pairs except one were born, raised, and are still living in Berlin, Germany. One male monozygotic pair was born and raised in Saxony and is now living apart from each other in different cities. Moreover, one of these siblings has moved to Trondheim, Norway, and had lived there for two years before the recordings for this study took place, while his brother is still living in Saxony. The siblings see each other only 3-4 times per year, but keep in contact through emails and telephone calls at least once per month. The twins that are living in Berlin also differ with respect to the amount of time they have recently spent together. All of the three dizygotic twin pairs are still living together and have not been apart from each other longer than a few weeks. Of the monozygotic twin pairs, one female pair is also still living together, and the other female pair is living next door to each other and sees each other nearly every day; the second male pair is living separately from each other and sees each other twice a month. The time the twins spend together can be considered an additionally factor that might

[3] A total number of 8 twin pairs (4 DZ and 4 MZ) were planned, but due to problems with the electromagnetic artiuculograph only 3 instead of the 4 DZ pairs could be recorded.

influence inter-speaker variability since the mutual influence of the twins and their shared social environment may play a role in shaping auditory goals. Thus, concerning the factor *shared social environment* or NURTURE our subjects can be divided into two groups: (1) seeing each other every day (all three dizygotic and two monozygotic twin pairs) and (2) living apart with different degrees of seeing each other (two monozygotic twin pairs). With regard to both factors NATURE and NURTURE we can divide our subjects into three groups: (A) genetically identical and shared social environment: two monozygotic pairs, (B) genetically identical and different social environments: two monozygotic pairs, and (C) genetically non-identical and shared social environment: three dizygotic pairs. This is summarized in Table 3 below. However, it should be noted that the most important period regarding speech acquisition was spent together for all pairs. All of the pairs grew up in the same social environment and lived together with their families until they were at least 18 years old. Thus, only the recent past is taken into consideration when we separate the twins into two groups based on different amounts of shared environment. Note that the factor *sex* could not be controlled for since the monozygotic pairs that live apart (group B) are also the two male pairs and the monozygotic pairs that share their environment (group A) are both female. This confinement has to be kept in mind when interpreting and evaluating the results, but since there is no reason to assume that gender has an impact on inter-speaker variability resulting from the NATURE-NURTURE issue, this restriction seems to be a minor problem.

3.1.1 Attitudinal and physical parameters

Another important influencing factor to control for is the attitude towards being a twin and the attitude towards the sibling. Studies in social psychology dealing with speech accommodation and mimickry have shown that a positive correlation between the degree of accommodation and liking exists (see for instance Chartrand & Bargh 1999 for social interaction and Aguilar et al., under review, for speech accommodation).

If a person has a negative attitude towards being a twin, he or she might be more likely to separate himself/herself from his/her sibling and assert an individual style, which might also be mirrored in the speaking style and speech characteristics in general. Therefore, separate interviews were conducted with each subject. When asked, all subjects tended to like being a

twin and saw more advantages than disadvantages to being a twin. The questionnaire used in the interviews is shown in the appendix (Table A.2). To quantify their statements, the twins were asked to make ratings on a 5-point Likert scale from 1 ("I don't like being a twin") to 5 ("I very much like being a twin"). Number 3 served as a neutral position with no positive or negative attitude towards being a twin. All subjects showed a strong positive attitude towards being a twin (only ratings of 4 or 5) and only one pair diverged on this point. Thus, the factor *attitude towards being a twin* can be neglected since it should not influence the results, as the pairs reveal no significant differences in their feelings towards being a twin. An overview of the characteristics of our subjects and the discussed influencing factors: *genetic identity* (zygosity), *shared environment* (amount of time spent together) and *attitude* (towards being a twin) is given in Table 3. Furthermore, a code name was generated for the different twin pairs that gives information about the genetic identity (MZ or DZ), the gender (m or f) and a running number (1 or 2). This code is given in the second column of Table 3 (*twin)* and will be used for further discussion.

Table 3: Overview of the twin pairs with information about the factors genetic identity, shared environment and attitude towards being a twin.

Subjects	Twin	Sex	Age	Genetic identity	Amount of time spent together	Attitude towards being a twin (1-5)	Group
AF HF	MZf1	f	34	MZ	Nearly every day	5 – 5	A
GS RS	MZf2	f	26	MZ	Live together	5 – 5	A
SL CL	MZm1	m	32	MZ	Twice a month	5 – 5	B
MI MA	MZm2	m	28	MZ	Three times a year	4 – 4	B
LR SR	DZf1	f	20	DZ	Live together	4 – 5	C
MG TG	DZf2	f	20	DZ	Live together	5 – 5	C
FM HM	DZm1	m	21	DZ	Live together	4 – 4	C

The factor *genetic identity* implies anatomical and physiological identity, and hence it is very reasonable to assume that the physiological and biomechanical properties of the vocal apparatus are rather similar in the monozygotic twin pairs and different in dizygotic pairs. To verify this assumption, the weight and height of the twins were determined to find out whether physical similarities vary between the twin types (i.e. whether the MZ pairs are more similar). It is known that a speaker's height correlates with his/her vocal tract length (Fitch & Giedd 1999). The length of the vocal tract influences the formants, and therefore a closer look should be taken at possible differences. The subjects of this study were asked about their height and weight (information given in Table 4). It can be seen that the height varies within the MZ pairs by only 1 cm and within the DZ pairs by only 2-3 cm. Larger differences appear in the weights of the twins: the MZ twin pairs vary by 1-5 kg, while the DZ pairs vary between 1-12 kg. It should be kept in mind that especially DZf2 stands out in terms of a large weight difference.

Table 4: Weight (in kilograms) and height (in meters) characteristics of the subjects.

Twin pair	Height (in m)		Weight (in kg)	
	twin1	twin2	twin1	twin2
MZf1	1.74	1.73	57	62
MZf2	1.64	1.64	49	46
MZm1	1.79	1.78	65	70
MZm2	1.72	1.72	63	62
DZf1	1.68	1.70	54	55
DZf2	1.65	1.68	51	63
DZm2	1.85	1.88	80	72

Thus, greater differences in physical characteristics turned out to be apparent within the DZ pairs while only negligible differences were found within the MZ pairs.

The shape of the palate is another even more relevant physiological factor that has an influence on speech production (Brunner et al. 2009). Results regarding palatal similarity are given in the following section.

3.1.2 Palatal shape: Silicone palate casts

A silicone dental and palate cast was taken to examine the overall shape of the palate and the steepness of the dome more closely. This was done to verify the assumption of identical physiology concerning the vocal apparatus in MZ pairs (in contrast to more variable palatal shapes within the DZ pairs). Environmental influences may have affected the growth of the palate: an intense use of a pacifier or the habit of thumbsucking during childhood can affect the shape of the palate. The varying use of pacifiers between siblings can potentially result in differences in the palatal shape even in MZ twins, and therefore the assumption of a more similar palatal shape in MZ twins than in DZ twins had to be tested and validated. Silicone molds, which are negatives of the palate and the upper teeth, were created with the help of a silicone paste that is used in dental laboratories. Pictures of the silicone molds of the eight female speakers are shown in the following figure.

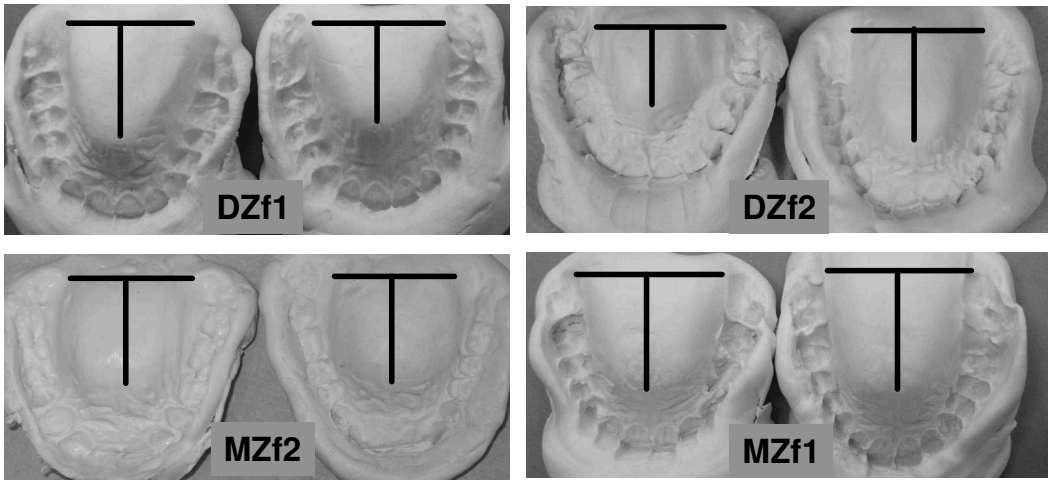


Figure 2: Examples of silicone palatal casts taken of all female subjects; above: DZ pairs, below: MZ pair; lines indicate measurement points (horizontal line = palate width, vertical line = palate length).

In general, it can be seen that the shape and size of the palate is different in the DZ pairs (above) but quite similar in the MZ pairs (below). It has to be noted that it turned out to be quite difficult to get imprints of high quality since the dried paste was hard to separate from the rack rail. Therefore, the imprints were sometimes taken off from the rack rail before they were totally dry. As a consequence, in some cases the imprints were slightly distorted, especially at the alveolar ridge (when the incisors were difficult to separate from the rack rail). This might explain the different rise of the alveolar ridge for MZf1 (pictures on the lower right). While this method of comparing overall palate shape clearly has some drawbacks, it can nevertheless be used for a basic visual comparison of the palate for a given twin pair. The pictures show greater variation in the palatal sizes of the DZ twins than of the MZ twins. For the pictures of the remaining male speakers see appendix (Figure A.1). To objectify the impression given by the pictures and to get a better estimation of the size of the palate of each speaker, the palates were also measured. Table 5 lists the three measurements that were taken from each palate cast: height, width and length of the palate.

Table 5: Measurements of the size of the palate (in cm) and differences within pairs; biggest differences in bold (Δh: difference in height, Δw: difference in width between the 4th molars, Δl: difference in palate length from the midpoint of the vertical line between the 4th molars to the point at which the palate starts to descend).

Twin pair	Height		Width		Length		Δh	Δw	Δl
	Twin1	Twin2	Twin1	Twin2	Twin1	Twin2			
MZm1	2.5	2.4	4.2	4.1	2.3	2.4	0.1	0.1	0.1
MZm2	2.6	2.4	4.4	4.3	2.3	2.0	0.2	0.1	0.3
MZf1	2.1	2.1	4.0	4.1	2.7	2.6	0.0	0.1	0.1
MZf2	1.8	1.8	3.9	4.0	2.8	2.6	0.0	0.1	0.2
DZf1	2.1	2.0	3.8	3.6	2.2	2.8	0.1	0.2	**0.6**
DZf2	2.2	2.5	3.8	3.4	2.5	3.0	**0.3**	**0.4**	0.5
DZm1	2.4	2.6	4.1	4.5	2.3	2.7	0.2	**0.4**	0.4

The height was measured as the distance between the base and highest point of the palatal cast. The width was measured as the horizontal distance between the 4th molars and is illustrated in Figure 2 as a black horizontal line. The length was measured perpendicular to this line from its midpoint to the point where the palate starts to descend (see vertical line in Figure 2). The differences in height, width and length do not exceed 0.3 cm for the MZ twins, while the DZ twins reveal greater differences, especially in the length of the palate: DZf1 and DZf2, show differences of 0.6 cm and 0.5 cm respectively. Also the width differs for the DZ twins: DZf2 and DZm1 vary in the width of the palate by 0.4 cm. The difference in palate height is biggest for DZf2. Hence, from the pictures and also from the measured data in Table 5 it is evident that the DZ twins reveal bigger differences in physiological properties, i.e. palatal sizes, than MZ pairs.

3.2 Articulatory and acoustic recordings

Articulatory and acoustic recordings were conducted to investigate inter-speaker variability within monozygotic and dizygotic twin pairs. During the recordings the subjects sat in a sound-attenuated room. The different stimuli were presented in a random order on a screen that could be seen through a window. The subjects were asked to read the individual tokens as soon as they heard a beep. The beep was used for synchronizing the articulatory and acoustic data.

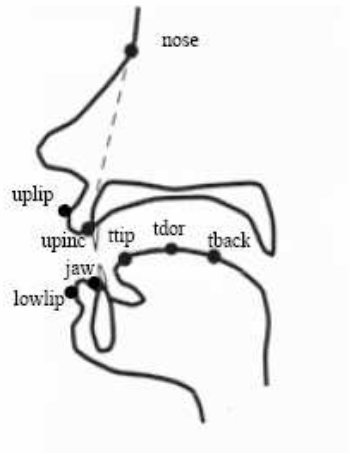


Figure 3: Positioning of the sensors on tongue (tip, dorsum, back), jaw, lower and upper lip, and the reference sensors at the upper incisors and the bridge of the nose.

Articulatory recordings were carried out using a 2-D electromagnetic articulograph[4] (EMA, Carstens AG 100). For the articulatory measurements eight coils in total were attached to the subject's tongue, lips and jaw. Two sensors, one at the upper incisors (upinc) and one at the bridge of the nose (nose), served as reference sensors to compensate for head movements. Three coils were glued midsagittally to the tongue: one approximately 0.5 cm behind the tongue tip (ttip), one approximately 5 cm behind the tip at the tongue back (tback), and a third one at equal distance in-between at the tongue dorsum (tdor). Another sensor was placed below the lower incisors in order to track jaw movements (jaw), and two further sensors were glued to the upper (uplip) and the lower lip (lowlip) to record lip movements (as shown in Figure 3). All receivers were monitored during recording to ensure that they were providing valid signals. For the purposes of the present study, the tongue receivers were of crucial importance. Thus, in the event one of these receivers failed during recording, it was replaced with the one from the upper lip. For the analysis of lip movement, the receivers from the lower lip still provided an assessment of the speaker's labial behavior (for example, for the production of /ʃ/). After the recordings the sensor on the tongue tip was removed and the contour of the palate was recorded by moving this sensor along the palate from back to front. This contour could be used afterwards to compare the shapes of the palates within the twin pairs in more detail (here the slope of the alveolar ridge could be inspected (see Section 3.2.2)). The occlusal plane was recorded by means of a custom made t-bar in order to define a comparable coordinate system: the occlusal plane defines the horizontal (x) axis and the zero-line for the y-axis. This was done by having the speaker bite on a T-piece onto which two sensors had been glued midsagittally. The articulatory data was preprocessed including correction algorithms for head movement, filtering of the data (low pass filter: bandwidth of 18 Hz with a damping of 50 dB at 52 Hz), rotation and translation of the position data, and synchronization with the acoustic data (the supplementary correction program and pre-processing software that was used is described in more detail in Hoole 1996a and 1996b). The sampling frequency of the processed articulatory data was 200 Hz. In addition to the articulatory recordings, the audio tracks were recorded for each speaker via a Sennheiser Mkh

[4] Three transmitter coils that are mounted on a helmet generate an alternating magnetic field at three different frequencies. The x and y coordinates of the sensors (horizontal and vertical positional data) are obtained by converting the distance between the sensors and the transmitters.

20 P48 microphone on one track of a digital audio tape (48 kHz sampling rate) for further acoustic analysis. On the second track the rectangular synchronization impulse heard as a beep by the speakers was recorded. With the help of this synchronization impulse the large audio file containing all sweeps was cut into different smaller wav-files containing only one target sentence. The wav-files were downsampled to 22 kHz sampling rate.

In total, 13 subjects were recorded by means of EMA. Due to problems with the electronic system of the articulograph during the recording of subject 14, the respective twin pair (DZm1) could be included in the articulatory analysis. Thus, articulatory data could be gathered from six twin pairs (4 MZ and 2 DZ), while for the acoustic analyses the data of seven twin pairs (4 MZ and 3 DZ) could be used.

3.2.1 Experimental setup and requirements

As one aim of the study was to compare articulatory movements between speakers with a nearly identical physiology (within the MZ pairs), it was crucial to use the same positions for the coils on the tongue. Therefore, several precautions were taken to get data from coil positions that were as similar as possible.

3.2.1.1 Tongue-coil templates

First, photographs were taken of the tongue with the glued coils on top of it; then, a template was created of the tongue with the coils of one of the twins to be used as a reference for the second twin. The template, with holes at the tongue positions of the first twin, was held on top of the tongue of the second twin and the positions of the coils were marked through the holes of the template. Examples of two templates for a female twin pair (left) and a male twin pair (right) are given in Figure 4. The templates were used for all speakers except the pair DZf2, since the difference in tongue size was too big.

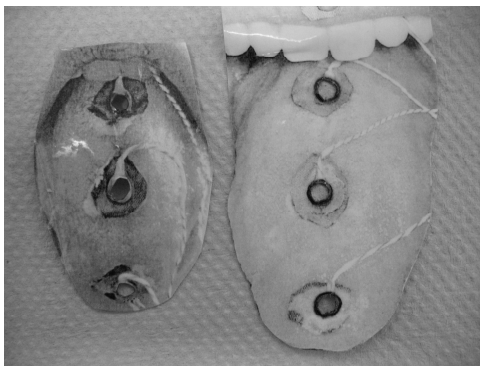


Figure 4: True to scale tongue-coil templates with three tongue coil positions for two different pairs (MZf2, Mzm2).

3.2.1.2 Measured coil distances

During the recordings the distances between the glued coils were measured and an effort was made to apply the same distances to both twins. Table 6 gives an overview of the distances between the glued coils on the tongue. For DZf2 (for whome no template could be used) a noticeable difference in the distances between the glued coils was found: the distance between coil 2 (tongue dorsum) and coil 3 (tongue back) was 2.4 cm for TG but only 1.8 for her sister MG. Hence, as expected, a bigger tongue for speaker TG can be assumed.

Table 6: Distances (in cm) between the three tongue coils measured from front (ttip) to back (coil 3) for each speaker.

Twin 1 – Twin 2	Twin pair	ttip – coil 1		coil 1 – coil 2		coil 2 – coil 3	
		Twin1	Twin2	Twin1	Twin2	Twin1	Twin2
SL CL	MZm1	0.6	0.5	1.8	2.0	2.0	1.9
MI MA	MZm2	0.6	0.65	2.4	2.3	2.0	1.9
AF HF	MZf1	0.4	0.4	2.0	1.9	2.4	2.5
GS RS	MZf2	0.55	0.5	1.6	1.7	1.7	2.0
LR SR	DZf1	0.6	0.5	2.0	1.9	2.1	2.0
MG TG	DZf2	0.6	0.5	2.0	2.0	1.8	2.4

For the monozygotic twin pair MZf2, a difference of 0.3 cm can be seen in the distance between coil 2 and coil 3. Here, the position of coil 3 (tongue back) seems to have a different location for the sisters. The more backward position of coil 3 for RS in comparison to GS has to be kept in mind for the following analysis of the results. For all other pairs the measured distances between the tongue coils match quite well within the pairs.

Additionally, the positions of the coils of the tongue concerning the midsagittal axis were assessed. To this end, the distances between the left and right margin of the tongue and the glued coil were measured. The results are given in Table 7.

Table 7: Measured midsagittal positions of coil 2 and coil: distances to the lateral edges (left and right) of the tongue in cm for each speaker.

		Coil 2				Coil 3			
Twin1 -Twin2	Twin pair	left		right		left		right	
		Twin1	Twin2	Twin1	Twin2	Twin1	Twin2	Twin1	Twin2
SL CL	MZm1	2.4	2.3	2.0	2.0	2.4	2.4	2.0	2.1
MI MA	MZm2	2.0	2.0	1.9	2.0	2.0	2.0	2.0	2.0
AF HF	MZf1	2.0	2.0	2.0	2.0	2.3	2.2	2.0	2.0
GS RS	MZf2	1.8	1.7	1.8	1.7	-	-	-	-
LR SR	DZf1	1.8	1.5	1.8	1.3	1.8	1.8	2.0	1.8
MG TG	DZf2	1.9	2.0	2.0	2.2	2.2	2.4	2	2.5

Here again, the measured distances are quite similar for the monozygotic twins. Both dizygotic twin pairs reveal some differences: speaker LR from DZf1 seems to have a wider tongue than her sister SR, since the distances from coil 2 to the left and right margins are 1.8 cm for LR but only 1.5 cm and 1.3 cm respectively for SR. DZf2 reveals the greatest differences in the width of the tongue back: here, TG again seems to have a bigger tongue, since the distances between coil 3 and the left and right margins are 2.4 and 2.5 respectively, but for her sister MG only 2.2 and 2.0. For both speakers of MZf2 no measurements could

be made for coil 3 since the tongues of these speakers were rather short and small and when the speakers tried to stretch out their tongue as long as possible it got very shaky. Therefore, no distance values for coil 3 are given for these speakers.

In general, as was hypothesized, our measurements concerning the size of the tongue and the positions of the tongue coils revealed a great anatomical similarity within the monozygotic twin pairs, but noticeable differences regarding the anatomy of the dizygotic twins.

3.2.2 Contour of the palatal shape

Since the palatal casts could only be used to get an impression of the overall size of the palates (width, length and height), the recorded contour of the palate was used to check and visualize the shape of the palates (e.g. the slope of the alveolar ridge). Figure 5 shows the adjusted palate contours of the twin pairs (midsagittal tracing, face to the left). The vertical line in each graph marks the highest point of the palate, which was taken as a reference for the adjustments. The horizontal lines under the graphs indicate the lengths of the palates. Hereby, the hypothesis of a more similar palatal shape of monozygotic twins than of dizygotic twins could be supported. The figure reveals the outstanding similarity of the palate contour of both female MZ pairs at the top of the figure. Differences can be seen in the slope of the alveolar ridge between the pairs MZf1 and MZf2, but within the pairs the slopes are identical. At first glance the palate contours of the male MZ pairs show some differences. However, when looking at the size dimensions of the palate, both speakers of MZm1 reveal a remarkably high palate (2 cm) and are identical in the distance from the highest point of the palate to the beginning of the incisors. The speakers of MZm2 are similar in terms of a high but much more slowly rising alveolar ridge.

In contrast, both DZ twins vary in the size and shape of the palate (cf. the different lengths of the lines under the respective figures). Especially DZf1 varies in the distance between the highest point of the palate and the incisors, and LR (grey) reveals a smaller palate than her sister SR (black). Within the pair DZf2, MG (grey) has a smaller palate than her sister TG (black), and the slope of the alveolar ridge is much steeper for MG (grey) than for her sister.

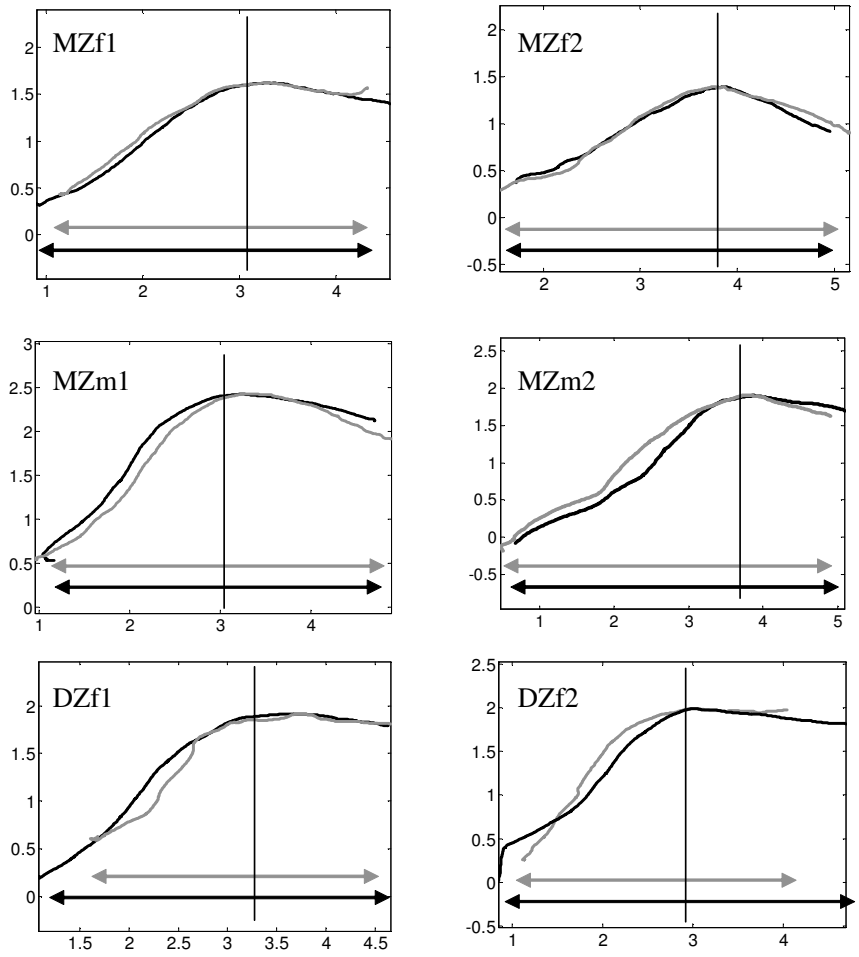


Figure 5: Palatal contours of all speaker;, each twin pair in different subplots, different speakers indicated by different colors (grey = twinA, black = twinB), axis scales in cm.

3.2.3 Adjusting the twins' articulatory data

To compare the articulation between the speakers of one twin pair, the articulatory data had to be adjusted for different sitting and head positions during the recordings. Therefore, the recorded palatal contours were used as a reference pattern to align the articulatory data of the twins. This was done with the help of a MATLAB (version R2007a) script written by Pascal Perrier: the palatal data of speaker 1 was used as a reference for the transformation of the palatal data of speaker 2. The script looks for vertical maxima of the palatal contours and aligns them. Then it looks for the horizontal minima of the translated palate, computes the angle between the two alveolar regions, and rotates the transformed palate around the highest

point to fit the position of the reference palate. Then these computed translation and rotation patterns are used to adjust the articulatory data of speaker 2. In this way the articulation of the two speakers can be compared best. Figure 6 shows the palatal contour and articulatory data of one recorded sentence for two speakers. The green color indicates the data of the reference speaker. The figure on the left side shows the raw data of both speakers (in green and blue); the figure on the right side shows the same data for speaker 1 (still green) but the translated and rotated data of speaker 2 (red). The articulatory data drawn involve the 3 tongue coils, the jaw, and the upper and lower lip.

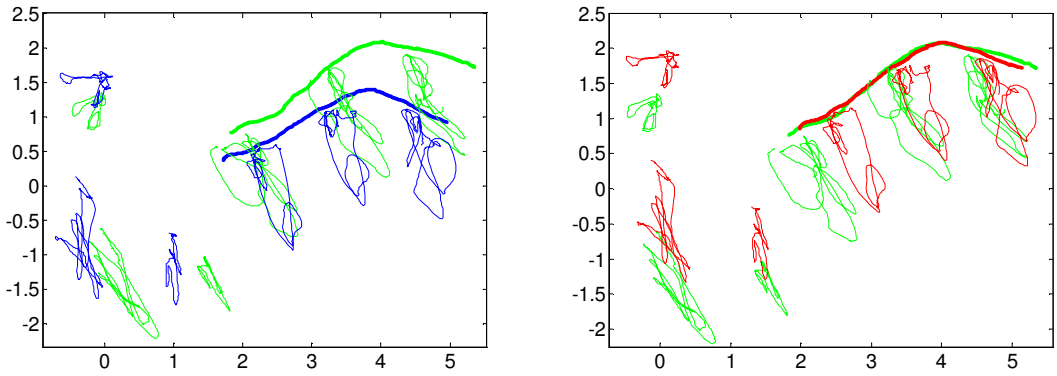


Figure 6: Raw (left) and adjusted (right) articulatory data of MZf2 (green: raw data of GS, blue: raw data of RS, red: adjusted data of RS).

This translation and rotation process was done for each twin pair; all articulatory data is displayed in the appendix (Figures A.2 - A.6). For all further analyses and the comparison of the twins' articulation, only the adjusted data was used. In all respective plots in the following chapters the siblings are indicated by the colors red and blue, and red is always used for the speaker with the rotated and translated data.

3.3 Speech material

The recorded speech material consists of a huge variety of different phonemes including vowels, sibilants, and plosives and was chosen based on studies regarding speaker-specific parameters in speech as well as the results of the pilot study discussed in 2.3. Altogether, the recording session for each speaker lasted approximately 1 hour and was divided into two subrecordings. The first subrecording consisted of 36 different sentences, which were repeated in a randomized order 4-10 times.[5] All possible target phonemes were part of a word (a verb or a name), and the word was inserted in a carrier sentence. The carrier sentence always started with "Ich" (*I*) and ended with "im Garten" (*in the garden*) or "am Montag" (*on Monday*), e.g. "Ich küsse Giba im Garten" (*I kiss Giba in the garden*). The second subrecording consisted of 10 repetitions of five different sentences, of which only one serves as a source for the current analysis, i.e. the investigation of the articulatory gesture /aka/ in the carrier word /kakadu:s/ (*cockatoos*). The target word was part of the sentence "Gestern sah ich bei Peter Kakadus und andere Vögel" (*Yesterday I saw at Peter's cockatoos and other birds*). The speech material under investigation is 1) the cardinal vowels /a, i:, u:/ (cf. Chapter 4) 2) the sibilants /s, ʃ/ (cf. Chapter 5), 3) the sequence /aka/ (cf. Chapter 6), and 4) fundamental frequency and voice quality parameters of the word 'wasche' (cf. Chapter 7). The investigated sounds, the respective carrier words and their phonological transcriptions are shown in Table 8. Altogether, seven phonemes plus the sequence /aka/ were investigated. Different numbers of repetitions were obtained of the various stimuli (ranging from 10 repetitions of the sequence /aka/ to 44 repetitions of the vowels /i:/ in 'liebe'). Altogether, 234 x 14 speakers = 3276 stimuli were recorded and served as speech material for this investigation. However, some repetitions had to be removed in the respective articulatory and/or acoustic analyses either due to corrupted data and measurement errors (especially of the articulatory recording of the electromagnetic articulograph) or due to possible coarticulatory effects of different phoneme contexts (see Section 5.2.6). A more detailed discussion of the number of analyzed stimuli for each of the different phonemes and each speaker is given in the relevant chapters.

[5] An overview of all recorded sentences is given in the appendix (cf. Table A.3).

Table 8: Speech material (investigated phonemes, carrier words and phonological transcription), analyzed target phonemes in bold.

Investigated phonemes (number of repetitions)	Carrier word	Phonological transcription
/a/ (40)	…wasche…	/ˈv**a**ʃə/
/u:/ (40)	…suche…	/ˈz**u**:xə/
/i:/ (44)	…liebe…	/ˈl**i**:bə/
/i/ (10)	…Hagi…	/ˈha:g**i**/
/i:/ (10)	…Giba…	/ˈg**i**:ba/
/s/(40)	…küsse…	/ˈkʏ**s**ə/
/ʃ/(40)	….wasche…	/ˈva**ʃ**ə/
/aka/ (10)	… Kakadus…	/ˈk**aka**du:s/

3.4 Acoustic analyses and labeling

Basically, the acoustic analyses can be divided into two parts. First, acoustic *targets* for particular phonemes (vowels, Chapter 4, and sibilants, Chapter 5) were analyzed and compared within the twin pairs; second, acoustic *transitions* between targets (formant transitions in the sequence /ʃə/) were investigated. In addition, an analysis of fundamental frequency measures and voice quality parameters was conducted for the stimulus/ˈvaʃə/, which is used in Chapter 7 regarding perceived similarity and acoustic correlates. A detailed description of the analyses and the compared parameters can be found in the relevant chapters.

The data of the acoustic recordings was cut into different audio files containing only one target sentence each. This was done for all 14 speakers. The chosen phonemes in the target words mentioned in Table 8 were then segmented and annotated manually with the help of textgrids in PRAAT (version 5.1, Boersma & Weenink 2009). Figure 7 gives an example of the labeled target word /vaʃə/ in one tier and the respective vowel /a/ and sibilant /ʃ/ in the second tier.

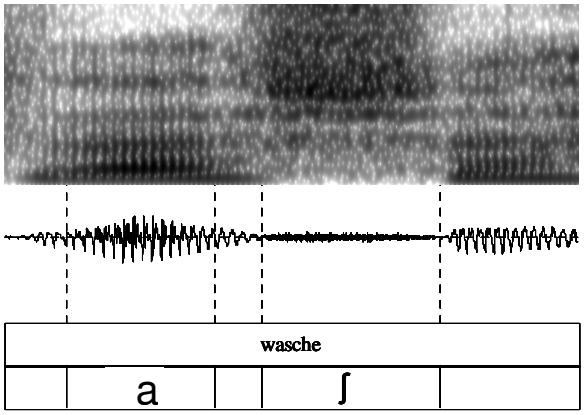


Figure 7: Spectrogram and oscillogram of a segmented and annotated vowel and sibilant in PRAAT.

As criteria for segmentation the on- and offset of the second formant were chosen for the vowels, and the formants were measured in the temporal midpoint of the label. The beginning and end of the fricatives were determined by the on- and offset of the frication period. Thus, as indicated in Figure 7, the end point of the vowel and the starting point of the following sibilant do not have to coincide. Additionally, segment boundaries were placed on the zero-crossings in the oscillogram.

3.5 Articulatory analyses

Corresponding to the acoustic analysis, two kinds of articulatory measurements were made to compare the articulation within the twin pairs. First, articulatory *target* positions for particular phonemes were determined (3.5.2), and second, articulatory *gestures* and their corresponding kinematic parameters (amplitude, velocity) were investigated (3.5.3). All measurements and analyses of the articulatory data were conducted with the help of *artmat*, a graphical interface in MATLAB 7.4.0 written by Christine Mooshammer. Before the articulatory analysis could be done, a closer look had to be taken at the reliability of the articulatory data. The articulatory recording is very sensitive to measurement errors due to different influencing factors: technical reasons (EMA-inherent difficulties, deviation from the midsagittal plane) and subject-related factors (different speech styles, articulatory movements) can play a role.

3.5.1 Reliability of the articulatory data

For maximum accuracy of the articulatory measurements it is important that the main axes of the transmitter and sensor coils are aligned in parallel. Misalignment – i.e. displacement of the sensors from the midline and rotation – results in measurement errors. Therefore, besides the positional data of each articulatory record, the amount and variability of rotational misalignment can be analyzed: a correctness factor (tilt x)[6] and its mean variability (SD) are provided for each coil by the EMA software. These values are inspected more closely for each speaker and coil to get an estimation of the reliability of the articulatory data. The higher the correctness factor, the more reliable is the measurement due to less displacement of the coils from the midline. The mean variability gives an estimation of how much the displacement of the coil varies over one articulatory record. Since two different recordings were made due to the varying speech material (cf. Section 3.6.), two correctness factors have to be investigated for each speaker. Table 9 gives an overview of the correctness values and their mean variability for all speakers and coils of the first recording. For some speakers the data for the upper lip coil is missing due to a failure of a receiver during recording. Thus, positional data of the upper lip is not investigated further. The provided values are calculated over the articulatory data for the coils that are investigated in the following analyses.

In general the correctness factors are very high and above 90% for all speakers and the lip coils, the tongue tip and the jaw. Speaker MA reveals a correctness factor of only 80.18% for the tongue back coil and 88.82% for the tongue dorsum coil. Here, the mean variability also shows a quite high value of 2.81 and 2.32 respectively. This has to be kept in mind for the further analysis of the tongue dorsum and back cursor for this speaker: for the investigation of the vowel targets, the tongue dorsum coil (for /i/) and tongue back coil (for /a/ and /u/) are crucial, thus the twin pair MZm2 (with speakers MA and MI) had to be excluded for the articulatory analysis of the vowels.

---

[6] Formula: correctness factor = corrected radius/radius (see Mooshammer 1998)

Table 9: Correctness factors (tilt x) and mean variability (SD) for all speakers and 6 coils for the first recording session; missing data (-) for upper lip coil (2 speakers) due to a broken coil.

Speaker	Upper Lip tilt x (SD)	Tongue Back tilt x (SD)	Tongue Dorsum tilt x (SD)	Tongue Tip tilt x (SD)	Jaw tilt x (SD)	Lower Lip tilt x (SD)
AF	-	98.84 (0.49)	97.48 (0.99)	96.84 (0.71)	99.35 (0.18)	96.35 (0.63)
HF	98.19 (0.36)	97.55 (0.95)	96.71 (0.85)	95.66 (1.13)	98.00 (0.01)	97.90 (0.53)
GS	98.33 (0.41)	94.71 (1.55)	98.94 (0.76)	96.69 (1.24)	99.38 (0.45)	97.58 (0.50)
RS	98.06 (0.21)	95.93 (1.44)	97.96 (0.49)	94.41 (1.27)	100.43 (0.07)	98.11 (0.28)
CL	98.06 (0.17)	100.31 (1.69)	98.35 (0.99)	104.12 (0.74)	100.13 (0.25)	97.86 (0.38)
SL	99.04 (0.19)	98.69 (0.69)	99.90 (0.97)	98.20 (0.66)	98.83 (0.15)	98.65 (0.47)
MI	98.05 (0.12)	93.00 (0.89)	95.65 (1.04)	94.43 (1.76)	91.18 (0.35)	92.66 (0.67)
MA	-	80.18 (2.81)	88.82 (2.32)	98.60 (1.57)	98.90 (0.28)	96.87 (0.40)
LR	97.40 (0.56)	99.46 (0.98)	92.66 (1.52)	98.08 (1.42)	96.66 (0.28)	98.35 (0.44)
SR	95.23 (0.67)	99.48 (1.17)	97.06 (0.84)	100.26 (1.48)	96.14 (0.10)	97.13 (0.41)
MG	98.46 (0.42)	96.78 (2.01)	93.65 (2.30)	97.20 (1.68)	95.74 (0.30)	98.76 (0.33)
TG	96.91 (0.37)	99.86 (2.05)	95.82 (1.38)	97.47 (0.68)	98.98 (0.03)	97.06 (0.44)

For the second recording correctness factors were inspected for all coils and all speakers with a particular focus on the tongue back and tongue dorsum coil since they are crucial for analyzing looping patterns in /aka/. Again speaker MA turned out to have very low reliability scores of the articulatory measurements. The high error scores reveal that the position of the tongue was not measured very precisely. This could be due to the fact that the speaker twists the tongue in such a way that the tongue coils deviate from the midsagittal plane. Figure 8 shows the correctness factors (tilt values) for the tongue back (tba) and the tongue dorsum (tdo) for all articulatory data of the second recording; the renditions of speakers MA and MI

are plotted in red, and MA is additionally marked by a circle. The data of speaker MA reveal correctness factors below 92% for the tongue dorsum and below 82% for the tongue back, while the other speakers show values between 88-99% for tdo and 89-100% for tba. Therefore, speaker MA, and thus again the twin pair MZm2, was excluded for the analysis of the tongue back movement during the articulatory gesture /aka/.

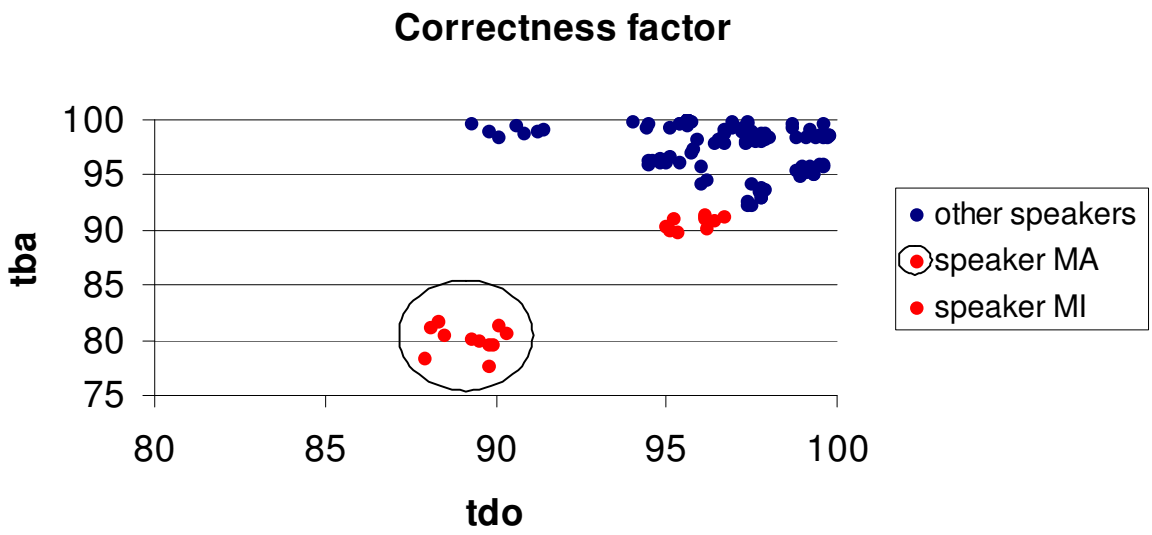


Figure 8: Mean correctness factors of tongue back and tongue dorsum coils for all speakers revealing the validity of the articulatory measurements.

3.5.2 Articulatory labeling of TARGET positions

Articulatory target positions were measured for each speaker and vowel. The articulatory target position is defined as the point in time (and hence the position of an articulator) when the tongue or the jaw has reached a certain extreme position or maximum after which point the movement direction changes. Additionally, the tangential velocity of this coil should be minimal at this time point. For each phoneme a particular articulator was chosen to define the achievement of this articulatory target. It was assumed that the target position had been reached when the velocity of this particular articulator was minimal. The chosen parameters can be found in the following table.

Table 10: Parameters defining articulatory targets.

Phoneme	Parameter
/a/	Lowest vertical position and minimum in velocity of the jaw
/i:/	Highest vertical position and minimum in velocity of the tongue dorsum
/u:/	Lowest horizontal position (= maximal protrusion) and minimum in velocity of lower lip
/s/	Highest vertical position of the jaw and the tongue tip
/ʃ/	Highest vertical position of the jaw and the tongue tip

Articulatory target positions of /a/ were reliably determined, and the point of maximal jaw opening was in most cases congruent with the lowest horizontal position of all tongue coils. For the realizations of the vowel /i:/ the tongue dorsum coil was significant, but less distinct, as the tongue was alredy moving upwards because of the preceding /l/. Still, target positions for /i:/ could be determined easily in most cases. Ascertaining the articulatory target position of /u:/ in the target word /ˈzu:xə/ proved to be the most difficult determination. Often, no minimum in the velocity of the tongue coils could be found, therefore the upper and lower lip and the acoustic signal in the oscillogram were also taken into account. The defined target positions for /s/ and /ʃ/ could for the most part reliably be determined since the jaw has to reach a maximum position to create an obstacle necessary for the production of frication noise. To cross-check, the position of the tongue tip coil was also taken into account. All measurements were carried out manually. The determined time points with the corresponding positions of the coils were saved.

3.5.3 Articulatory labeling of GESTURES

As a second measurement, the articulatory *gesture* /aka/ was investigated. More precisely, the horizontal and vertical movement of the tongue back during the sequence /aka/ was analyzed. The whole looping gesture consists of a closing gesture (from /a/ to /k/) and an opening gesture (from /k/ to /a/) of the tongue back. The start of the looping pattern is marked by the beginning of the closing gesture from the first /a/-target to the velar closure of

the /k/. The reaching of the /a/-target is determined by the lowest vertical position of the tongue back and thus the minimal tangential velocity of the tongue back. The end point of the gesture is marked by the end of the opening gesture from /k/ to the second /a/. Again the minimal tangential velocity of the tongue back, and thereby the lowest vertical position of the tongue back, determines the reaching of the /a/-target. Figure 9 gives an example of a labeled gesture: start and end of the gesture are marked by vertical lines and light asterisks on the x-axis (time in s). The oscillogram, the articulatory data of the tongue back in horizontal (tbackX) and vertical (tbackY) direction, and the corresponding tangential velocity of the tongue back coil (tbackTV) are shown.

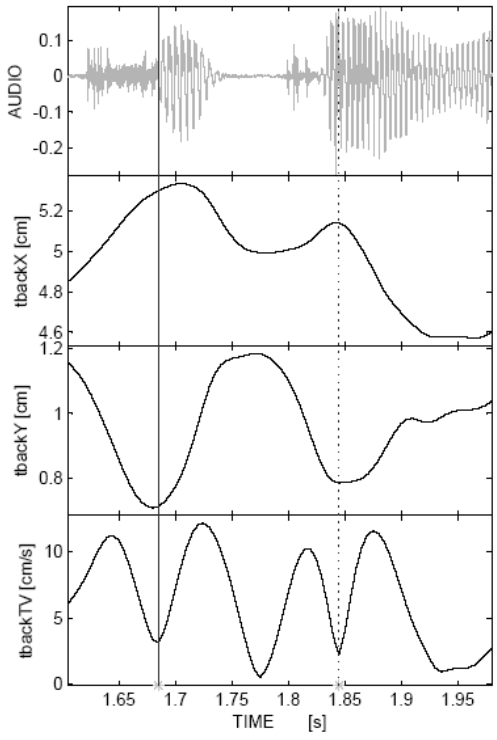


Figure 9: Oscillogram and articulatory movement of the tongue back during the sequence /aka/ (start and end of the sequence are marked by vertical lines), horizontal and vertical movement of the tongue (tback X, tbackY in cm) and the tangential velocity of the tongue back (tbackTV, in cm/s).

With the following chapter the result section of the present study starts. Here, speaker-specific patterns are investigated in twins' speech, and the role of the two influence factors NATURE and NURTURE on inter-speaker variability is evaluated. First, the analysis of the vowels will be discussed.

4 INTER-SPEAKER VARIABILITY IN VOWELS

4.1 Articulatory inter-speaker variability in vowels

In the current chapter the possible influence of NATURE (and in particular individual anatomical and physiological characteristics) on the articulation of vowels will be investigated. In speech research it is discussed how the speech signals we are able to produce and perceive are limited by our physiology, i.e. NATURE (see Section 1.2, Lindblom 1984; for an overview see Fuchs et al. 2007). In general, consonants are described as being more influenced by anatomical and physiological constrictions than vowels, due to their tongue-palate contact patterns and (in accordance with that) their anatomical restrictions. From the discussion in Chapter 1 it is evident that somatosensory feedback is more important in consonants than in vowels, and thus a stronger influence of NATURE (palatal shape, tongue size and physiology, vocal tract geometry) on consonants than on vowels is assumed. However, several studies have also shown a) that physiology can be a reason for differences in intra-speaker variability between different vowels and b) that individual physiological characteristics can lead to inter-speaker variability.

Mooshammer et al. (2004) and Brunner et al. (2005) have found in their production studies on vowels that articulatory variability is reduced when the amount of linguo-palatal contact is high, suggesting that articulatory variability in vowels is constrained by the interaction of the tongue and vocal tract boundaries (e.g. in the production of high vowels). In an earlier study, Perkell & Nelson (1985) investigated the existence of a physiological "saturation effect" in the production of /i/ and /a/ in American English.[7] The authors found that the articulatory variability is lowest in the direction perpendicular to the vocal tract midline for both vowels. For /i/ the authors suggest "that the sides of the tongue blade are being pushed against and

[7] But see Buchaillard et al. (2008) for a critical discussion of the saturation effect hypothesis.

restrained by the hard palate" (p. 1894). This points to biological restrictions in speech production. The influence of anatomical properties (NATURE), in particular the shape of the palate, on intra- and inter-speaker variability was investigated in a study by Brunner et al. (2009). The authors assume that speakers with a flat palate are more constrained in their articulatory variability, since slight variation in the tongue position has a larger impact on the area function and hence on the acoustics than in speakers with a dome-shaped palate. They investigated articulatory and acoustic variability in the production of vowels in 32 speakers of seven different languages (e.g. English, German, Polish, Bulgarian, and Norwegian) by means of electropalatography (EPG). Their results support the hypothesis of a relation between vertical variability and palate shape. Speakers with flat palates showed reduced articulatory variability. For speakers with dome-shaped palates the degree of articulatory variability was more irregular: some revealed a significantly larger amount than speakers with flat palates, but this was not the case for all speakers. No effect of palate shape on acoustic variability was found, since speakers with flat palates did not differ in their acoustic variability from speakers with dome-shaped palates. The authors suggest that speakers adapt their articulatory variability to their individual morphology. By this means, speakers make sure that their acoustic variability remains within a compatible range for perception. Similar findings were reported previously by Perkell (1997) and Mooshammer et al. (2004). Perkell (1997) found differences in articulatory variability associated with different palatal vaults. They compared six speakers, who produced /i/, /ɪ/, and /ɛ/, and found that the speaker with the shallowest vault showed the least variability in tongue height for the three vowels.

The abovementioned studies are especially relevant with regard to the subject group under investigation in the present analysis (MZ and DZ twins). Following the results of the studies discussed above and thus the assumed influence of NATURE on intra- and inter-speaker variability in the production of vowels, one hypothesis could be that DZ twins (who differ in their palatal shapes) reveal more differences than MZ twins (who show identical palatal shapes) in their realizations of articulatory targets. This question will be addressed in the analysis below.

An additional source of inter-speaker variability in articulation is tongue muscle recruitment and thus the corresponding tongue shape. It has been found that different articulatory strategies can be used to reach the same acoustic output and here the shape of the tongue

could be an interesting factor to investigate. Not much is known yet about individual differences in tongue shape during articulation; in research, the focus mainly lies on a description and quantification of the general tongue shape in relation to different phonemes (vowels and/or consonants), positional conditions of a certain phoneme, or differing phoneme contexts (Harshman et al. 1977, Stone & Lundberg 1996, Hoole 1999, Davidson 2006). Nevertheless, a noticeable amount of inter-speaker variability in tongue shape during articulation is assumed and will therefore also be investigated in this study.

Inter-speaker variability in the production of the tense-lax contrast in 8 German front vowels was investigated by Hoole & Kühnert (1996) by means of electromagnetic articulography in seven speakers. While they found that all of their speakers showed the same overall articulatory pattern regarding the realization of the phonological oppositions *Vowel Height, Tenseness* and *Rounding*, they also found that speakers differed in their precise amount of jaw involvement in the tense-lax opposition. Some of their speakers showed significant jaw height differences for almost all tense-lax pairs, while others showed hardly any. Similar findings were presented by Johnson et al. in an earlier study (1993). They investigated variability in the vowel production of five speakers of American English by means of x-ray micro-beam data. They found a high degree of intra-speaker consistency in terms of the locations of the pellets at the midpoint of the vowel. Additionally, the movement consistency was investigated by looking at the location of the tongue dorsum pellet at three points in time. Again, very little variability was found between the different repetitions. Thus, consistent productions within speakers and hence low intra-speaker variability can be assumed in vowel targets and in gestures. However, individual differences were found in the production of the tense-lax distinction. Some speakers varied tongue height with a fixed jaw position while others coordinated tongue and jaw in the same manner or even in opposite directions. Moreover, speakers differed in their production strategies between the different tense-lax pairs. Johnson et al. give several possible explanations for the observed articulatory differences. Different vocal tract anatomies – and here in particular the degree of palate doming – may require different articulatory strategies. Ladefoged et al. (1972) found in their study a negative correlation between the range of jaw position at the midpoints of the vowels and an index of palate doming, i.e. speakers with flat palates showed a greater range of jaw positions. However, Johnson et al. could not find such a relationship in their study and point to a more

complex interaction that needs further research. The authors suggest the possibility that articulatory strategies are not directly correlated to physical requirements, since speech sounds can be produced by a variety of articulatory gestures. Thus, a person's articulatory strategy may only partly be determined by anatomy and might additionally be influenced by unique habits and idiosyncratic patterns. From their findings on speaker-specific articulation patterns the authors conclude that the acoustic output is the most important goal in the organization of speech production and argue for an auditory theory of speech production. This would point to the significance of auditory goals in vowels and a superior impact of the factor NURTURE. In accordance with these results, with respect to the present analysis no difference in the amount of inter-speaker variability between MZ and DZ twins in the vowel targets would be expected.

Regarding the influence of NURTURE and learning, an additional factor that should always be kept in mind is that speaker-specific variability has to be seen in the light of communicative demands. Communication is a two-sided process with a speaker on one side and a listener on the other. The aim of the speaker should be to be understood by the listener with the least effort possible (parsimony of the system). The aim of the listener is to receive the information the speech signal carries (cf. Ladefoged 1984). The speech signal itself consists of different segments with different degrees of importance. Words under focus and stressed syllables are the most crucial parts of the coded transferred information. Therefore, it can be assumed that these segments are spoken with more effort, and reveal larger articulatory gestures that are longer in duration (de Jong et al. 1993, de Jong 1995). It may also be assumed that these stressed syllables correspond to learned auditory goals, and that the unstressed syllables are generally shorter in duration, more influenced by coarticulation processes, less articulatorily distinct and more variable (de Jong 1998, Mooshammer & Geng 2008). In the aforementioned pilot study (see Section 2.3), an impact of the factor *stress* on inter-speaker variability in plosives could be found: it was more likely to find differences in VOT (voice onset time) and VDC (voicing during closure) within all twin pairs but especially in MZ pairs in *stressed* syllables than in *unstressed* syllables. Thus, it can be hypothesized that unstressed syllables are more sensitive to physiological factors and more influenced by the individual vocal apparatus (NATURE). Thus, regarding the present study, this means that the articulatory target (and thus also the acoustic output) of /i/ in an unstressed syllable is

assumed to be more similar in MZ twins than in DZ twins (while no difference is hypothesized between MZ and DZ twins in a stressed syllable). Here, the speech material under investigation is vowels that are considered to be even more influenced by the factor stress.

A further impact factor that will be addressed in the following analysis is that vowels and consonants interact in terms of their articulatory target (e.g. Alfonso & Baer 1982, Parush et al. 1983, Geng et al. 2003). Thus, as velar stops have been observed to be influenced by anatomical restrictions and velar stops and vowels have revealed a great deal of coarticulatory behavior, a stronger influence of NATURE on vowels surrounded by velar consonants might be hypothesized. For the current subject group this means that the factor *consonant context* of the vowel (i.e. /i:/ following a velar stop vs. following a liquid) may affect the difference in inter-speaker variability between MZ and DZ twins.

To summarize, while it is indeed assumed that vowels are oriented towards auditory targets (and hence NURTURE), from the findings reported above the influence of a speaker's individual anatomy and physiology (and hence NATURE) on the production of vowels cannot be neglected. Especially differences in palatal doming seem to be an influencing factor. In addition, the *consonant context* has to be kept in mind, since several studies have shown an interaction in articulation between velar consonants and vowels. Futhermore, the factor *stress* will be investigated since it is assumed that stressed syllables are less influenced by physiology. To shed some light on these points the current investigation studies inter-speaker variability in vowels in the speech of MZ and DZ twins, taking the factors NATURE and NURTURE into account.

4.1.1 Hypotheses

Regarding the abovementioned issues three hypotheses are made which will be investigated in the following section.

(1) Articulation strategies are influenced by the speaker's individual physiology, and thus DZ twin pairs reveal larger differences in the production of vowels (in their *articulatory target positions* and in their *tongue shapes*) than MZ twin pairs.

(2) The *consonant context* matters: the physiological influences of the tongue and the shape of the palate show a greater impact on the production of /i:/ following a velar consonant than following a liquid. Thus, the difference in the amount of inter-speaker variability between MZ and DZ twins is greater in /i:/ following /g, k/ than in /i:/ following /l/.

(3) The factor *stress* affects the impact of physiology, and thus the amount of inter-speaker variability in MZ and DZ twins. It is hypothesized that MZ twins and DZ twins differ in the amount of inter-speaker variability in *unstressed* but not in *stressed* syllables, mirroring a greater influence of physiology on the production of an *unstressed* syllable.

4.1.2 Method

The following analysis focuses on the production of the three corner vowels /i:/, /u:/ and /a/ in German.[8] For each speaker and vowel, articulatory target positions were defined as described in Section 3.5.2. For /a/, the lowest vertical position (and hence the minimum in velocity) of the jaw was chosen as the parameter that determines the reaching of the articulatory target position. For /i:/, the highest vertical position (and hence the minimum in velocity) of the tongue dorsum and for /u:/, the lowest horizontal position (and hence minimum in velocity) of the lower lip were taken as defining parameters. The determined time points with the corresponding articulatory positions were stored. In this way two dimensional data points (in the horizontal and vertical dimension) could be gathered and compared within the twin pairs.

Three analyses have been carried out in order to investigate the three hypotheses presented above. The first one deals with the impact of NATURE on inter-speaker variability in vowel *targets*. As a first step, a qualitative analysis of the articulatory targets of the corner vowels /a/, /i:/ and /u:/ for each subject is presented and scatterplots are used for a first visual inspection of the inter-speaker variability within the twin pairs. In a second step, statistical

[8] For further discussion on the phonological representations of the German vowels see Hall (1992) and Wiese (1996).

tests in R (ANOVA and post hoc Tukey tests[9]) have been conducted with SPEAKER as independent variable and horizontal/vertical position of the respective tongue coil (i.e. tongue back for /a/ and /u:/, tongue dorsum for /i:/) as dependent variable. In this way the target positions of each vowel are compared within each twin pair. In addition, a closer look is taken at the tongue shape by analyzing mean tongue shape plots and calculating the slope of the tongue between a) the tongue tip coil and the tongue dorsum coil and b) the tongue dorsum coil and the tongue back coil.

The second hypothesis focuses on the impact of the *consonant context* (or *coarticulatory effects)*; therefore, the articulatory analysis looks into the amount of inter-speaker variability in /i:/ following a velar consonant vs. /i:/ following a liquid. The third hypothesis deals with the factor *stress* and its possible influence on the amount of inter-speaker variability; thus the statistical tests look for significant differences in the articulation of /i:/ in a stressed position vs. /i/ in an unstressed position.

Articulatory data could be obtained from six pairs (out of seven); the male DZ pair is missing in this analysis. Due to bad reliability scores for the tongue dorsum and tongue back coils of speaker MA, the twin pair MZm2 also had to be excluded from the articulatory analysis of the vowels (cf. 3.5.1). Thus, 3 MZ pairs and 2 DZ pairs take part in this analysis. The speech material consists of the stressed vowels /a/, /i:/ and /u:/ in a non-focused position. The vowels stem from the verbs /ˈli:bə/ (1st p. sg. 'to love'), /ˈvaʃə/ (1st p. sg. 'to wash') and /ˈzu:xə/ (1st p. sg. 'to search for'), which were part of the carrier sentences (e.g. "Ich suche/liebe/wasche G(i/a/u)ba/Ha(g/k)(i/a/u) im Garten"). In this way the number of renditions of the vowels could be increased: each speaker repeated the target vowels 40 times in different carrier sentences presented to them on a monitor. For the second analysis, concerning the degree of coarticulation with neighboring consonants (i.e. /l/ vs. /g/), /i:/ was additionally investigated in the nonsense-word (name) /ˈgi:ba/. The word was presented in some of the carrier sentences and served to investigate inter-speaker articulatory variability of /i:/ in the syllable /gi:/ in contrast to the variability in the syllable /li:/. Furthermore, the

[9] The post hoc test has advantages over a normal t-test, since the Tukey test adjusts the results to the amount of t-tests that are made.

influence of the factor *stress* on inter-speaker variability in vowels was analyzed by investigating and comparing the production and variability of /i:/ in a stressed syllable (/ˈgi:ba/) vs. /i/ in an unstressed syllable (/ˈha:gi/).

Table 11: Overview of the number of analyzed items for each speaker and phoneme; mean and standard deviation (SD) for each phoneme.

Speaker	Number of analyzed items				
	/a/ /ˈvaʃə/	/i:/ /ˈli:bə/	/u:/ /ˈzu:xə/	/i/ /ˈha:gi/	/i:/ /ˈgi:ba/
MZf1a	40	45	40	10	9
MZf1b	38	35	38	8	14
MZf2a	39	46	40	10	10
MZf2b	38	48	40	9	10
MZm1a	39	48	40	10	10
MZm1b	40	47	40	10	10
DZf1a	34	46	38	9	7
DZf1b	38	47	39	8	7
DZf2a	30	34	32	6	7
DZf2b	37	45	36	8	9
MEAN	37.3	44.1	38.3	8.8	9.3
(SD)	(3.1)	(5.2)	(2.6)	(1.2)	(2.0)

Table 11 gives an overview of the speech material used. The number of repetitions differs slightly among the speakers since some articulatory data had to be excluded from the analysis; therefore, mean values and standard deviations of the analyzed repetitions are also given in the table. Altogether, 4.3% of the data had to be excluded. It has to be mentioned that due to the experimental setup and time restrictions during the EMA-recordings, the vowels /a/, /i:/ and /u:/ were iterated approximately 40 times, while the vowels /i:/ or /i/ in the /g/-condition were iterated just 9 times (on average) in each stress condition. Note also that in contrast to the verbs containing the vowel renditions, the target words /gi:ba/ and /ha:gi/ were in a focused position.

4.1.3 Results of the articulatory analysis of vowel TARGETS

4.1.3.1 Qualitative analysis of articulatory TARGET positions of /a/, /i:/ and /u:/

As a first step a closer look is taken at the articulatory realization of the vowel targets /a/, /i:/ and /u:/ for each speaker pair. Graphical plots are displayed visualizing the articulatory target positions for each repetition of the respective vowel with an interpolated line. The vertical and horizontal positions of the coils on the tongue tip, the tongue dorsum and the tongue back are measured as explained in Section 3.5.2. For each coil, the mean and standard deviation of the horizontal and vertical positions were calculated. Black ellipses are drawn around the midpoint of each coil with a size of two standard deviations for each axis. The ellipses were calculated by a principal component analysis with two main components: the highest amount of variability served to define the direction and length of the first axis, and the second axis is perpendicular to the first. In addition to the tongue coils, the position of the lower lip, which can give some information on the amount of lip protrusion for the realization of /u:/ and the position of the jaw for the realization of /a/, is shown in the figures.

Figure 10 displays the articulatory target positions of the twin pair MZf1. Different speakers are indicated by different colors (AF = red, HF = blue). Remember that red is always used for the speaker with the rotated and translated data. The figure shows that the articulatory targets of /i:/ and /u:/ are remarkably similar for this speaker pair in terms of the position and the shape of the tongue. For /u:/ a difference in lip protrusion can be observed: speaker AF (red) reveals a higher and more fronted position of the lower lip coil, and thus uses more lip protrusion in producing/u:/. The most differences between the articulatory target positions of the two speakers can be seen for /a/: the jaw seems to be lower, and thus the mouth slightly more open, for HF (blue), and the tongue position of her sister AF (red) is higher. Note that the tongue shape is still quite similar and the difference in tongue height might be a consequence of the different degree of jaw opening.

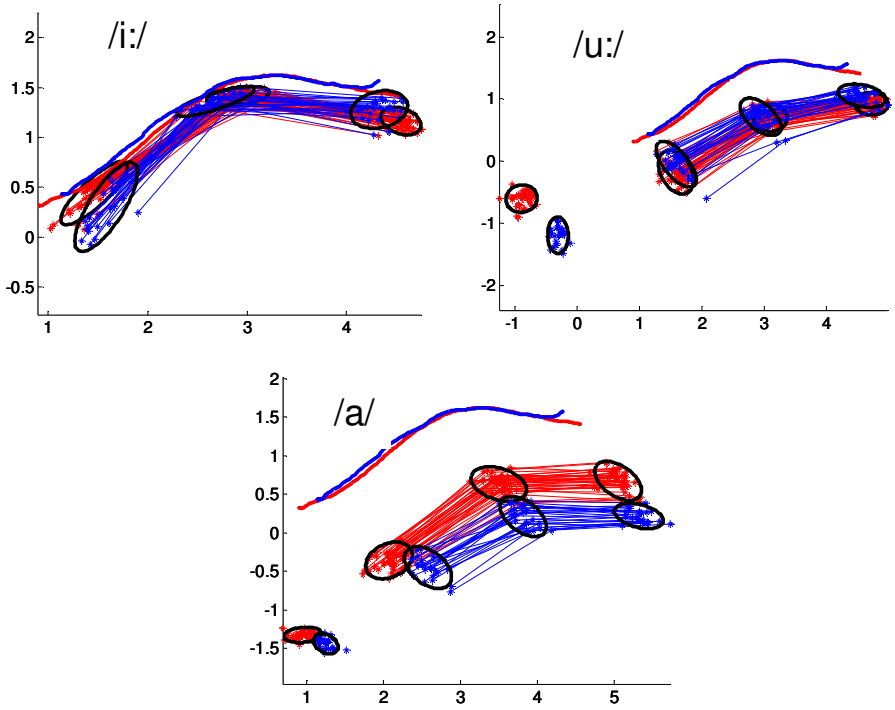


Figure 10: Articulatory target plot of /i:/, /u:/ and /a/ for MZf1 (HF = blue, AF = red). The black ellipses around the midpoint of the coil positions have a size of two standard deviations for each axis.

The second female MZ pair (MZf2) reveals more differences in articulation than MZf1 (cf. Figure 11). While the articulatory target position of /i:/ is still quite similar, the tongue position of /u:/ differs: GS (blue) shows a higher tongue position and more lip protrusion than her sister RS (red). RS reveals a tongue back position that is also high, but there is a steeper decline in the tongue contour towards the tongue tip. The articulatory /a/-target differs between the sisters in the same way it does for MZf1. Again, the speakers differ in the amount of jaw opening and tongue height, while the shape of the tongue is similar. Interestingly the two twin pairs reveal differences in the amount of intra-speaker variability for /i:/: both speakers of MZf2 show very little articulatory variability between the different renditions of /i:/ (the least among the three vowels), while the intra-speaker variability of both speakers of MZf1 is moderate and nearly the same for /i:/ and /u:/.

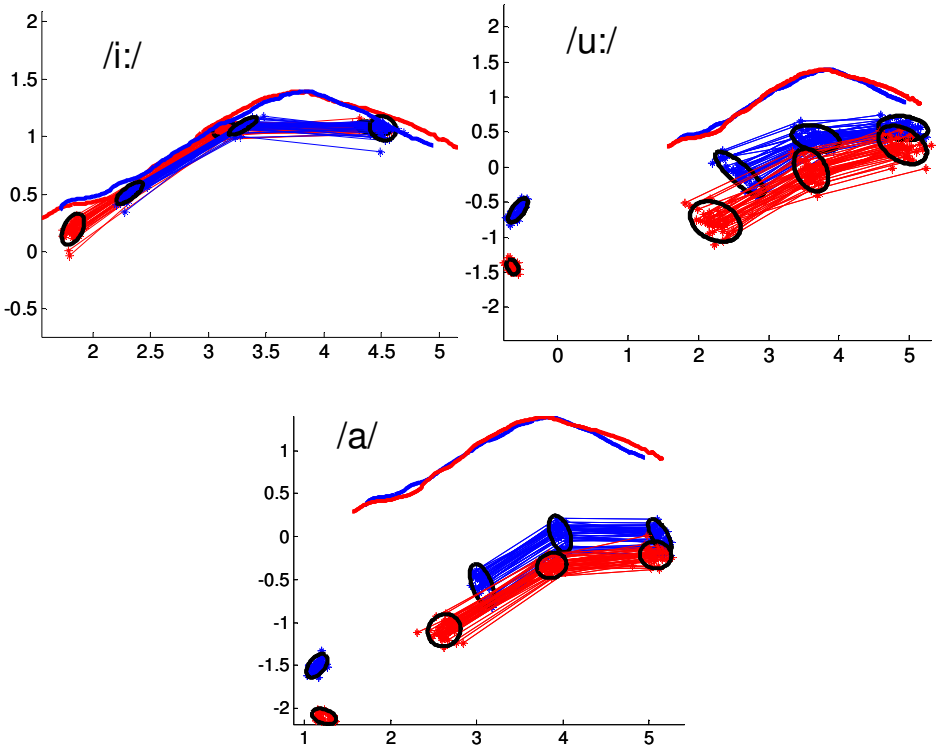


Figure 11: Articulatory target plots of /i:/, /u:/ and /a/ for MZf2 (GS = blue, RS = red). The black ellipses around the midpoint of the coil positions have a size of two standard deviations for each axis.

The speakers of the male twin pair MZm1 reveal differences in all vowel productions (cf. Figure 12). For /i:/, it is obvious that SL (red) shows very little intra-speaker variability while his brother CL (blue) shows quite a lot. These findings are in accordance with the abovementioned study by Brunner et al. (2009), in which the authors assume that speakers with a dome-shaped palate (like our pair MZm1) may choose the amount of articulatory variability but speakers with a more flat palate (like MZf2) in general show less variability, because of constraints on the variability range of the acoustic output. For the articulatory target of /u:/, the shape of the tongue and the height of the tongue back are quite similar between the speakers, while differences can be found in terms of the amount of lip protrusion: speaker CL (blue) reveals a much higher position of the lower lip coil, thus more lip protrusion can be assumed. Also, the articulatory targets for /a/ differ slightly between the speakers: the tongue dorsum and tongue back of speaker SL rest at the same height, but speaker CL (blue) reveals a lowered tongue back and therefore a more dome-shaped tongue.

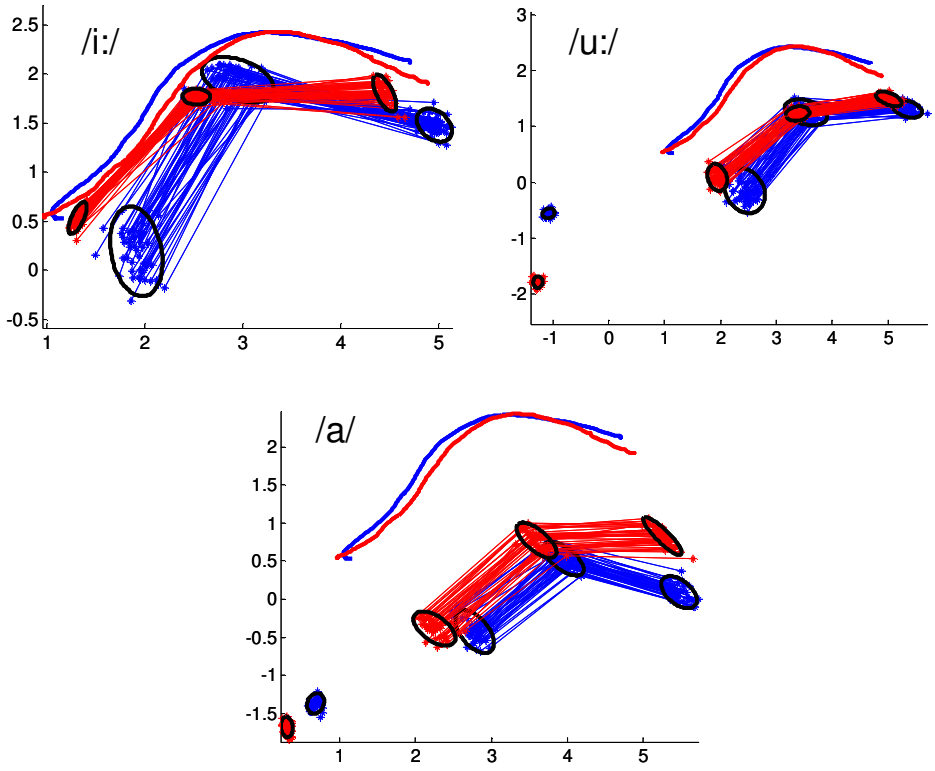


Figure 12: Articulatory target plot of /i:/, /u:/ and /a/ for MZm1 (CL = blue, SL = red). The black ellipses around the midpoint of the coil positions have a size of two standard deviations for each axis.

Figure 13 shows the articulatory target plots of the first DZ twin pair, DZf1. First of all, differences in palate size and shape can be seen with speaker LR (red), who has a smaller palate than her sister. Differences in articulatory strategies are obvious for all vowels. The shape of the tongue differs between the sisters in /i:/, /a/ and /u:/; in particular, the height of the tongue back coil is different. Speaker SR (blue) shows a higher position of the tongue back and thus a steeper and straighter tongue shape than her sister. The tongue back of speaker LR (red) is slightly lower than that of her sister and the tongue contour is more domed. Additionally, for /u:/, the speakers differ in total tongue height and in the amount of lip protrusion. For /a/, it is noteworthy that speaker LR (red) shows a remarkably high amount of intra-speaker variability, especially in terms of the tongue dorsum coil.

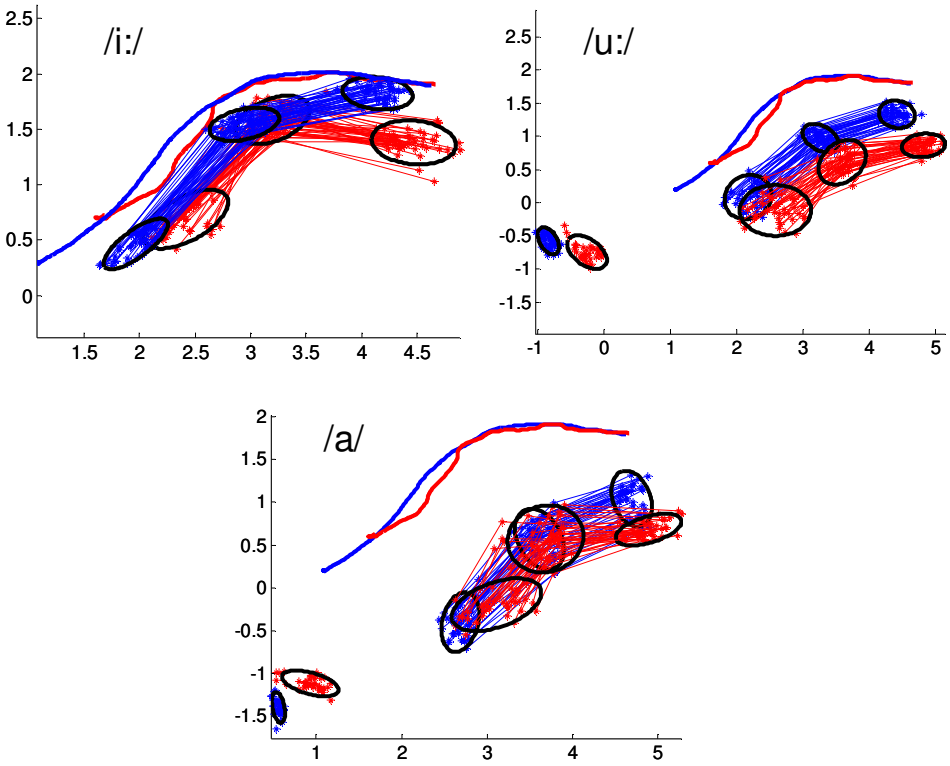


Figure 13: Articulatory target plot of /i:/, /u:/ and /a/ for DZf1 (SR = blue, LR = red,). The black ellipses around the midpoint of the coil positions have a size of two standard deviations for each axis.

The second DZ pair, DZf2, also reveals differences in all articulatory vowel targets (cf. Figure 14). For /i:/ the differences are least obvious, but even here the tongue position is different, mirroring the shape of the palate and especially the slope of the alveolar ridge. The tongue of speaker MG (blue) is more fronted, and the black ellipse that is drawn around the tongue dorsum coil and marks the intra-speaker variability runs parallel to the alveolar ridge. The tongue dorsum of speaker TG (red) is more retracted and varies horizontally but also vertically. For /u:/ and /a/, the speakers differ especially in terms of the shape of the tongue, but also regarding the target position of all tongue coils. The measured tongue of speaker TG seems to be straight and increases in height in a steady rise from the tongue tip to the tongue back coil. The tongue of speaker MG (blue) on the other hand is more bent and very steep between the tongue tip and the tongue dorsum.

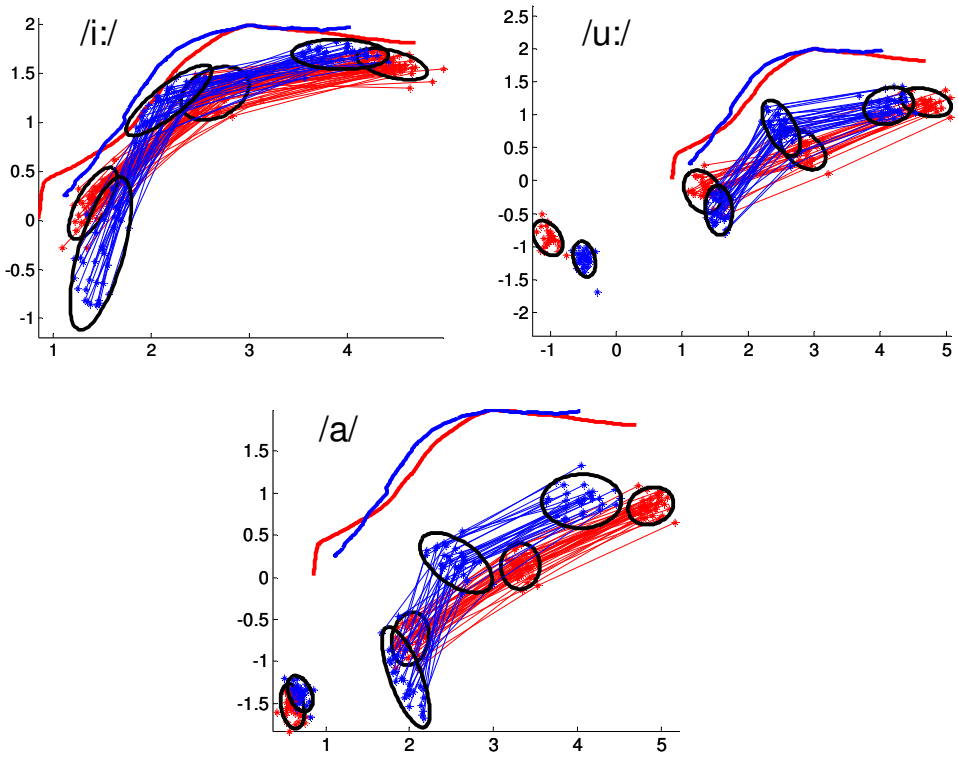


Figure 14: Articulatory target plot of /i:/, /u:/ and /a/ for DZf2 (MG = blue, TG = red). The black ellipses around the midpoint of the coil positions have a size of two standard deviations for each axis.

To sum up, differences in articulatory target positions can be found for all pairs but to different degrees. The twin pair MZf1 shows the fewest differences with nearly congruent articulatory targets for /i:/ and /u:/. From the qualitative analysis so far, a tendency towards more similarities in MZ twin pairs might be assumed (and here in particular within the female twin pairs), but this seems to be the case primarily in terms of the shape of the tongue.

4.1.3.2 Quantitative analysis of articulatory TARGET positions of /a/, /i:/ and /u:/

To get a better estimation of the degree of inter-speaker variability within the twin pairs, in a next step the differences between the two speakers of a twin pair were analyzed quantitatively. This was done by comparing the vertical and horizontal positions of the respective tongue coil that was used to define the reaching of the vowel target within the twin pairs. Separate ANOVAs and post hoc Tukey tests with a) VERTICAL TONGUE POSITION and b)

HORIZONTAL TONGUE POSITION as dependent factors and SPEAKER as independent factor were calculated for each vowel. A detailed overview of the calculated ANOVAs and post hoc tests is given in the appendix (Tables B.1 and B.2). As described previously, different coils were used as references for determining the articulatory targets of the vowels (cf. Table 12): the tongue dorsum coil (tdo) was used for /i:/ and the tongue back coil (tba) for /u:/ and /a/. Thus, the positions in a horizontal (X) and vertical (Y) direction of these coils at the target positions were compared within the twins. Table 12 gives an overview of the amount of articulatory inter-speaker variability within the pairs in the production of /a/, /i:/ and /u:/. The numbers give the differences between the measured mean tongue coil positions in cm, and significant differences within a pair are printed in bold.

Table 12: Differences in target tongue positions (in cm) of the three vowels within the twin pairs, significant differences ($p < .01$) in bold.

Twin pair	Coil position	/a/ (tba)	/i:/ (tdo)	/u:/ (tba)
MZf1	vertical (Y)	**0.445**	0.044	**0.114**
	horizontal (X)	**0.250**	**0.193**	**0.165**
MZf2	vertical (Y)	**0.223**	0.029	**0.235**
	horizontal (X)	0.049	**0.168**	0.006
MZm1	vertical (Y)	**0.743**	**0.171**	**0.176**
	horizontal (X)	**0.226**	**0.428**	**0.326**
DZf1	vertical (Y)	**0.491**	0.011	**0.591**
	horizontal (X)	**0.235**	**0.126**	0.035
DZf2	vertical (Y)	0.057	0.077	0.049
	horizontal (X)	**0.824**	**0.469**	**0.583**

In general, the most similarities in the target positions within the pairs were found for the vowel /i:/ and the fewest for the vowel /a/. This can be explained in terms of vowel-dependent intra-speaker variability (see Brunner et al. 2005, Mooshammer et al. 2004), with high vowels showing less articulatory variability than low vowels. If less articulatory *intra*-speaker variation can be expected for /i:/ than for /a/, less *inter*-speaker variation can be assumed in siblings with similar physiology and palate shapes. The pairs with the least inter-

speaker variability in all vowels are MZf2 and DZf2 (but note that the difference in the vertical position of the tongue dorsum of /i:/ of DZf2 (not bold) reached significance with $p < 0.02$). The most differences were found with MZm1. From the plots above the pair with the most similar articulatory targets was expected to be MZf1, since nearly congruent articulatory target plots were found. Nevertheless, the mean target positions of the respective tongue coils very often differed significantly. This contradiction suggests that the differences in mean coil positions might not be the best way to compare articulatory targets. A more appropriate way might be the tongue contour which will be investigated in the following section.

In general, the results above cannot support the assertion that similar physiology leads to similar articulatory target positions in the investigated vowels, since no clear difference between MZ and DZ twins can be found in terms of significant target distances.

4.1.3.3 Tongue shapes

Since the tongue shapes seemed to be an interesting and perhaps more promising factor than the single target positions, they were considered more closely. Figure 15 shows the mean tongue shapes during the realizations of each vowel for the MZ pairs; different speakers are marked by different colors.

Both female MZ pairs reveal remarkably similar tongue shapes for /i:/. MZf1 also shows parallel tongue shapes for /u:/, while the speakers of MZf2 show parallel contours from the tongue tip to the tongue dorsum but slightly different slopes for the tongue back. For /a/ both pairs reveal differences in tongue height but similar slopes for the front and the back parts of the tongue. The male MZ pair on the other hand reveals differences in tongue shapes. The front part of CL's tongue (blue) has a steeper upward slope than his brother's and the back part of the tongue declines for /i:/ and /a/ while his brother's remains raised.

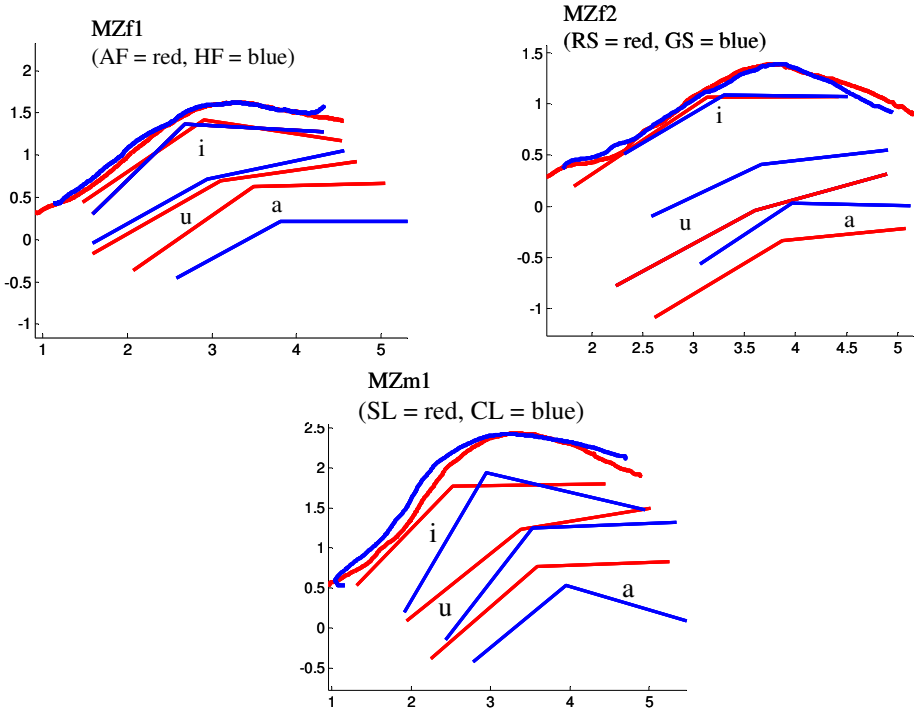


Figure 15: Mean tongue contours for the MZ pairs and all vowels, different speakers marked by different colors.

Both DZ twin pairs show differences in tongue shapes, but the speakers of DZf2 in particular turn out to be very different in their tongue contours (cf. Figure 16). The tongue of MG (blue) always rises very steeply at the front part and increases even more up to the tongue back. This extremely steep slope of her front tongue could already be observed before. The tongue shape of her sister TG inclines much more gently and regularly over the whole tongue. For DZf1, differences appear between the tongue shapes of /a/ and /i:/, especially in the back part. The tongue of SR (blue) rises up to the tongue back, whereas the tongue of her sister stays stable (for /a/) or declines (for /i:/) from the tongue dorsum to the tongue back. For /u:/ the tongue shapes are quite similar and parallel.

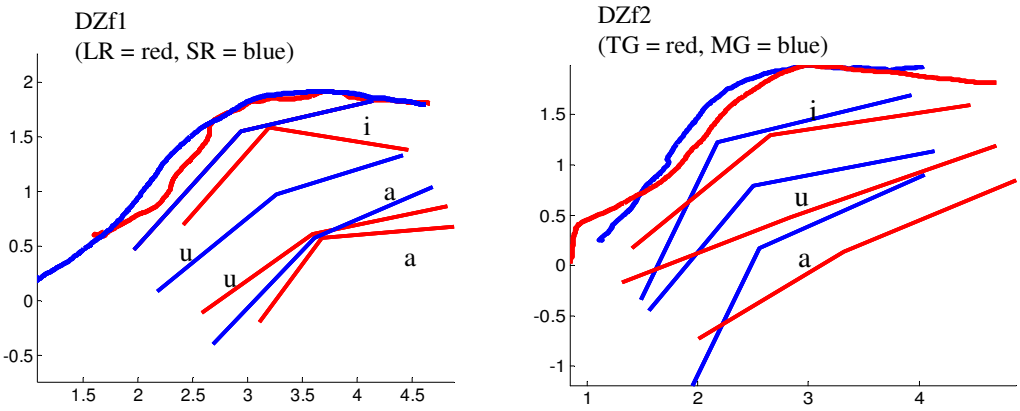


Figure 16: Mean tongue contours for the DZ pairs and all vowels, different speakers marked by different colors.

To quantify the differences in tongue shapes the slope of the tongue was calculated a) between the tongue tip coil and the tongue dorsum coil (henceforth called SlopeA) and b) between the tongue dorsum coil and the tongue back coil (SlopeB). The following formula (with x and y representing the horizontal and vertical positions of the respective two tongue coils) was used to calculate the two slopes (m) for each speaker:

$$m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}.$$

In a next step the absolute difference between the slopes was calculated for each twin pair, vowel, and both slopes. Figure 17 displays the calculated differences in a bar plot. The black lines separate the three vowels; the dotted lines separate DZ and MZ twins. The different colors mirror the two slopes, with SlopeA describing the shape of the tongue from the tip to the dorsum and SlopeB reflecting the tongue contour from the dorsum to the back.

For all vowels the female MZ pairs reveal the smallest amount of total differences, and DZf2 reveals the highest amount of total differences. DZf2 especially stands out in terms of differences in SlopeA. This was mentioned before and mirrors the very steep rise of the tongue contour from the tongue tip to the tongue dorsum of speaker MG of this pair. The other DZ pair, DZf1, also shows a high amount of differences in slopes, especially for /a/ and /i:/ and SlopeB. This finding echoes the different tongue back slopes seen in the figures above, with a stable or even declining tongue back contour of LR. Thus, if we only looked at

the female pairs, an effect of zygosity on tongue contour could be assumed; however the analysis of the male MZ pair does not point in this direction, especially for the vowels /i:/ and /u:/. Here, MZm1 shows the second highest amount of total differences, and this is based, in contrast, on SlopeA: speaker SL showed a steeper rise than his brother from the tongue tip to the tongue dorsum. There are different possible explanations for this result. Two possible factors have been mentioned before in the qualitative analysis of the target positions: 1) the high palate of this pair, which goes hand in hand with differences in the amount of articulatory variability, could explain the differences (cf. Brunner et al. 2009) and 2) differences in speech rates and degrees of articulatory precision between the two speakers might be responsible. Both factors will be discussed in more detail in the summary section of this chapter.

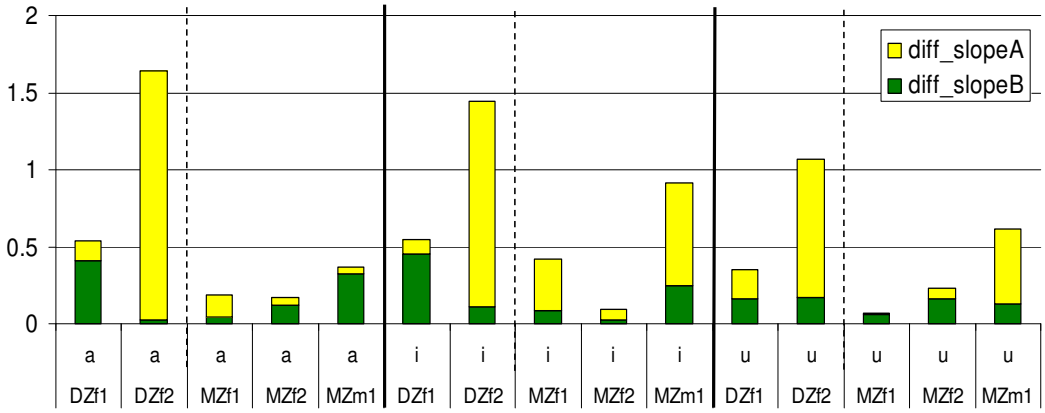


Figure 17: Absolute differences in tongue shape for each pair and vowel; SlopeA describes the front part of the tongue (from tongue tip to tongue dorsum), SlopeB the back part (from tongue dorsum to tongue back).

To sum up, the results do not corroborate hypothesis 1 presented above. More similar articulatory target positions in MZ twins than in DZ twins could not be found in the quantitative analysis, and thus no support for a crucial influence of physiology and anatomy on the articulation of stressed corner vowels was found. However, an interesting tendency could be observed related to tongue shape: particularly the female MZ twins revealed strikingly similar tongue shapes in the qualitative analysis while the female DZ twins did not. Nevertheless, this finding was not supported by the analysis of the male MZ pair.

4.1.4 Influence and interaction of the factors stress and consonant context

The above results are based on vowels in stressed syllables irrespective of the consonantal environment. To investigate the possible effects of stress/consonant context, the following analysis was carried out. Here, only the vowel /i:/ is taken as speech material. As mentioned in the introduction of this chapter, the effect of physiology is suggested to be stronger in unstressed syllables. Therefore, an articulatory analysis was conducted regarding the production of /i:/ and /i/ in three conditions: 1) /i/ produced in the unstressed syllable /gi/, 2) /i:/ produced in the stressed syllable /gi:/, and 3) /i:/ produced in the stressed syllable /li:/. Since the focus lies on the target tongue positions, the horizontal and vertical positions of the tongue dorsum coil were again measured and then compared between the speakers. ANOVAs and post hoc Tukey tests revealed significant differences, and Table 13 shows the results of the articulatory inter-speaker variability within the twin pairs. Bold numbers correspond to significant differences (numbers are given in cm) (cf. Tables B.3 and B.4 in the appendix).

From the results given in Table 13, it can be said that the factor *stress* has an impact on the articulatory inter-speaker variability in the production of /i:/ and /i/ within twin pairs. None of the MZ pairs shows significant differences in their target positions of the unstressed /i/. Of the two DZ pairs, one pair reveals significant differences in the horizontal position of the tongue dorsum ($p < .01$). This finding supports hypothesis 3 and points to a greater influence of physiology on the production of an unstressed syllable than on the production of a stressed one. For the two stressed conditions, each pair revealed at least one significant difference. Interestingly, for the DZ pairs the differences were more common in the stressed syllable /gi:/ than in the stressed syllable /li:/ (even though the larger sample size would favor significance in the /li:/-condition). For the MZ pairs, there was either no difference in variability between the two stressed conditions (MZf1, MZm1) or more differences in the syllable /li:/ (MZf2) were found. This is interpreted in terms of a stronger influence of physiology and biomechanics on articulation in the production of the vowel /i:/ following a velar consonant (in the syllable /gi:/) and corroborates hypothesis 2. The impact of NATURE on velar stops will be discussed again in more detail in Chapter 6.

Table 13: Differences in target tongue positions (in cm) of the vowel /i:/ or /i/ within the twin pairs, significant differences (p < .01) in bold.

Twin	Stress	Tongue dorsum Y	Tongue dorsum X
MZf1	Unstressed /gi/	0.0628	0.1618
	Stressed /gi:/	0.0404	**0.4844**
	Stressed /li:/	0.0446	**0.1931**
MZf2	Unstressed /gi/	0.1516	0.0763
	Stressed /gi:/	0.0491	0.1277
	Stressed /li:/	0.0290	**0.1677**
MZm1	Unstressed /gi/	0.1403	0.1887
	Stressed /gi:/	**0.2351**	**0.4352**
	Stressed /li:/	**0.1712**	**0.4280**
DZf1	Unstressed /gi/	0.1510	0.0092
	Stressed /gi:/	**0.1312**	**0.2342**
	Stressed /li:/	0.0383	**0.2590**
DZf2	Unstressed /gi/	0.0628	**0.3254**
	Stressed /gi:/	**0.1699**	**0.3013**
	Stressed /li:/	0.0728	0.4700

4.1.5 Summary and conclusion

Of the three hypotheses at the beginning of this section two could be supported by the findings from this analysis. First, the factor *lexical stress* plays a role when the influence of physiology (NATURE) on articulatory inter-speaker variability is investigated. No strong effect of zygosity was found in stressed syllables since MZ and DZ twins did not differ in their amount of significant differences in horizontal and vertical target positions. In the

investigated unstressed syllables, on the other hand, MZ twins did not show any significant difference, while one DZ twin pair did.

Second, the influence of NATURE (in terms of tongue physiology and shape of palate) on the production of vowels was found to be greater when the vowels were preceded by a velar consonant. In detail this means that the articulatory target position of the vowel /i:/ was more similar in MZ twins than in DZ twins when the vowel followed a velar consonant.

While no general influence of zygosity on articulatory inter-speaker variability with MZ twins being more similar than DZ twins in their articulatory targets (i.e. horizontal and vertical positions of tongue dorsum for /i:/ and tongue back for /u:/ and /a/) could be found, the investigation of the tongue shapes revealed at least a tendency towards this assumption. The female MZ pairs were very similar in the calculated slopes of the tongue, while both DZ twins revealed larger differences. This finding is restricted by the results of the male MZ pair, who also showed great inter-speaker variability in tongue shapes. One reason for this difference might be the high palate of this pair: as indicated in the introduction speakers with high palates have been found to be more flexible in their articulatory variability while speakers with flat palates in general show a more limited range of articulatory variability (see for example Brunner et al. 2009). This could already be observed in the articulatory target plot for /i:/ of this pair (cf. Figure 12) and may have an influence on the tongue contours. Another reason might be the different degrees of articulatory precision of the speakers of the male MZ pair revealed. Speaker SL (red in the plots) read the target sentences quite slowly and very precisely although he was asked to read them normally and without hyperarticulation. This difference between the speakers might also have an effect on the mean tongue shapes.

Hence, overall no strong effect of zygosity on inter-speaker variability in articulatory targets in the three stressed vowels /a/, /i:/ and /u:/ was found. This indicates a lesser influence of physiology (NATURE) than expected in the production of vowels and points to the assumption that learned auditory targets and shared social environment play an important role. This hypothesis will be investigated in more depth in the following acoustic analysis of the vowels.

4.2 Acoustic inter-speaker variability in vowels

Concerning the influence of biology (NATURE) on vowel production there is a long research tradition addressing the fact that the individual length of the vocal tract influences its filter characteristics. Thus, not only do different articulatory vowel configurations result in different formant patterns, but also different speakers vary in their formant values (Fant 1960). First of all a longer vocal tract leads to lower formant frequencies, resulting in the fact that children's formants (generally) being highest in frequency, followed by female and then by male speakers (Peterson & Barney 1952, Lindblom & Sundberg 1971). In addition, studies on the influence of aging on formant frequencies have revealed a similar effect. Luchsinger & Arnold (1965) found that the respiratory system and digestive tract lower with increasing age. The resulting lengthening of the vocal tract should lead to the known formant lowering. Note, though, that while some studies have found a formant lowering with increasing age (Xue & Hao 2003, Linville & Rens 2001), others have not (Labov 1994). Concerning different formants, it has been shown that parameters in higher spectral regions (i.e. formants such as F3 and F4) are more likely to show speaker-specific differences than parameters in lower spectral regions (Stevens et al. 1968, Sambur 1975, Lewis & Tuthill 1940, Ramishvili 1966, Dukiewicz 1970). In addition to the impact of gender and age, individual differences in speakers' vocal tracts might result in speaker-specific formant patterns. This will be investigated in the present chapter by analyzing twins' vowel productions.

Of course, as mentioned previously (see Section 1.3), NURTURE can also affect speaker-specific formant patterns. To explain the great amount of variation in speech production and to emphasize the role of NURTURE, Peterson & Barney already stated in 1952 that "…both the production and the identification of a vowel sounds by an individual depend on his previous language experience" (Peterson & Barney 1952, p. 184). Interestingly, not only the speaker-specific production of speech is noted, but also possible inter-speaker variability in the identification of a speech sound is mentioned. The ability to identify a sound or to distinguish between two similar sounds is dependent on the learned phoneme inventory of a speaker (cf. Section 1.3.2) and plays a crucial role in research on second language acquisition.

Moreover, a language's phoneme inventory can influence the allowed amount of token-to-token variability in this language (Lavoie 2002, Manuel 1990, Jongman et al. 1985). The study

of Manuel (1990), for example, could show that the coarticulation of vowels is constrained by the phoneme inventory of the language. In detail, he found that the languages Ndebele and Shona (with the phonemic vowels /i, e, ɑ, o, u/) revealed greater anticipatory coarticulation for the target vowel /ɑ/ than did the language (Sotho), which has a more crowded mid and low vowel space (with the phonemic vowels /i, e, ɛ, ɑ, ɔ, o, u/).

In addition, the influence of a speaker's auditory acuity on speech production was explained above (see also Section 1.3.2). Several studies have shown that auditory acuity affects speaker-specific realizations of phonemes and phoneme contrasts (Newman 2003, Perkell, Guenther et al. 2004, Perkell, Matthies et al. 2004, Perkell et al. 2008, Ghosh et al. 2010).

The subject group under investigation here was also examined by Loakes (2006) in her dissertation. Thus, her study can give some helpful insights into the current topic of acoustic inter-speaker variability in twins. Among other parameters she examined individual differences in the formant patterns of several vowels in the speech of five male twin pairs (4 MZ and 1 DZ) and found F3 to be the most speaker-specific formant. Furthermore, lax vowels turned out to be more speaker-specific than tense vowels. These results support the findings of an earlier study on speaker-specific acoustic parameters in vowels (Loakes 2004), where F2 and F3 of /ɪ/ showed the most inter-speaker variability in twins' speech (3 MZ and 1 DZ pair). No focus was put on the difference in zygosity in her study, which might be due to the small group of DZ twins (i.e. one pair). However, from the discussed results it seems that no difference in inter-speaker variability between the MZ and DZ twin pairs was found.

The abovementioned pilot study (cf. Section 2.1), which investigated inter-speaker variability in 4 MZ twin pairs and 1 DZ twin pair, revealed that the central and back vowels [ɑ, ɑ:, ʊ, u:, o:] showed more significant differences than the front vowels [ɪ, i:, ɛ, e:, ʏ, y:] concerning all twin pairs. Comparing the MZ twin pairs with the DZ twin pair, the DZ pair only demonstrated a higher probability of showing differences in F3, but in general no significant difference in the amount of inter-speaker variability could be found between the twin types. The pilot study and the studies from Loakes (2004, 2006) strongly point to the influence of shared social environment (NURTURE) in vowel production: overall, the siblings were more similar than unrelated speakers in their vowel formants, but DZ twins did not show more differences than MZ twins in their acoustic outputs. However, since only one pair represented

the group of DZ twins in each of these studies, we must be cautious when drawing conclusions.

The possible influence of the factors *stress* and *consonant context* (and in particular, the coarticulation with a preceding *velar consonant*) on the production of vowels has been mentioned and described above in the introduction of the articulatory analysis (cf. 4.1). Given the results of the articulatory analysis, these factors have also been taken into account in the following analysis.

4.2.1 Hypotheses

Based on the articulatory analysis and the results of the previously mentioned literature, the following three hypotheses will be investigated further:

(1) DZ twin pairs need not naturally show more differences in their acoustic outputs regarding vowels than MZ twins, as they adjust their speech production to each other and auditory goals are assumed to be crucial (NURTURE).

(2) The physiology of the tongue and the shape of the palate (NATURE) have a greater influence on the production of the syllable /gi:/ than on the syllable /li:/. Thus, acoustic outputs of /i:/ following a velar stop are more similar in MZ than in DZ twins.

(3) The factor *stress* affects the impact of physiology (NATURE), and thus the amount of inter-speaker variability in MZ and DZ twins. Parallel to the articulatory analysis, it is hypothesized that MZ twins and DZ twins differ in the amount of inter-speaker variability in *unstressed* but not in *stressed* syllables, mirroring the greater influence of physiology on the production of an *unstressed* syllable.

4.2.2 Method

The investigated vowels were segmented and annotated as described in Section 3.4 and the formants F1-F4 were measured semi-automatically in the middle of the segmented interval in PRAAT with a positive time step of 0.01, a maximum number of 5 formants, a maximum formant value of 5500 Hz (for females) and 5000 Hz (for males), a window length of 0.025s and a pre-emphasis of 50 Hz. Each measured formant value of every analyzed vowel was checked manually and corrected if necessary. Parallel to the articulatory analysis scatterplots with dispersion ellipses (two standard deviations) of F1-F2 variation were calculated for each subject and vowel. After that, statistical analyses (ANOVAs and post hoc Tukey tests) in R (version 2.8.1) were run to look for significant differences in mean formant values within the pairs. In addition, following the approach of the articulatory analysis, the influence of a) the factor *stress* and b) the factor *consonant context* (i.e. coarticulation with a preceding liquid or velar stop) on the inter-speaker variability in the production of /i:/ in twin pairs was investigated.

For the acoustic analysis the data of all seven twin pairs could be used, thus 4 MZ and 3 DZ pairs were investigated. The speech material and the number of repetitions that could be used for the acoustic analysis are the same as for the articulatory analysis described in Section 4.1.2. Table 14 gives an overview of the averaged renditions for each target phoneme.

Table 14: Overview and number (average per subject) of analyzed items with their stress condition.

Vowel	Stress condition	Target word	∅ N per subject (SD)
/a/	stressed	/ˈvaʃə/	37.3 (3.1)
/i:/	stressed	/ˈli:bə/	44.1 (5.2)
/u:/	stressed	/ˈzu:xə/	38.3 (2.6)
/i:/	stressed	/ˈgi:ba/	9.3 (2.0)
/ɪ/	unstressed	/ˈha:gi/	8.8 (1.2)

4.2.3 Results of the acoustic analysis of vowel TARGETS

4.2.3.1 Qualitative analysis of the acoustic TARGETS /a/, /i:/ and /u:/

To get a first impression of the vowel spaces of each subject and the twin pairs in particular, F1-F2 scatterplots for each pair and the three vowels /a/, /i:/ and /u:/ in the stressed conditions are displayed. The following figures show the scatterplots for the seven twin pairs (Figure 18 shows the 4 MZ pairs and Figure 19 the 3 DZ pairs). Each measured F1-F2 value is marked by a single dot. Ellipses were calculated and drawn to illustrate the intra-speaker variability of each vowel. The two colors (blue and red) distinguish the two speakers of a twin pair.

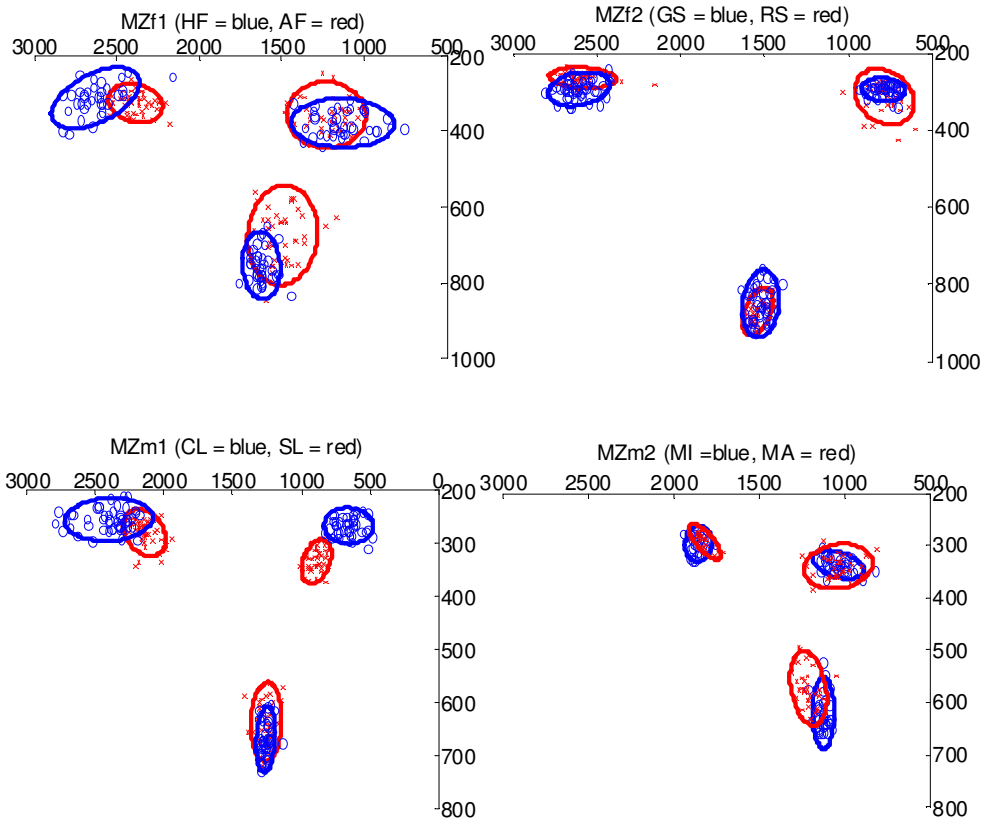


Figure 18: Scatterplots of F1 (negative y-axis) and F2 (negative x-axis) for the female MZ pairs (above) and the male MZ pairs (below) and the vowels [a], [i:], [u:]. The ellipses have a size of two standard deviations for each axis. The two colors (blue and red) distinguish the two speakers of a twin pair.

Overall, speaker-specific formant patterns are apparent: the speakers differ in the acoustic distance between the vowels, the intra-speaker variability of each vowel and the general shape of the vowel space defined by distances in the F1 and/or F2 dimension. However, when speakers within the twin pairs are inspected, similar patterns also arise. Figure 18 shows that the vowel spaces of MZf2 are most similar; here, the ellipses overlap nearly 100%. Also, MZm2 reveals congruent ellipses for /i:/ and /u:/, but slight differences for /a/ in terms of a higher F2 and a lower F1 for speaker MI. For MZm1 and MZf1, differences within the pairs can be assumed in the mean formant values of /i:/ and /a/ for MZf1, and /i:/ and /u:/ for MZm1. Similarities in the sizes of the ellipses and hence the intra-speaker variability are strikingly apparent: overall, variability in F1 and F2 is relatively small for MZm1, MZm2 and MZf2, but it is considerable for MZf1.

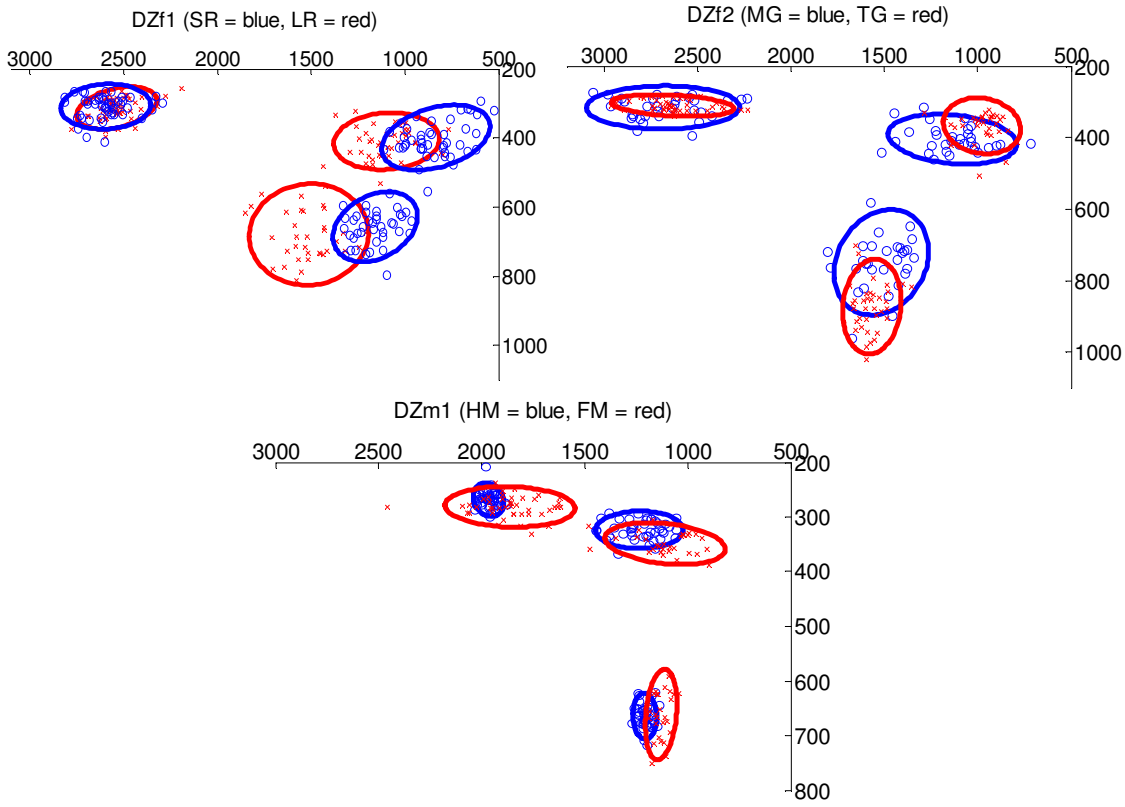


Figure 19: Scatterplots of F1 (negative y-axis) and F2 (negative x-axis) for the female DZ pairs (above) and the male DZ pair (below) and the vowels [a], [i:],[u:]. The ellipses have a size of two standard deviations for each axis. The two colors (blue and red) distinguish the two speakers of a twin pair.

In the scatterplots of the female DZ twins (upper part of Figure 19), F1-F2-ellipses of /i:/ are quite similar within the pairs. For /a/ and /u:/, differences can be seen, but especially for DZf1 in F2. As was shown in Figure 5 above, DZf1 shows a difference in the sizes of the palates and it can be assumed that the different sizes of the vowel spaces are influenced by this anatomical difference: LR (red), who has a smaller palate also displays a smaller vowel space than her sister SR (blue). The vowel spaces of DZm1 are quite similar in terms of acoustic difference (in F1 and F2) between the vowels, but differences in mean values might be found nevertheless.

4.2.3.2 Quantitative analysis of the acoustic TARGETS /a/, /i:/ and /u:/

To look for statistically significant differences in formants within the twins, mean formant values of F1-F4 of the three vowels were measured for each subject and compared with the corresponding sibling. Table 15 gives an overview of the mean formants for each speaker.

Table 15: Mean formant values (F1-F4) of /a/, /i:/ and /u:/ for each speaker.

Twin pair	Speaker	Vowel	Mean_F1	Mean_F2	Mean_F3	Mean_F4
MZf1a	AF	a	675	1492	2603	3968
MZf1b	HF	a	754	1621	2799	3918
MZf2a	GS	a	852	1521	2568	3569
MZf2b	RS	a	872	1538	2559	3918
MZm1a	CL	a	671	1256	2579	3612
MZm1b	SL	a	636	1252	2255	3450
MZm2a	MA	a	574	1211	2308	3067
MZm2b	MI	a	621	1124	2192	3229
DZf1a	SR	a	659	1164	2266	3841
DZf1b	LR	a	674	1475	2269	4083
DZf2a	MG	a	749	1516	2775	3756
DZf2b	TG	a	873	1566	2313	4041
DZm1a	HM	a	665	1208	2369	3665

DZm1b	FM	a	661	1129	2358	3361
MZf1a	AF	i:	325	2384	3538	4279
MZf1b	HF	i:	311	2626	3530	4525
MZf2a	GS	i:	294	2604	3249	4608
MZf2b	RS	i:	265	2585	3349	4423
MZm1a	CL	i:	256	2404	3421	4287
MZm1b	SL	i:	280	2141	2782	3863
MZm2a	MA	i:	294	1818	2640	3685
MZm2b	MI	i:	297	1856	2433	3372
DZf1a	SR	i:	309	2595	3161	4108
DZf1b	LR	i:	312	2538	3133	4341
DZf2a	MG	i:	315	2684	3268	4272
DZf2b	TG	i:	310	2632	3082	4310
DZm1a	HM	i:	268	1966	2898	3421
DZm1b	FM	i:	281	1862	2846	3431
MZf1a	AF	u:	356	1232	2741	4009
MZf1b	HF	u:	378	1129	2812	3990
MZf2a	GS	u:	296	788	3070	4015
MZf2b	RS	u:	314	783	2334	4044
MZm1a	CL	u:	268	654	2622	3716
MZm1b	SL	u:	335	886	2520	3585
MZm2a	MA	u:	340	1037	2336	3153
MZm2b	MI	u:	337	1036	2121	3626
DZf1a	SR	u:	400	834	2460	3917
DZf1b	LR	u:	412	1103	2732	4179
DZf2a	MG	u:	404	1136	3332	4048
DZf2b	TG	u:	367	977	3056	3776
DZm1a	HM	u:	323	1233	2158	3234
DZm1b	FM	u:	349	1109	2269	3273

For all vowels, separate ANOVAs were calculated for each FORMANT as dependent variable and for SPEAKER as independent variable. A detailed overview of the calculated ANOVAs with F values, degrees of freedom and the corresponding post hoc tests is given in the appendix (Tables B.5 and B.6). Table 16 shows the significant differences found between speakers of the same twin pair. The MZ pairs showed on average 5.5 significant differences (of 12 possible differences [3 vowels x 4 formants]) in F1-F4 of the three vowels, and the DZ pairs 6.3. Within the MZ pairs, the least inter-speaker variability in formants was found for the twin pair that shares genetics as well as environment (MZf2, as indicated previously in the scatterplots). The male MZ pairs show more differences; they are also the pairs which see each other only twice a month (Mzm1) or three times a year (Mzm2). Concerning the number of differences, the male DZ pair (DZm1) that lives together even comes before these MZ pairs. As hypothesized, the results point to a shared environment as the greatest impact factor on the acoustics of stressed vowels and support the findings of the pilot study.

An influence of vowel height on acoustic variability, as assumed in earlier literature in terms of less inter-speaker variation in /i:/ due to the strong influence of physiology on the production of this vowel, could not be found. In contrast to the similarities in the articulation of /i:/ (see Section 4.1), the acoustic analysis revealed many differences for /i:/. This indicates that in higher vowels less articulatory variance is necessary to achieve differences in the acoustic output. In fact, the MZ twins showed the most differences in the formants of /i:/ and the fewest in /u:/, whereby inter-speaker variability within the DZ twins was largest in /u:/ and smallest in /i:/.

Table 16: Significant differences in F1-F4 within the twin pairs of /a/, /i:/, /u:/ (post hoc Tukey test in R, significance level < .01).

Twin pair	/u:/	/a/	/i:/	No. of differences Total /12
MZf1		F1 F2 F3	F2 F4	5/12
MZf2	F3	F4	F1 F4	4/12
MZm1	F1 F2	F3	F1 F2 F3 F4	7/12
MZm2	F3 F4	F1 F2	F3 F4	6/12
DZf1	F2 F3 F4	F2 F4	F4	6/12
DZf2	F1 F2 F3 F4	F1 F3 F4	F3	8/12
DZm1	F1 F2	F2 F4	F2	5/12

Since the size and form of the vocal tract are considered to have a strong influence on the speaker-specific higher formants, it was expected that the MZ pairs would show less inter-speaker variability in F3 and F4 than the DZ pairs. However, the results do not point in a clear direction: MZf2 and MZm1 show 3 differences in the higher formants, and MZf1 only 2, but MZm2 even have 4; the female DZ twins also have 4 and 5 differences, whereas the male DZ pair shows only one significant difference in F4. Thus, no clear conclusion can be drawn from this.

4.2.3.3 Vowel spaces

An additional analysis was carried out to investigate the shape of the F1-F2 vowel spaces of the different speakers in greater depth. The following plots give insight into the size and the horizontal and vertical dimensions of the vowel spaces. The mean formant values of /a/, /i:/ and /u:/ are plotted and the dots are joined by a line. Each pair is shown in a separate plot and different speakers are again marked by their respective colors.

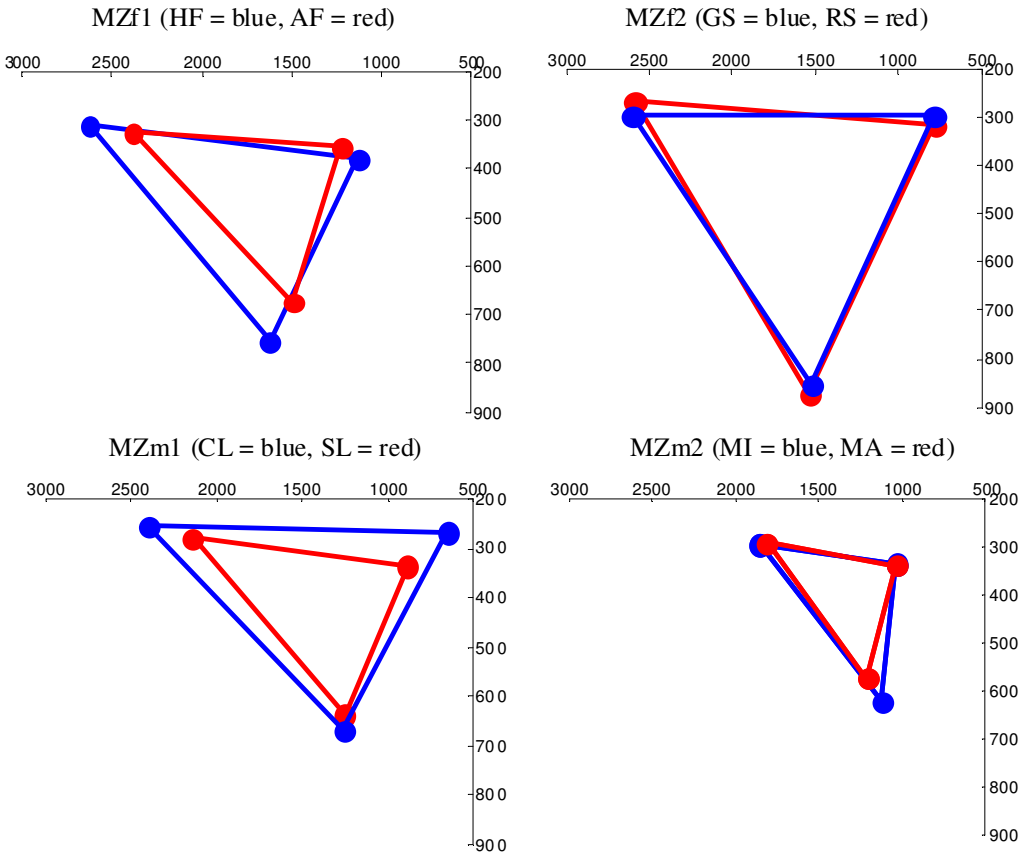


Figure 20: Vowel spaces of the MZ pairs, each pair in a separate plot, different speakers marked by different colors.

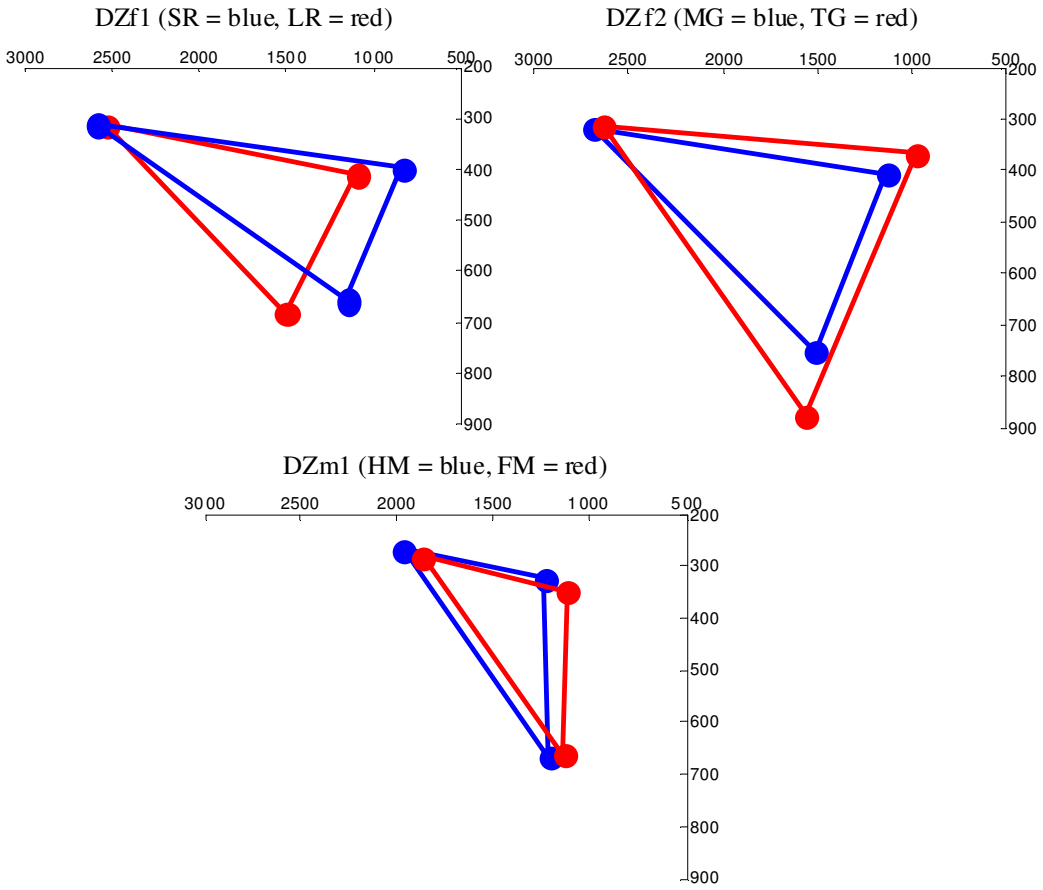


Figure 21: Vowel spaces of the DZ pairs, each pair in a separate plot, different speakers marked by different colors.

In general, a remarkable amount of inter-pair variability can be seen in the plots in terms of overall size and horizontal and vertical dimensions of the vowel spaces. Some speakers reveal very small vowel spaces, like MZm2, and some very large, with many acoustic differences in terms of F1-F2 values, like MZf2. This variation points to differences in articulatory effort and precision between the pairs, since it is known that hyperarticulation leads to a larger vowel space in F1-F2 dimensions. Indeed, it was already noted during the recording session that the speech of this male MZ pair is characterized by a very informal style and a low precision. It is known that a casual speaking style that is accompanied by hypoarticulation can shrink the vowel space (van Bergem 1993). In addition, there is a great deal of inter-pair variation in the relation between the horizontal and the vertical dimensions of the vowel

space. Some speakers show more acoustic distance in the horizontal (F2) dimension (like DZf1), and some more in the vertical (F1) dimension (like DZm1).

When we compare speakers within the pairs, differences appear, too. At first glance, the pairs MZf2, MZm2 and DZm1 seem to be most similar. The female DZ pairs and here especially DZf1 reveal obvious differences. As was mentioned previously, speaker LR (red) has a smaller palate and tongue than her sister SR (blue); here she also displays a smaller vowel space in the vertical dimension and hence in the acoustic difference in F2.

To quantify the shape of the vowel spaces and to measure the relation of the horizontal to the vertical dimension of the vowel spaces, Euclidean distances (ED) were measured between the vowels for each speaker. The ED between /i:/ and /u:/ determined the size of the horizontal dimension, and the ED between /u:/ and /a/ determined the vertical dimension. Then, the relation of the ED_horizontal to the ED_vertical was calculated. This calculated relation coefficient then gives a measurement of how the vowel space is shaped. The higher the coefficient, the larger the horizontal dimension of the vowel space (in relation to the vertical dimension).

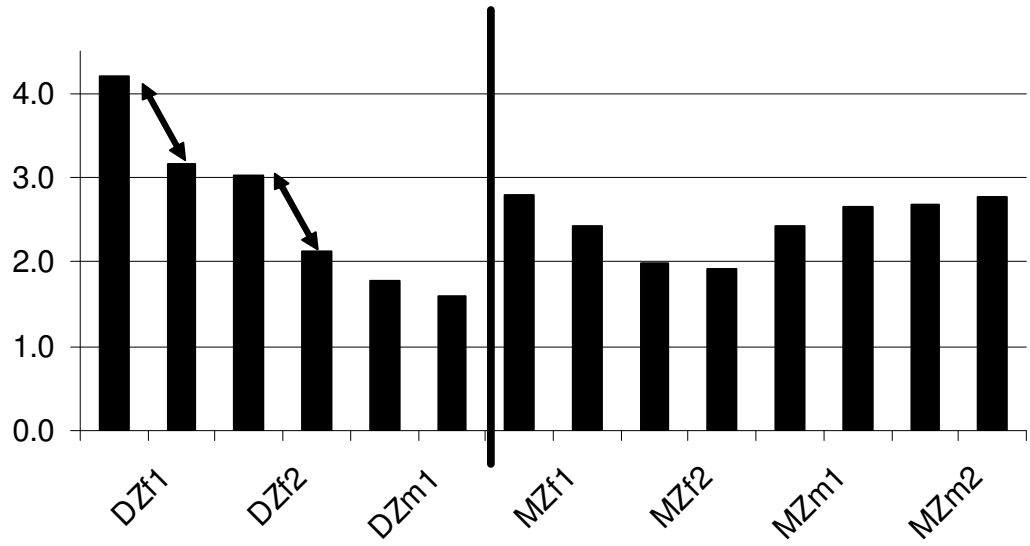


Figure 22: Relation coefficient (Euclidean distance between /i:/ and /u:/ divided by Euclidean distance between /a/ and /u:/) of the vowel space for each speaker, siblings plotted next to each other, DZ pairs on the left side of the black line, arrows mark speaker pairs with greatest differences.

The figure clearly shows that two pairs stand out in their differing relation coefficients. DZf1 and DZf2 reveal remarkable differences in the height of the bars, whereas all other pairs show very similar relation coefficients, mirroring similar shapes of the vowel spaces. Thus, 2 of 3 DZ pairs reveal differences in terms of acoustic vowel spaces, while all four MZ pairs show similarity. This finding points to an influence of zygosity on the realization of the acoustic contrast of the F1-F2 vowel space as defined by the three point vowels /a/, /i:/ and /u:/. Thus, even though MZ and DZ twins do not differ in their acoustic inter-speaker variability of the vowel targets, an influence of zygosity and hence NATURE could be found in the realization of the vowel contrasts. This can be interpreted in the following way: while auditory goals and the influence of NURTURE are crucial in realizing vowel targets, there still exists an impact of physiology (maybe linked to vocal tract length) on the overall configuration, size and shape of the acoustic vowel space.

4.2.4 Influence and interaction of the factors stress and consonant context

This section focuses on the second and third hypotheses, i.e. the factor *stress* and the coarticulatory influence of a *velar stop* on the acoustic inter-speaker variability of the vowel /i:/ are discussed. In addition to the production of /i:/ in the stressed syllable /li:/, the vowel is analyzed and compared in the syllables /gi:/ and /gi/, in a stressed and in an unstressed position respectively.

Note that (as in the articulatory analysis) due to the speech material the number of repetitions differs between the three conditions. For the analysis of the formants of the vowel in the stressed and unstressed syllables /gi:/ and /gi/ only 9 repetitions per condition (on average) could be taken into account. The mean formants of /i:/ in /li:bə/ could be investigated in approximately 40 repetitions. Therefore, different sample sizes were used for the statistics, and it has to be considered that these variations influence the probability of finding significant differences. Tests with larger sample sizes are more reliable, and it is more probable to find significance on a lower p-level. Thus, in the interpretation of the results this difference has to be kept in mind. Table 17 gives an overview of the significant differences found within the pairs in the three conditions: /i:/ after a liquid and after a velar stop and /i/ in an unstressed

syllable. Results of ANOVAs and post hoc tests are given in the appendix (Tables B.7 and B.8.)

Comparing the group of the MZ twins with the group of the DZ twins, it is noteworthy that there is a clear majority of significant differences in the formants of /i:/ produced in the stressed syllable /li:/ for the MZ twins, but a quite equally distributed number of differences in the formants for all three conditions for the female DZ twins. The male DZ pair (DZm1) only reveals a significant difference in F2 in /li:/. As noted above, more differences were expected for the stressed /li:/-condition because of the larger sample size. In spite of this fact, the female DZ twins (who differ in the size of palate and tongue) reveal more differences in F1 and F2 in both /g/-conditions, reflecting a stronger influence of physiology on the first two formants of a vowel following a velar consonant.

Table 17: Significant differences in formants within the twin pairs for the three conditions: /i/-/i:/ produced in the unstressed syllable /gi/, in the stressed syllable /gi:/, and in the stressed syllable /li:/ ($p < .05$).

Twin pair	Stressed /i:/ in /li:bə/	Stressed /i:/ in /gi:ba/	Unstressed /i/ in /ha:gi/
MZf1	F2 F4	F1	
MZf2	F1 F4	F1	
MZm1	F1 F2 F3 F4	F3	F2 F3
MZm2	F3 F4		F3 F4
DZf1	F4		F1 F2
DZf2	F3	F1	F1 F2
DZm1	F2		

When comparing the two /g/-conditions, the results support our hypothesis of an interaction of physiology and the factor *stress*: both female MZ twin pairs, who revealed strikingly similar palatal contours (cf. Figure 5), show differences in the stressed condition (F1), but not in the unstressed condition. Both female DZ twin pairs reveal more inter-speaker variability in the unstressed than in the stressed syllable, pointing to auditory goals as being crucial in stressed vowels. However, the results are not totally congruent, since MZm1 reveals a significant difference in F2 in the unstressed but NOT in the stressed condition.

4.2.5 Summary and conclusion

In the beginning of the chapter, three hypotheses were formulated. The first one addressed the greater influence of a shared environment (NURTURE) over an identical physiology (NATURE), and assumed that zygosity does not affect inter-speaker variability in stressed vowels, since they are oriented towards auditory goals. The results support this hypothesis, as no differences in the amount of acoustic inter-speaker variability in the formants of the stressed vowels /i:/, /u:/ and /a/ between the MZ and DZ twin pairs were found. The MZ pair that lives together revealed the fewest differences regarding formants, but one DZ pair that also lives together showed less acoustic inter-speaker variability than the MZ pairs who see each other only twice a month or even less. However, additional analysis revealed that zygosity seems to play a role in the similarity of F1-F2 vowel spaces. Two of the three DZ pairs revealed differences in their F1-F2 relation as defined by the Euclidean distance between the vowels, while the MZ pairs showed very similar configurations of their acoustic spaces in terms of the horizontal and vertical distances.

The second hypothesis was based on the findings from the articulatory analysis and suggested that *lexical stress* could be a possible influencing factor in inter-speaker variability. More inter-speaker variability should be found in stressed than in unstressed syllables within MZ twin pairs since a greater influence of physiology on the production of an unstressed syllable was assumed. Supporting evidence was found that there is an interaction between NATURE and the factor *stress:* physiology seems to have a stronger influence on the production of the vowel when it is produced in an unstressed syllable. Both female DZ twin pairs revealed more differences in formants in the unstressed condition (i.e. /i/), and the 2 female MZ twin pairs with the remarkably similar palatal shapes, showed more differences in formants in the stressed condition (i.e. /i:/).

The third hypothesis, which assumed a stronger effect of identical physiology on the acoustics of a vowel that follows a velar consonant than one that follows a liquid, could also be supported, since the female DZ pairs reveal more differences in the /g/-conditions than in the /l/-condition, whereas the MZ twins showed a similar number of differences or even fewer in the /g/-conditions.

To sum up, it can be said that a shared environment (NURTURE) plays a very important role in acoustic inter-speaker variability in vowels. However, there are several factors that contribute to this variability and intensify the impact of the identical physiology of the vocal apparatus (NATURE) of MZ twins, namely, the production of a *velar consonant preceding the vowel* and the factor *stress*. Moreover, the specific shape of the vowel space, which is defined by the relation of the acoustic distances on the F1 and F2 dimensions between the point vowels /a/, /i:/ and /u:/, seems to be affected by NATURE, since 2 of the 3 DZ twin pairs but none of the MZ twin pairs revealed differences.

4.3 Limitations and further research

Concerning the articulatory analysis, further investigations of the tongue shape should be made. Here, data from 2D-EMA-recordings is used, but it should be noted that the shape of the tongue is only interpolated and cannot be measured directly through EMA. During the recordings, only the positional data of the tongue coils can be obtained and therefore a clear conclusion cannot be drawn. Nevertheless the results point in a clear direction and justify further research using other techniques like magnetic resonance imaging (MRI) or ultrasound, where the shape of the whole tongue can be inspected.

In regard to the analyzed speech material some remarks have to be made. It has to be considered that due to the larger sample size, the overall probability of detecting significant differences is greater in the /l/-condition than in both /g/-conditions. Moreover, /li:bə/ is in a non-focused position, whereas /ha:gi/ and /gi:ba/ are under focus. Nevertheless, here the center of attention is on a comparison between inter-speaker variability in MZ vs. DZ pairs and not on a comparison between the three conditions for all speakers; thus the requirements are equally balanced and comparable.

Furthermore, in addition to the analysis of static vowel targets, an investigation of dynamic patterns could be promising. Research in the field of forensic phonetics has shown that speaker-specific characteristics might be more common in coarticulation patterns than in targets, since it has been suggested that the individual physiology is mirrored in the way a speaker manages to move from one target to the next one, while the actual target is influenced by shared auditory goals (Nolan et al. 2006, Kühnert & Nolan 1999). Thus, further

analysis in formant transitions in MZ and DZ twins would be very interesting and could contribute to this discussion.

In general the validity of the results is limited by the speech material, but more importantly, due to the time-consuming articulatory recording, it is restricted by the relatively small group of speakers, i.e. pairs. This is especially the case in the articulatory analysis since only the data of 3 MZ and 2 DZ pairs could be used. However, even this limited number of pairs could reveal the existence of pair-specific patterns, and thus studies with only one pair representing the DZ twins should be interpreted very carefully. In the following chapter the focus will be placed on consonants instead of vowels. In particular, the amount of inter-speaker variability in sibilants will be discussed.

5 INTER-SPEAKER VARIABILITY IN SIBILANTS

5.1 Articulatory inter-speaker variability in sibilants

The impact of physiological constraints (NATURE) on speaker-specific articulation in sibilants is the topic of the present chapter. It has already been mentioned that this influence might be stronger on consonants than on vowels since tactile feedback through tongue-palate contact and spatial restrictions due to anatomical and physiological boundaries are crucial in the production of consonants, especially in the case of sibilants (cf. Section 1.2, Stone 1995, Honda et al. 2002, Brunner 2009). For the production of /s/ the front part of the tongue is situated at the dento-alveolar ridge; in addition, a high jaw position is needed to create an obstacle. The airstream is forced through a short midsagittal groove along the anterior tongue blade, and the friction noise is generated when the airstream hits the upper incisors (Shadle 1985). To produce a /ʃ/ a (longer) groove is formed and the tongue is situated at the anterior palatal region to create a space underneath the tongue that functions as a sublingual cavity. The larger and more complex resonant cavities result in a fricative noise with lower spectral energy than for /s/ (Perkell et al. 2006). Thus, the difference in articulation between /s/ and /ʃ/ is the different size of the front cavity, in particular its length (Hughes & Halle 1956), and the additional use of the sublingual cavity for the production of /ʃ/. Tactile cues are considered to be important for the distinction of the sibilants, since a contact between tongue and lower incisors is assumed as a somatosensory goal for /s/ but not for /ʃ/ (Perkell et al. 2006). Note, though, that laminal and apical /s/ productions have been reported for several languages, as for example English (Ladefoged & Maddieson 1996). The difference is that for the laminal production the tongue blade is used to build the constriction (and hence the tip of the tongue rests against the lower teeth), while for the apical production the tongue tip is used. For /ʃ/ an acoustic coupling of the sublingual cavity and the lip cavity is possible when the lips are protruded (Toda et al. 2010).

Hence, these articulation strategies do not seem to be obligatory. Several studies have shown a large degree of inter-speaker articulatory variation. In the palatographic and linguographic study of Dart (1998), individual articulatory behavior in the production of coronal consonants in 20 speakers of American English and 21 speakers of French was investigated. Results suggest that articulatory variability is more speaker dependent than language related, since speakers of one language do not produce the coronal consonants at the same place neither on the palate nor on the tongue: dental-alveolar, laminal and apical productions were found independent of language background.

Fuchs et al. (2007) investigated speaker-specific articulation strategies in /s/ and /ʃ/ in 6 German speakers by means of electropalatography. While not much variation was found within speakers, a great deal of inter-speaker variation regarding the place of constriction for both sibilants but especially for /ʃ/ could be shown. No consistent differences between /s/ and /ʃ/ in either the articulatory or the acoustic domain were found, thus the authors suggest a highly speaker-specific behavior and motor equivalence: speakers who realize /ʃ/ with a front place of articulation need to produce lip protrusion, since otherwise they risk a perceptual confusion with /s/. Lip gestures are therefore required for those speakers' productions that do not produce substantial differences in the place of articulation of the tongue.

Toda (2006) studied sagittal contour tracings from magnet resonance imaging (MRI) during the production of /s/ and /ʃ/ in seven speakers of French. In particular, she looked at the different articulatory strategies for realizing the /s/-/ʃ/ contrast and observed two speaker-specific strategies: 1) tongue position adjustment, and 2) tongue shape adjustment. For the first strategy the tongue position differs between /s/ and /ʃ/, mainly in its horizontal position: for /ʃ/ the tongue is more retracted than for /s/, and no lifting or doming appears of the tongue back, but lip protrusion is apparent for /ʃ/. For the second strategy less lip protrusion is found for /ʃ/ but a great deal of tongue doming appears; here, the biggest articulatory difference between /s/ and /ʃ/ is the shape of the tongue.

Thus, different articulatory behavior has been described in the production of /s/ and /ʃ/ and the reason for this seems to be the speaker himself: articulation is highly speaker specific. Whether this inter-speaker variability arises through different motor strategies and whether

these differences are influenced by the speakers' physiology are the issues of the present chapter.

5.1.1 Hypotheses

As discussed above, speaker-specific articulation parameters are assumed in the production of sibilants. Reasons for this inter-speaker variability could be differences in the anatomy and physiology of the speaker. Therefore, the following two alternative hypotheses concerning the impact of NATURE or physiology on the articulation of the sibilants /s/ and /ʃ/ are tested:

(1) Physiology has a major impact on the articulation of sibilants.

MZ twins with nearly identical palatal shapes and shared biomechanical parameters of the tongue muscles are more similar in their articulation than DZ twins.

(2) Physiology has NO major impact on the articulation of sibilants.

MZ twins are as similar as DZ twins in their articulatory realizations of the sibilants.

5.1.2 Method

As a first step the articulatory target positions of /s/ and /ʃ/ were determined and compared within the twin pairs. Then, the mean tongue position for each of the two sibilants was calculated and the Euclidean distances were measured for the three tongue coils between the two speakers of each twin pair. In a third step the articulatory realization of the /s/-/ʃ/ constrast was investigated and the inter-speaker variability in realizing this contrast in regard to the horizontal and vertical variation in the tongue tip when producing the two sibilants was analyzed. It should be noted that even if the front part of the tongue is seen as the main articulator in producing the sibilants it does not necessarily have to be the tongue tip. However, since the coil on the tongue dorsum is assumed to be even farther away from the real place of constriction, and no data for the points between tongue tip and tongue dorsum could be gathered, the tongue tip coil was used for all speakers to investigate the target position of the phonemes and to determine the position of the constriction.

5.1.2.1 Speech material

The analyzed sibilants were part of slightly different carrier sentences. /ʃ/ was taken from the target word /vaʃə/ in the sentences: *'Ich wasche Hagi/Haga/Hagu/Haku im Garten'* ('I wash Hagi/Haga/Hagu/Haku in the garden'). /s/ was analyzed in /kʏsə/ in the carrier sentence: *'Ich küsse Kiba/Giba/Gaba/Guba im Garten'* ('I kiss Kiba/Giba/Gaba/Guba in the garden'). The number of repetitions of /s/ and /ʃ/ differs slightly between the speakers due to missing values or bad data that had to be excluded from the analysis (i.e. 5.5% of the data). On average, 37.6 repetitions of /s/ in /kʏsə/ and 38 repetitions of /ʃ/ in /vaʃə/ for each of the 12 speakers (6 twin pairs) could be examined for the articulatory analysis. Note that articulatory data could only be collected from 2 of the 3 DZ pairs as reported above.

Table 18: Number of analyzed items for each speaker and the sibilants /s/ and /ʃ/.

	Number of analyzed items	
Speaker	/s/	/ʃ/
MZf1a	39	40
MZf1b	39	38
MZf2a	39	39
MZf2b	44	39
MZm1a	40	40
MZm1b	40	40
MZm2a	39	33
MZm2b	31	35
DZf1a	32	30
DZf1b	38	38
DZf2a	35	43
DZf2b	35	41
MEAN	37.6	38
SD	3.68	3.64

5.1.2.2 Articulatory analysis of TARGET positions

For each speaker and sibilant articulatory target positions were measured. The point at which the target positions for /s/ and /ʃ/ were reached was in most cases clear and easy to define (see Section 3.8.1). As described in the beginning of this chapter the movement of the jaw is crucial for the production of the sibilants and the tongue tip serves as an articulator. Thus, to determine the target positions of /s/ and /ʃ/ the position of the jaw served as a reference. In cases where the determination was ambiguous the position of the tongue tip was taken into account. All measurements were carried out manually. The determined time points with the corresponding positions of the coils were saved. For statistical analyses ANOVAs and post hoc Tukey tests (with factor SPEAKER and HORIZONTLA or VERTICAL TONGUE POSITION as dependent variable) were calculated. Details of the statistical analyses are given in appendix C (cf. Tables C.1 and C.2).

5.1.2.3 Articulatory distances between speakers' /s/ and /ʃ/ productions

To get an estimation of the articulatory distance between the target positions of the siblings, the mean position for each of the three tongue coils was measured and the Euclidean distance was calculated for each pair and tongue coil (regarding the vertical (x) and horizontal (y) position of the coils). The following equation shows the calculation of the Euclidean distance (ED) between the two points *a* and *b* with the two dimensions *x* and *y*.

$$ED(a,b) = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2}$$

5.1.2.4 Comparing the realization of the /s/-/ʃ/ CONTRAST

To compare the mean tongue positions for /s/ and /ʃ/, Euclidean distances were calculated for each tongue coil between the mean position of /s/ and /ʃ/ for each speaker and then compared within the twin pairs. After that, the articulatory distance between /s/ and /ʃ/ was analyzed by comparing the amount of horizontal and vertical distance between the articulatory mean target positions of /s/ and /ʃ/.

5.1.3 Results of the articulatory analysis of sibilant TARGETS

5.1.3.1 Inter-speaker variability in the articulatory TARGETS of /s/

In the following plots articulatory target positions of the fricative /s/ for the three tongue coils (tongue tip, tongue dorsum, tongue back) can be seen. Each line connects the measured position for the three tongue coils and represents one rendition of the production of /s/. As in the analysis of the vowels mean positions and standard deviations were calculated for the horizontal and vertical positions of the tongue coils. The black ellipses (calculated by a principal component analysis with 2 main components and a radius of 2 standard deviations for each direction) show the intra-speaker variability around the mean value. Each plot shows one twin pair; the different colors indicate the two speakers of the particular pair. Again, as in the analysis of the vowels, the color red was used for the speaker with the rotated and translated data.

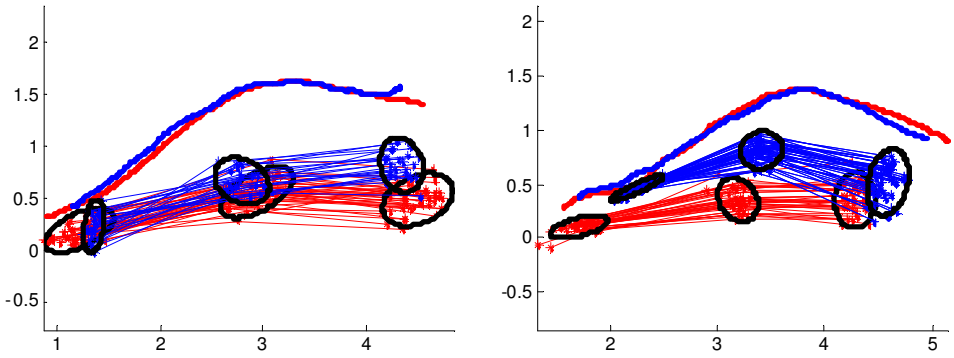


Figure 23: Tongue positions of articulatory targets of /s/ MZf1 (HF = blue, AF = red) and MZf2 (GS = blue, RS = red); left = front.

Again, the strikingly similar palatal shapes of both MZ female twin pairs can be seen. However, differences appear in their articulatory inter-speaker variability. MZf1 stands out in terms of very similar articulatory targets in the production of /s/. No differences were found for the vertical position of the tongue tip. In fact, the mean y-position measured for the two speakers was nearly the same (namely, 0.187 cm and 0.183 cm).

MZf2, in contrast, reveals differences in the articulation of /s/: all tongue coils differ in their target positions; the tongue tip, which is assumed to be the relevant articulator, shows significant differences in the horizontal ($F(11, 434) = 253.89$, $p < 0.001$) and vertical ($F(11, 434) = 290.68$, $p < 0.001$) positions. Two possible explanations could be postulated for these differences: first, the positions of the tongue coils may vary between the speakers, or second, the speakers may differ in their articulatory strategies. Great effort was made to position the coils exactly on the tongues within the pairs (with the help of photographs and tongue-coil templates, cf. Section 3.2.1 and Figure 4). Nevertheless, this speaker pair revealed differences in the position of the tongue back coil (see Section 3.2.1.2): the coil on the tongue back was positioned slightly more back for speaker RS than for her sister GS. However, this cannot explain the different tongue positions in Figure 23, since RS (red) reveals a more *fronted* tongue position. In addition, differences were also found between the sisters in the acoustics of /s/ (cf. Section 5.2.3: RS reveals a higher COG than her sister), thus a difference in the articulation strategy can be assumed. Bordon & Gay (1979) describe two different articulations of /s/: 1) the tongue tip touches the back of the lower teeth, and 2) the tongue tip is higher in the mouth behind the upper teeth. The graph seems to indicate an apical production of /s/ for speaker GS (blue) with the tongue tip placed behind the upper incisors, as has been reported previously for some speakers (Ladefoged & Maddieson 1996). When they were asked (independently), both speakers stated that they produce the /s/ with a laminal constriction.[10] The graph seems to be somewhat misleading (perhaps due to the limits of a two dimensional graph of a three dimensional gesture), yet a difference in the articulatory position between the sisters is clearly present, even if it does not seem to reflect the abovementioned apical articulatory strategy.

Regarding intra-speaker variability, it can be said that both pairs are quite similar, and the form and direction of the drawn ellipses for each coil are also very alike, especially for MZf2. By looking at the amount and direction of variability of the tongue tip of both speakers of MZf2, it can be seen that the main axes of the drawn ellipses of the tongue tip coil are

[10] The speakers were asked to report the place of the tongue tip during their production of /s/ and both stated that the tip was behind their lower incisors. In addition, they were asked to take a breath through their mouth while holding the articulatory position of /s/; here, they stated that their tongue got cold in the middle part and not at the tongue tip. Both of these statements point to a laminal articulation of /s/ with the tongue tip placed behind the lower teeth.

oriented parallel to the palate: the tongue tip varies most in the horizontal position for both speakers, while the speakers of MZf1 also reveal articulatory variability of the tongue tip in the vertical direction. This might point to an influence of the slope of the alveolar ridge: speakers with a more flat rise (MZf2) tend to vary the most along the horizontal dimension.

The next figure shows the articulatory targets for the male MZ pairs. When we look at the plot of MZm1, it is obvious that the speakers differ in their intra-speaker variability. Speaker CL (blue) shows much more articulatory variability than his brother SL (red). The same finding was already observed for /i/ (see Section 4.1) and again points to the assumption that speakers with dome-shaped palates (such as our pair MZm1) may choose the amount of articulatory variability, in contrast to speakers with flat palates, who are more limited in their articulation (such as MZf2) (see Brunner et al. 2009). In terms of inter-speaker variability, MZm1 differs indeed in the horizontal position of the tongue tip, but not in the vertical position. Even though the variation in the tongue tip position in the production of /s/ is high for CL (blue) and low for SL (red), the mean height of the tongue tip varies by only 0.06 cm (0.61 cm and 0.67 cm). The shape and form of the tongue during the target position of /s/ differ slightly between the brothers: the tongue of CL is concave (the tongue dorsum is much higher than the tongue back), whereas the tongue of SL is straighter (nearly no difference in height between the tongue dorsum and the tongue back). The shapes of the tongues of the two speakers of MZm2 are quite similar but not the same. MI (blue) shows a slightly more arched tongue than his brother. Both speakers have similar tongue heights but differ in their horizontal position. Speaker MA (red) reveals a more fronted position of the tongue tip than his brother: the horizontal and vertical position of the tongue tip differs significantly between the speakers.

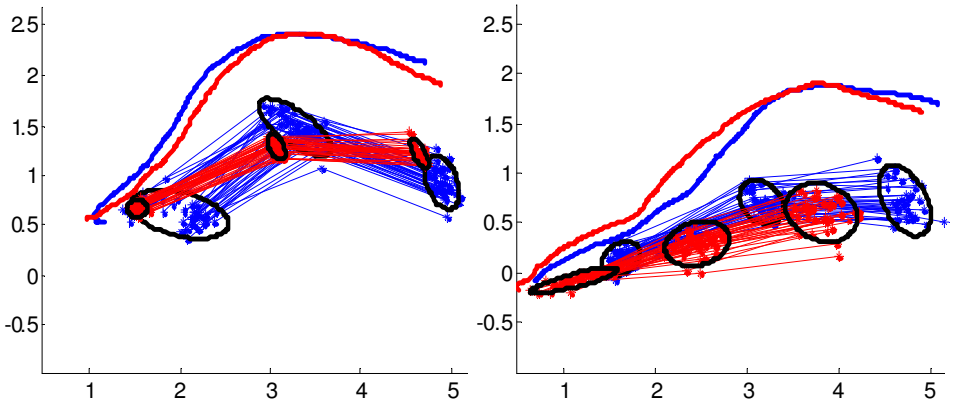


Figure 24: Tongue positions of articulatory targets of /s/ for MZm1 (CL = blue, SL = red) and MZm2 (MI = blue, MA = red).

Figure 25 shows the articulatory target positions of /s/ for the two female DZ twin pairs DZf1 (LRSR) and DZf2 (TGMG). Both pairs reveal inter-speaker variability in all tongue coils. The position of the tongue tip is significantly different in the horizontal and vertical dimension within the pairs ($p < 0.001$). The position of the tongue of LR (red) is lower than that of her sister, but the shape of the tongue is quite similar for this pair. The DZf2 pair, on the other hand, reveals great differences in the shape and form of the tongue. Speaker MG (blue) bends the tongue, whereas the tongue of her sister TG (red) is straight for the production of /s/. Moreover, they differ in the articulator responsible for the production of the fricative. TG produces the /s/ with a constriction at the tongue tip (or at least at the front part of the tongue) as all the other speakers did. In contrast, MG seems to use a more backward articulation strategy (with a place of articulation between the tongue tip and the tongue dorsum) to build the (laminal) constriction. A possible reason could be the extreme steepness and the small size of the palate of MG (and also the small tongue, cf. 3.2.1.2). In addition, it can be noticed that DZf1 differs in terms of intra-speaker variability, thus, the ellipse indicating the positions of the tongue dorsum for LR (red) is much bigger than that of her sister SR (blue). Here, the difference in the palatal contour at the place of constriction could be a possible reason.

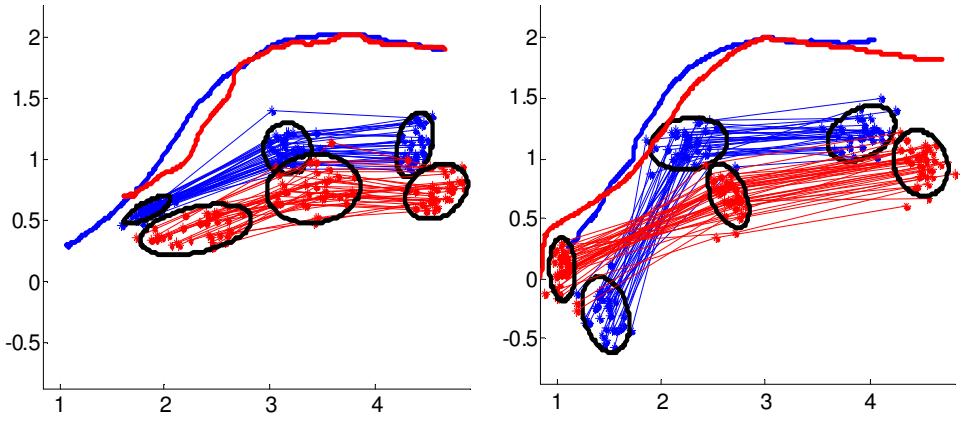


Figure 25: Tongue positions of articulatory targets of /s/ for DZf1 (SR = blue, LR=red) and DZf2 (MG = blue, TG = red).

To summarize, it can be said that no strong evidence could be found for an effect of zygosity (and hence identical physiology or NATURE) on inter-speaker variability regarding the articulatory target position of /s/ since no difference between MZ and DZ pairs was found: two MZ pairs and one DZ pair reveal obvious differences in their articulatory target positions (MZm2, MZf2, DZf2), two pairs are quite similar (MZm1, DZf1), and only one pair is very similar (MZf1). In this case, hypothesis 2 would be supported. However, as we have seen in the analysis of the vowels the shape of the tongue should also be kept in mind, and in this regard the MZ pairs resemble each other, whereas the speakers of DZf2 reveal clear differences. This result again points to the assumption that physiology and biomechanics indeed influence the shape and form of the tongue more than the precise articulatory position. Here, the individual physiology of the tongue muscles might be more important. However, only data from three midsagittal tongue coils can be gathered from the 2D articulatory measurements and the form of the tongue is only interpolated. In addition, especially for the production of sibilants, the 3D shape of the tongue and here the forming of a midsagittal groove along the anterior tongue blade is most interesting. Since no 3D data is available for the present study, no further focus will be placed on this here.

5.1.3.2 Inter-speaker variability in mean TARGET positions of /s/

To quantify the amount of inter-speaker variability and to get a better impression of the general variation for all twin pairs the mean tongue positions for each speaker were measured and then compared with the respective sibling by calculating the Euclidean distance for each tongue coil and twin pair. The following figure gives an overview of the measured Euclidean distances between the two speakers of each twin pair for the three tongue coils (tongue tip, tongue dorsum, tongue back[11]). The different coils are represented by different colors. The figure emphasizes that there is no clear difference between MZ and DZ pairs in their articulatory similarity in the production of /s/. These results support hypothesis 2: shared physiology does not have a major influence on the articulatory targets of /s/.

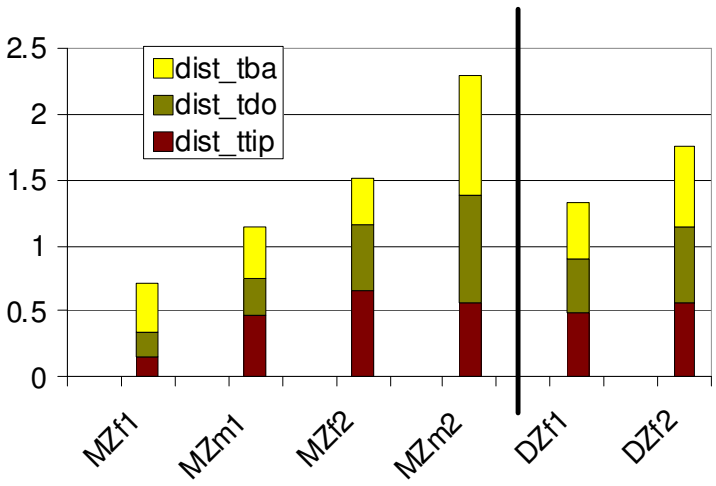


Figure 26: Euclidean distances for the three tongue coils between mean target positions of /s/ of each twin pair; the black line separates MZ and DZ pairs.

[11] Note that due to bad reliability scores of the articulatory data of speaker MA from MZm2 for the tongue dorsum and the tongue back coil, the data has to be treated carefully (cf. 3.5.1). However, when only the tongue tip coil is considered the results do not change.

5.1.3.3 Inter-speaker variability in the articulatory TARGETS of /ʃ/

For the production of /ʃ/ the rounding and protrusion of the lips can be crucial. Therefore, the following figures show in addition to the three tongue coils the target positions of the upper and lower lips. Note that due to problems during the recording session the upper lip sensor of AF (MZf1, red) and MA (MZm2, red) could not be used in the analysis and are therefore missing in the respective plots. Again the plots for the female MZ pairs (MZf1 and MZf2) are presented first. They reveal differences in the degree of inter-speaker variability: while the pair MZf1 seems quite similar in the shape and position of the tongue in terms of articulatory targets, the speakers of MZf2 again differ in their tongue positions: the speaker GS (blue) shows a more backward and raised tongue position than her sister. Similar to the production of /s/, this cannot be explained by different coil positions, since RS revealed a slightly more backward position of the tongue back coil (cf. 3.2.1.2). Again, the shape and form of the tongue are quite similar. Like in the production of /s/, here again the twin pair MZf1 reveals no significant difference in the vertical position of the tongue tip.

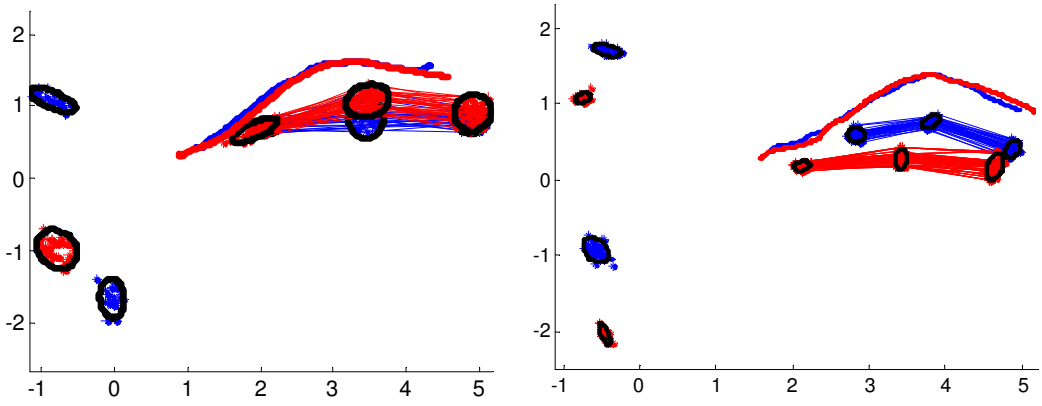


Figure 27: Tongue and lip positions of articulatory targets of /ʃ/ for MZf1 (HF = blue, AF = red) and MZf2 (GS = blue, RS = red).

For the male MZ pair MZm1, articulatory inter-speaker variability at the target position can be found in terms (x- and y-positions of ttip). However, the following figure also shows the quite congruent shape and bulge of the tongue. Again, speaker SL (red) reveals less intra-speaker variability than his brother CL, although the difference is less than for the production of /s/. MZm2 reveals very similar articulatory targets; in fact, no significant differences were found for the x- and y-positions of the tongue tip. Moreover, the shapes of the tongues are similar.

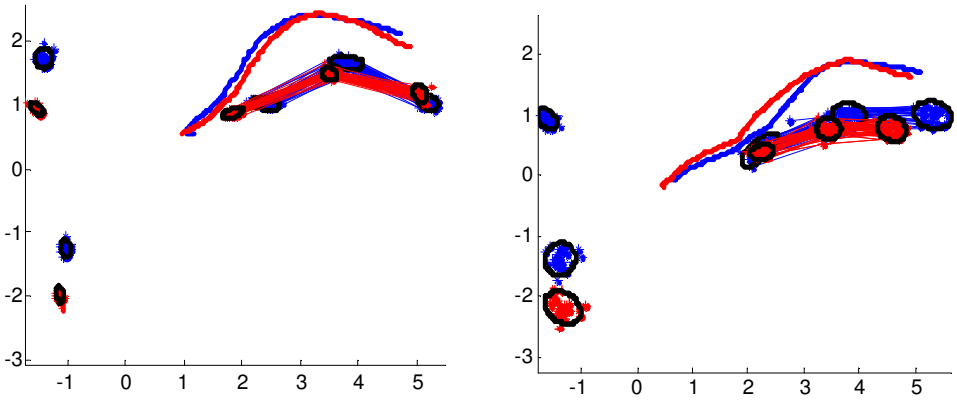


Figure 28: Tongue and lip positions of articulatory targets of /ʃ/ for MZm1 (CL = blue, SL = red) and MZm2 (MI = blue, MA = red).

Both DZ twin pairs reveal large articulatory inter-speaker variability and remarkable differences between the siblings in the vertical and horizontal positions of the tongue tip ($p < 0.001$) .Again, as could be seen previously in the production of /s/, DZf1 differs in the height of the tongue: the speaker SR (blue) produces /ʃ/ with a much higher tongue position and also a greater degree of lip protrusion than her sister. Another difference between these speakers is the high intra-speaker variability of LR in the horizontal position of the tongue dorsum. The most interesting pair regarding inter-speaker variability again turns out to be DZf2: the shape and position of the tongues are even more different than we have already seen for the production of /s/. MG (blue) produces /ʃ/ with a strongly bent tongue, whereas the tongue of TG (red) is straight. The extremely steep shape of the tongue of MG during articulation as well as in the "stationary position" during a speech break could already be observed during the recording session and is not due to differences in the coil positions

(between this speaker and all other speakers). Again TG uses the tongue tip for the necessary constriction, but as Figure 29 clearly shows, MG produces the fricative with the tongue dorsum. The articulatory strategies for the fricatives vary the most for this pair.

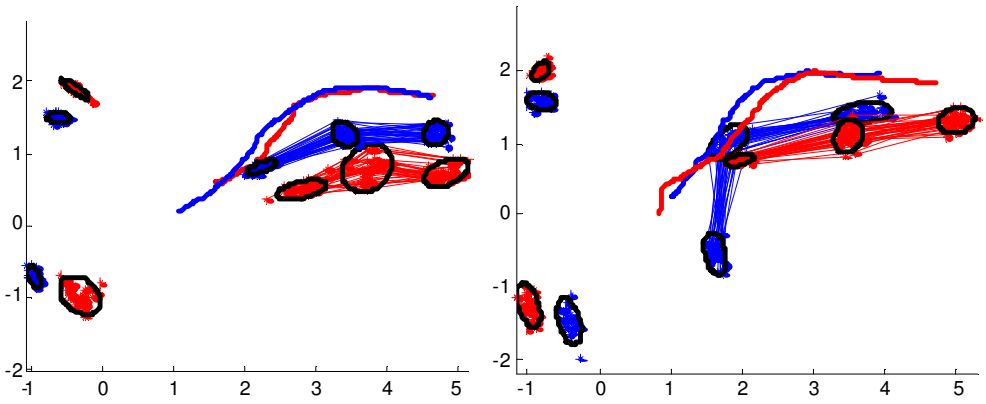


Figure 29: Tongue and lip positions of articulatory targets of /ʃ/ for the pairs DZf1 (SR = blue, LR = red) and DZf2 (MG = blue, TG = red).

To sum up, again differences in articulatory positions could also be found for the MZ pairs, but this time 3 of the 4 MZ pairs revealed similarities in articulation while both DZ twins varied in their articulation. At least a tendency towards an impact of physiology (and hence NATURE) on the articulatory targets of /ʃ/ was found, pointing to hypothesis 1. Note that the form and shape of the tongue again differ more for the DZ pairs than for the MZ pairs. All identical twin pairs reveal similar tongue shapes in terms of a straight or more bent shape.

5.1.3.4 Inter-speaker variability in mean TARGET positions of /ʃ/

Again, the mean tongue positions of each speaker and the Euclidean distances between the speakers of a twin pair for all tongue coils were measured. Figure 30 gives an overview of the results.[12] This time, in contrast to /s/, a difference in inter-speaker variability between MZ and DZ pairs can be seen. The average inter-speaker variability for MZ pairs is lower than

[12] Here again, the data for the tongue dorsum and the tongue back coil of speaker MA from MZm2 has to be treated carefully (cf. 3.5.1). Nevertheless, when only the tongue tip coil is considered, the results do not change. This holds true for Figure 31 as well.

that for DZ pairs. Especially one DZ pair (DZf2) shows a very high degree of variability, as could already be observed in the previous plots and which results from the different production strategy for the sibilants of MG (by using the part between the tongue tip and tongue dorsum to realize the constriction instead of the tongue tip like the other speakers).

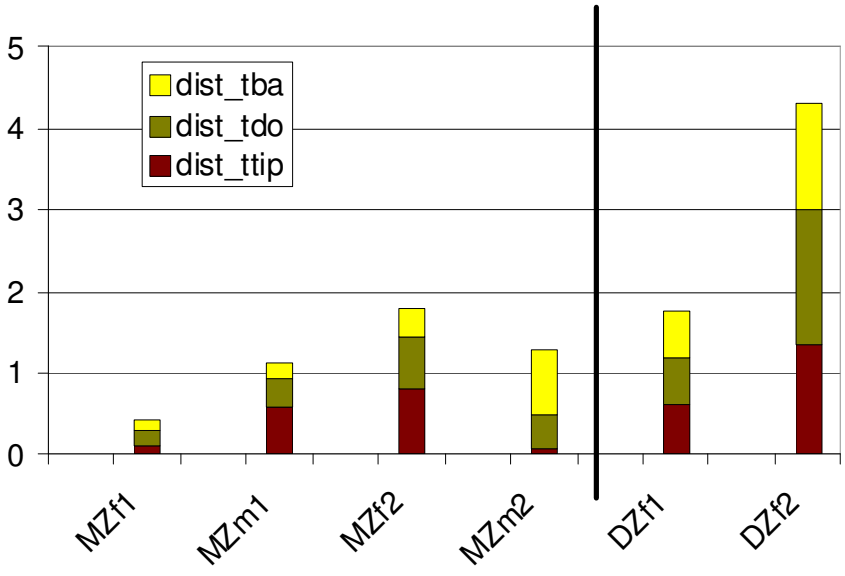


Figure 30: Euclidean distances for the three tongue coils between mean target positions of /ʃ/ of each twin pair; the black line separates MZ and DZ pairs.

Hence, in the articulation of /ʃ/ an influence of physiology on the realization of articulatory targets can be assumed. However, a Welch two-sample t-test with zygosity as independent factor and total ED (total amount of Euclidean distance for all tongue coils) as dependent factor failed to show significance (t = 1.4349, df = 1.095, p-value = .37). Note that a significant result with extremely small group sizes as in this case is very rare anyway, but nevertheless a tendency can be observed.

5.1.4 Summary of the articulatory analyses

To summarize, no clear difference between MZ and DZ twins in inter-speaker variability in the articulatory targets of sibilants could be found. Therefore, hypothesis 2, and thus no major impact of physiology (NATURE) on articulatory targets seems to be corroborated. However, results indicate a greater influence of physiology on the production of /ʃ/ than on

/s/, since here the MZ twins revealed greater similarities. Additionally, a more similar contour of the tongue during the production of /s/ and /ʃ/ could be observed for the MZ twins, pointing to an influence of shared biomechanical properties on the tongue shape. However, as stated previously, the interpolated tongue contours should be interpreted very carefully.

5.1.5 The articulatory realization of the /s/-/ʃ/ CONTRAST

In the following section the different production strategies in realizing the contrast between the two sibilants are investigated. In the speech corpus /ʃ/ was produced in the target word /vaʃə/, and /s/ in the target word /kʏsə/. Lip protrusion is assumed to still be apparent in the production of /s/ in /kʏsə/ through the transition from the preceding rounded vowel /ʏ/. Therefore, lip rounding and protrusion are not taken into account and all following plots only show the positions of the tongue coils.

First, the differences between the mean target tongue positions of the two sibilants of each speaker were analyzed. This was done by calculating the Euclidean distances between the mean positions of the three tongue coils for each speaker. Figure 31 visualizes the results. Speakers of the same twin pair are plotted next to each other. The figure shows the articulatory distance between /s/ and /ʃ/ for each speaker in terms of ED for each tongue coil. The graph indicates that the amount of articulatory difference between the two sibilants can vary within both MZ and DZ twins, since the greatest inter-speaker variability was found for MZm2 and DZf2. Hence, from this analysis no effect of zygosity and thus NATURE or physiology can be assumed on the articulatory realization of the /s/-/ʃ/ contrast; these results support hypothesis 2.

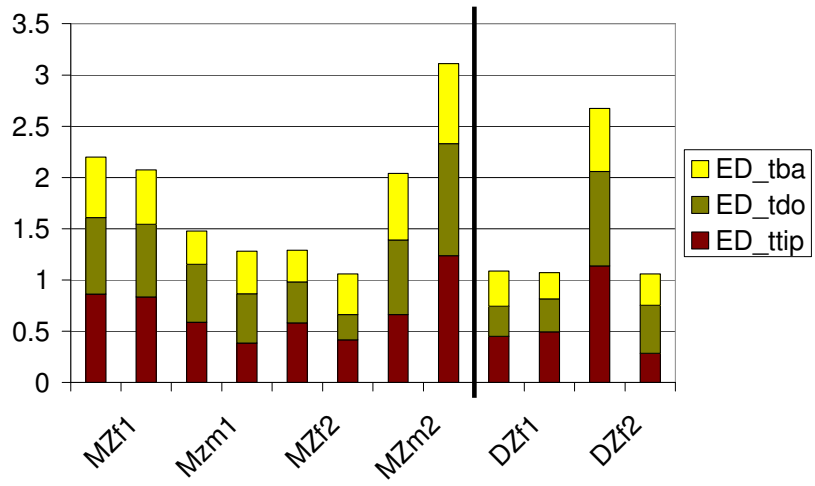


Figure 31: Euclidean distances (ED) between mean articulatory targets of /s/ and /ʃ/ for each tongue coil; speakers of the same twin pair are plotted next to each other; the black line separates MZ and DZ pairs.

A second analysis was done to take a closer look at the details of the production strategies used for this contrast. The following graphs in Figures 32 and 33 show the realizations of /s/ and /ʃ/ for each speaker separately.

Figure 32 shows the plots for all MZ twins; the productions of /s/ are plotted in dark green, and the productions of /ʃ/ in light green. Speakers of the same twin pair are plotted next to each other. By looking at the position of the tongue tip during the production of /s/ and /ʃ/ (marked in the graphs by black ellipses) the similarities between speakers of the same pair are obvious. Note that the ellipses are not calculated but are only drawn to aid in the visual examination. Speakers with a steep palate and an immediately rising palatal contour (MZm1) retract but also raise the tongue tip for the production of /ʃ/. They have to do this in order to not lose contact with the palate, which is necessary to build the constriction for the sibilant. Speakers with a shallow rise of the palate show more horizontal than vertical variation in the tongue tip between /s/ and /ʃ/ (MZm2, MZf1), or nearly no vertical difference at all (MZf2).

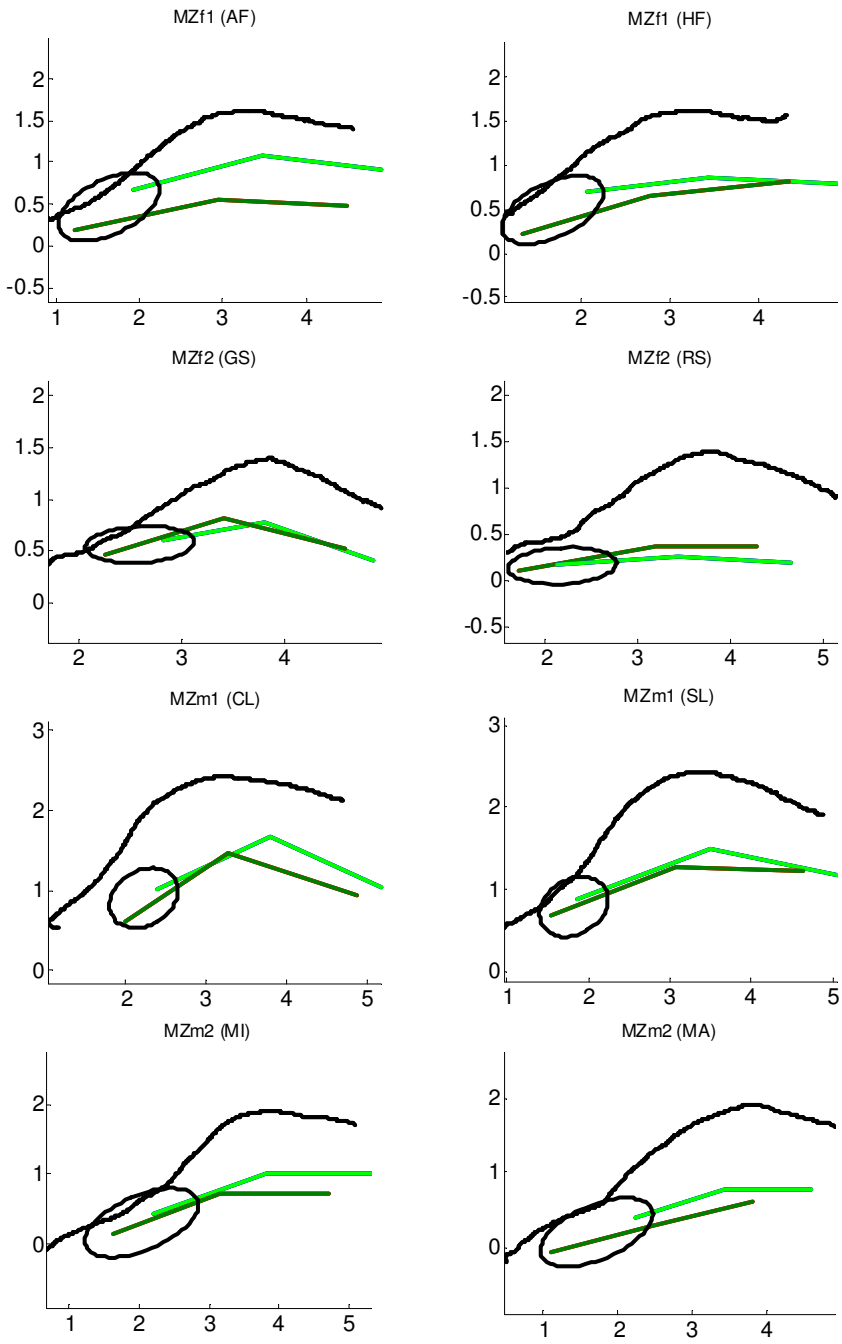


Figure 32: Mean articulatory target positions of /s/ (dark green) and /ʃ/ (light green) of the MZ twins; different plots show different speakers, ellipses visualize the amount of (horizontal and vertical) variation in the tongue tip.

.

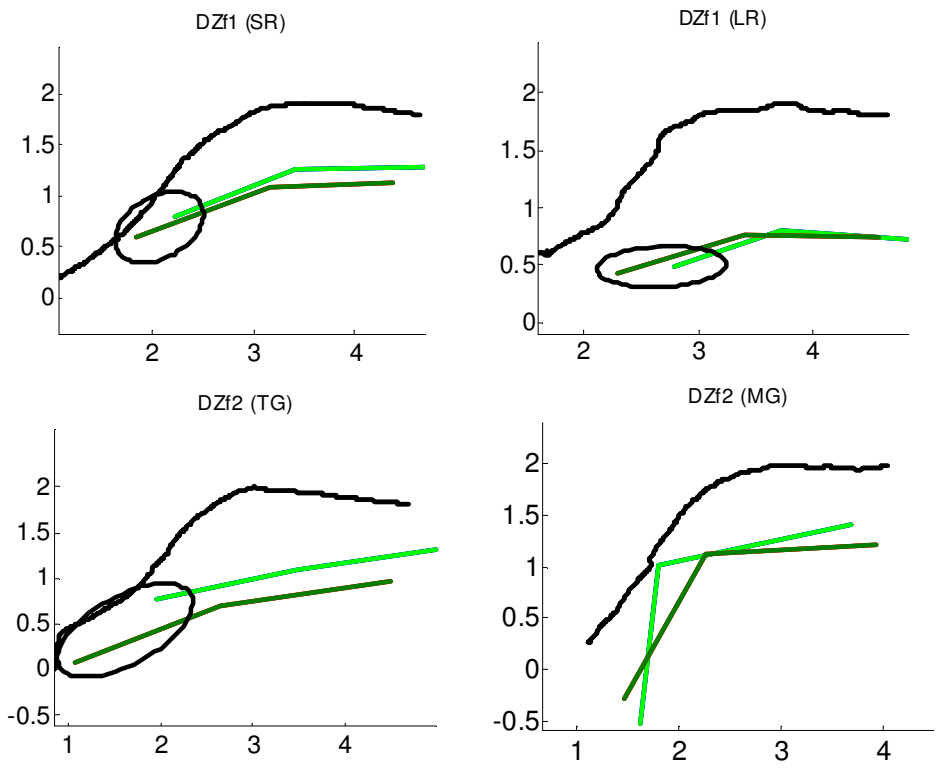


Figure 33: Mean articulatory target positions of /s/ (dark green) and /ʃ/ (light green) of the DZ twins; different plots show different speakers, ellipses visualize the amount of (horizontal and vertical) variation in the tongue tip, no ellipse is drawn for speaker MG (of DZf2) since she did not seem to use the tongue tip.

The plots of the DZ twins can be seen in Figure 33. As reported previously, both pairs reveal differences in the shape of the palate, but DZf2 in particular show obvious differences in the size and form of the palate contour. Moreover, differences in the amount of horizontal and/or vertical variation in the tongue tip between /s/ and /ʃ/ can also be found within the pairs. The first speaker of DZf1 retracts and raises the tongue tip for /ʃ/, whereas her sister mainly retracts it. The twin pair DZf2 reveals differences in the articulation strategy: the first twin uses the usual retracting and raising strategy. Her sister (MG), in contrast, differs in the shape and doming of the tongue; it is perhaps for this reason that she uses the part between the tongue tip and tongue dorsum to build the constriction and not the tongue tip like all the other speakers. In general, the production of the sibilants is quite similar for all speakers except MG. As discussed in the beginning of the chapter Toda (2006) points out that two articulatory strategies exist to realize the /s/-/ʃ/ contrast, namely the *tongue position strategy* and the *tongue adjustment strategy*. The use of the first strategy can be observed in nearly all of our

speakers: for the production of /ʃ/ the tongue is only retracted, without any differences in the shape and doming of the tongue (this is often accompanied by a protrusion of the lips, but this cannot be compared with the production of /s/ in our data, as lip protrusion is also assumed in the production of /s/ due to the preceding rounded vowel). In Figures 32 and 33 it can be seen that all speakers but MG use this strategy: the tongue is in a more fronted position during /s/ (dark green) and retracted for /ʃ/ (light green), but the shape of the tongue is more or less straight (depending on the speaker or pair), and stays in this form in both conditions. Speaker MG realizes the sibilants by using an articulation strategy that is more similar to the abovementioned tongue adjustment strategy, where the tongue is more domed for /ʃ/.

To quantify the differences in how the speakers realize the /s/-/ʃ/ contrast, the following graph visualizes the percentage of horizontal (red: dist_ttipX) and vertical (green: dist_ttipY) variation in the tongue tip for the production of /s/ and /ʃ/. Speakers of the same pair are again plotted next to each other. The four MZ pairs on the left side of the black line reveal very similar relative percentages of horizontal and vertical variation, but both DZ pairs show differences, as indicated by the arrows.

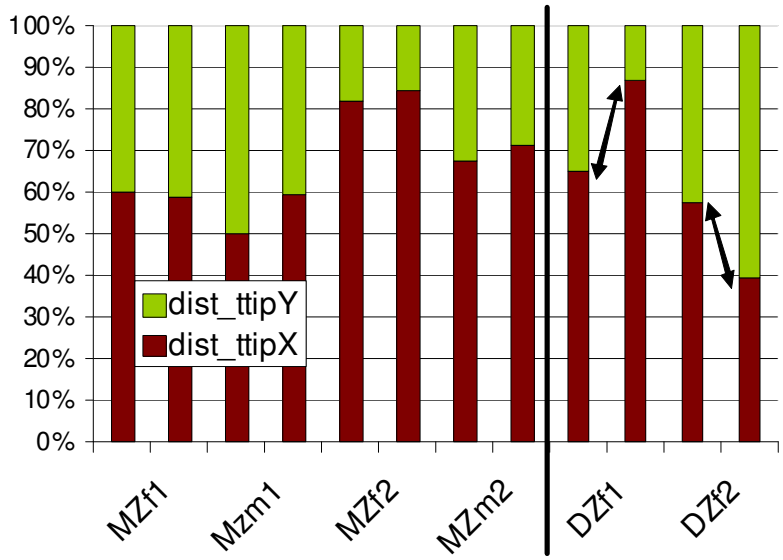


Figure 34: Percentage of horizontal and vertical variation in tongue tip in realizing /s/ and /ʃ/ for each speaker. Speakers of the same twin pair are plotted next to each other; the black line separates MZ and DZ pairs; the arrows mark the large difference between the DZ speakers.

The results of this analysis support hypothesis 1 and the influence of physiology (NATURE) on the articulatory realization of the /s/- /ʃ/ contrast, specifically, on how the tongue tip position varies between the two sibilants in a horizontal and/or vertical direction. A further interpretation of these results is that speakers with steep palates may have to raise their tongue tip to produce /ʃ/ in contrast to /s/, whereas speakers with flat palates mainly retract the tongue and only horizontal variation may be necessary.

5.1.6 Conclusion

No strong evidence could be found for the influence of physiology (NATURE) on the mean articulatory TARGETS of /s/ and /ʃ/, which leads to a corroboration of hypothesis 2. A tendency towards a greater influence of physiology on the production of /ʃ/ than of /s/ could be observed, but the results failed to show significance. MZ twins are not significantly more similar in their articulatory targets of /s/ and /ʃ/ than DZ twins. However, results indicate an influence of NATURE and physiology on the realization of the /s/-/ʃ/ CONTRAST, supporting hypothesis 1. The degree of horizontal and/or vertical variation between the productions of /s/ and /ʃ/ depends on the physiology of the speaker; in particular, this has been interpreted with respect to their palatal shape. This leads to the assumption that speaker-specific characteristics in articulation based on differences in physiology are more salient in the realization of phoneme contrasts (and perhaps gestures) than in the articulatory targets themselves.

5.2 Acoustic inter-speaker variability in sibilants

In this section the influence of NATURE and NURTURE on speaker-specific acoustic outputs of sibilants is examined. Therefore, as a first step, an evaluation of the possible parameters that characterize the acoustics of sibilants is made. Several acoustic parameters have been found to play a role in the acoustics of sibilants (see, among others, Hughes & Halle 1956, Forrest et al. 1988, Jongman et al. 2000, Newman 2003). Besides crucial spectral properties of /s/ and /ʃ/, noise duration and amplitude are also important parameters, and the spectral properties of the transition between the fricative and the following vowel have been shown to be significant as well (Gordon et al. 2002). Most researchers (Hughes & Halle 1956, Forrest et al. 1988, Evers et al. 1998, Jongman et al. 2000, Jones & Munhall 2003, Newman 2003) have focused on spectral properties and agree in the assumption that the frication noise is the primary acoustic parameter for distinguishing /s/ and /ʃ/. While there is an emphasis on the spectral properties of the sibilants, there is still discussion on the appropriate way to measure these spectral properties. Several acoustic parameters have been revealed as being significant in distinguishing the spectral properties of the sibilants: the centroid or *Center of Gravity* (COG) is the mean frequency of the spectrum with the highest energy (Forrest et al. 1988); *skewness* describes the energy distribution over the whole frequency range of the spectrum (Forrest et al. 1988); *kurtosis* reveals the peakedness of the distribution (Forrest et al. 1988); the parameter *PEAK* gives the frequency with the highest amplitude (Newman 2003, Jongman et al. 2000); and *slope* describes the rise or fall of the spectral envelope (Evers et al. 1998, Jones & Munhall 2003).

While there is obviously discussion in the literature about the most important acoustic parameter for distinguishing /s/ and /ʃ/, it seems to be even more difficult to find acoustic parameters that can differentiate between the productions of different speakers. However, this is the most relevant issue for the present analysis. Specific acoustic variability has been shown in various studies. A cross-linguistic acoustic study of voiceless fricatives in seven languages was conducted by Gordon et al. (2002). Duration, center of gravity, and overall spectral shape were investigated. Moreover, formant transitions from adjacent vowels were analyzed for a subset of the data. The authors found the overall *spectral shape* and the *coarticulation effects on formant transitions* to adjacent vowels to be the most important parameters

in differentiating the fricatives. Most importantly, they found speaker-specific acoustic variability in terms of *average spectra* and *average center of gravity* in all fricatives but especially in /s/.

Flipsen et al. (1999) give a comprehensive overview of 21 studies carried out during the last 40 years analyzing the acoustics of /s/ (among others Hughes & Halle 1956, Bauer & Kent 1987, Nittrouer et al. 1989, Katz et al. 1991, Tjaden & Turner 1997). From their review it becomes clear that the studies vary in many methodological and experimental parameters, such as the sample size, the speakers' ages, the linguistic context, the sampling frequency, the analysis range, and, as a result, also in the measured parameters PEAK and centroid. Flipsen et al. (1999) emphasize that all of these factors can have a significant effect on the inter-speaker variability found; nevertheless only a few of the reported studies controlled for these factors. For example, some studies have found different oral cavities between male and female speakers (Daniloff et al. 1980) which leads to differences in the acoustics of /s/. Furthermore, more aspiration noise has been found for female speakers (Klatt & Klatt 1990). In addition to physiological details, social factors and gender-specific aspects can affect the acoustics of /s/. Strand (1999) showed in her study that listeners identified sibilants differently depending on what they thought the speaker's gender was, pointing to stereotyped perception of social grouping. Research on gay speech and here the phenomenon of lisp also reveals the influence of social factors in speakers and listeners on the acoustics and perception of /s/ (Levon 2006, Munson 2007, Mack 2011, Bowen 2002).

Returning to the study of Flipsen et al (1999), their goal was to create an acoustic reference data base which addresses the diverse ways in measuring the acoustics of /s/. They recorded 26 adolescents (from 9 to 15 years old) and several measurement issues were examined. Among other things they investigated the first four spectral moments (mean or *centroid*, *standard deviation, skewness*, and *kurtosis*, following Forrest et al. 1988). They found that the best acoustic characterization of /s/ is obtained from the midpoint of the sibilant, represented on a linear scale, it is parameterized by the first and third spectral moment (i.e. mean and skewness), influenced by phonetic context, and also by sex.

In the production and perception study of Newman et al. (2001) the degree of acoustic variability among 20 English speakers in their /s/ and /ʃ/ productions was measured. They

also followed the method proposed by Forrest et al. (1988) and measured the first four moments of the spectral distribution. It again turned out that *centroid* and *skewness* were particularly useful for distinguishing the phonemes for all speakers. Results showed that the fricative's noise spectrum varies speaker dependently: inter-speaker variability in the acoustic parameters was apparent for both phonemes, but a higher degree of variability could be shown for /s/. The speakers also revealed variability in their acoustic distinction between /s/ and /ʃ/: some speakers produced sibilants with overlapping *centroids* and *skewness* values, while others showed /s/ and /ʃ/ productions that were quite distinct from one another. The authors were also able to link the differences in intra-speaker variability regarding the /s/-/ʃ/ distinction with the individual perceptual ability of distinguishing them.

The study of Ghosh et al. (2010) investigates the relation between auditory acuity, somatosensory acuity and the magnitude of the produced /s/-/ʃ/ contrast. They used several plastic domes with grooves of different spacings to measure the speakers' somatosensory acuity of the tongue tip. Auditory acuity was determined by calculating the subjects' *just noticeable difference* (JND), which corresponds to the difference in spectral mean of two synthetic stimuli that were distinguished by the participant. In addition, they measured the Euclidean distances between each speaker's /s/ and /ʃ/ production in a 3D space defined by the acoustic parameters *mean, skewness* and *kurtosis*. They found speaker-specific behavior in producing the sibilant contrast depending on the speakers' auditory and somatosensory acuity, pointing to the role of auditory and somatosensory goals in sibilant production.

In a follow-up study to Newman et al. (2001), Newman (2003) further investigated the correlation between speech perception and production to evaluate the importance of different acoustic parameters. The author emphasizes that it is often not quite clear which auditory cues and/or acoustic parameters listeners use for different phonemic distinctions, since many parameters are correlated, and thus it is difficult to distinguish between them. To investigate the auditory cues listeners use, several acoustic parameters of different phoneme categories were measured and correlations between the listeners' perceptual prototypes and their average productions were made and analyzed. For stop consonants, VOT turned out to show high correlations, whereas for voiceless fricatives *spectral PEAKS* showed the highest correlations with listeners' ratings. Thus, PEAKs seem to be a crucial parameter for listeners' perceptual categorizations as well as their productions of different fricatives. Other studies have also

shown the existence of a link between the production and perception of sibilants and that differences in production might be related to differences in perception (see Section 1.3, and among others Perkell et al. 2006, Jones & Munhall 2002). Thus, auditory goals that arise from different learning conditions and social environments can be crucial for inter-speaker variability.

In the literature, there is agreement about the fact that the spectral characteristics of the fricative noise are the most important acoustic cue in sibilants, and several parameters have been used to measure them (among others, centroid, PEAK, skewness, kurtosis and overall spectral shape). An analysis of all these parameters would go beyond the scope of this study, thus it has been necessary to concentrate on certain parameters. The acoustic parameters that are often used in relevant studies and which have been shown to be significant are COG (or centroid/mean), PEAK, and the overall shape of the mean spectrum. Therefore, the further analysis focuses on these acoustic parameters. Furthermore, to parameterize the shape of a spectrum a ***D****iscrete* ***C****osine* ***T****ransformation* (DCT, Watson & Harrington 1999, Guzik & Harrington 2007) is made. Guzik & Harrington (2007) showed in their study that the DCTs provide a very effective separation between the four fricative types in Polish. Jannedy et al. (2010) found DCTs to be a reliable parameter to differentiate the very similar acoustic spectra of /ç/ and /ʃ/ in Berlin German. The DCT values can roughly be compared with the spectral moments of Forrest et al. (1988). A more detailed description of how the DCTs are determined is given in the method section of this chapter.

One issue in speech research is the question as to wheather coarticulatory strategies are more idiosyncratic and physiologically determined than targets (Kühnert & Nolan 1999, cf. Section 1.3). It has been mentioned previously that transitions between sibilants and vowels are relevant (Gordon et al. 2002). Regarding twins' speech, studies have found that coarticulation parameters are more similar in MZ twins than in normal siblings or unrelated speakers (for coarticulatory behaviour in /r/ and /l/ see Nolan & Oh 1996, for coarticulation patterns in terms of F2 vowel onsets and F2 vowel targets in /b d g h/-V sequences see Whiteside & Rixon 2003, cf. Section 2.2.2.2). Coarticulatory behavior in sibilant-vowel sequences has not been analyzed so far in twins. Therefore, the following analysis takes this into account as well and investigates both sibilant targets and transitions.

5.2.1 Hypotheses

As we have seen in the introduction, inter-speaker variability can have a variety of different causes, including social or gender-specific factors. Note that these factors will not be taken into account in this analysis, as the main emphasis of this study is the NATURE-NURTURE issue. Regarding this topic, speaker-specific acoustic parameters of sibilants 1) may be due to differences in articulation that arise from differences in the physiology of the speakers or 2) may be influenced by different auditory targets. Following these assumptions two alternative hypotheses (H1a and H1b) can be made:

H1a: Acoustic inter-speaker variability is due to differences in physiology. (NATURE has an influence on the acoustic characteristics of /s/ and /ʃ/.)

> MZ twins are *more similar* than DZ twins in the acoustic characteristics of their sibilants (because MZ twins are more similar in their physiology than DZ twins).

H1b: Acoustic inter-speaker variability is due to differences in learned auditory targets. (Auditory targets (and NURTURE) are the most important factor determining the speaker-specific characteristics in the production of /s/ and /ʃ/. Physiology only plays a minor role.)

> MZ twins are *as similar* as DZ twins in the acoustic characteristics of their sibilants (because MZ and DZ pairs share the same degree of their social environment).

In addition, the possible different impacts of targets on the one hand and transitions on the other hand with regard to speaker-specific characteristics are investigated and the following assumption (H2) is made:

H2: Speaker-specific physiology is mainly reflected in speaker-specific TRANSITIONS, since transitions are subject to biomechanical restrictions of the individual speaker's physiology.

> MZ twins are *more similar* than DZ twins in transitions but not in targets.

5.2.2 Method

To investigate the acoustic outputs of /s/ and /ʃ/ and to look for differences within the twin pairs, the parameters COG (Center of Gravity) and PEAK were measured. COG gives a value for the spectral mean of the fricative. It is a correlate of the place of articulation and the length of the front cavity (Hughes & Halle 1956): COG is higher for front tongue articulations and is thus expected to be higher for /s/ than for /ʃ/. PEAK is not a mean value but the frequency with the highest amplitude. As COG gives a mean value and averages over frequencies, it is sometimes not very precise and reliable, since it will give a mean frequency between several maxima or within a plateau; therefore PEAK was used as a second parameter. To investigate the overall shape of the fricative spectra the mean spectra of /s/ and /ʃ/ for each speaker were compared in form and steepness within the twins and correlations were calculated. Furthermore, a ***D****iscrete* ***C****osine* ***T****ransformation* (DCT) was conducted, which decomposes the speech signal into a set of half-cycle frequency cosine waves (Watson & Harrington 1999). The resulting amplitudes of the different cosine waves are the DCT coefficients and correspond to the cepstral coefficients of a spectrum. Thus, the first three DCT coefficients are proportional to the mean of the spectrum, its linear slope and curvature. Average DCT coefficients were calculated for each speaker and sibilant and compared within the twins.

Following the articulatory approach above, the acoustic realization of the phoneme contrast was investigated. To do this, Euclidean distances were calculated between the sibilants – defined by a) the acoustic parameters COG and PEAK and b) the DCT coefficients that parameterize the spectra – and compared within the twin pairs. In addition, coarticulatory parameters were taken into account: transitions between the sibilants and the following vowel (schwa) were analyzed by measuring formants at the end of the fricative and 40 ms after the fricative, thus within the vowel. The differences between the formant transitions were calculated and compared between the speakers.

5.2.2.1 Speech material

The speech material is the same as in the articulatory analysis (see Section 5.1): /ʃ/ was taken from the target word /vaʃə/ and /s/ from the target word /kʏsə/, both embedded in different carrier sentences (*Ich wasche Hagi/Haga/Hagu/Haku im Garten, Ich küsse Kiba/Giba/Guba im Garten*). For the analysis of the formant transitions between the sibilant /s/ and the following schwa, only the sentences with *Kiba* and *Giba* were taken into account to avoid any coarticulatory effects due to different vowel contexts. Again, the numbers of analyzed items vary among the speakers due to excluded measurement errors (i.e. 14.7%). Note that for DZm1 only 12 renditions of /s/ could be recorded. Nevertheless, for the acoustic analysis data from all seven twin pairs (including DZm1) could be analyzed.

Table 19: Number of analyzed items for /s/ and /ʃ/ for each speaker (differing number of analyzed items for transitions for /s/ in brackets) with mean and standard deviation (SD).

	Number of analyzed items	
Speaker	/s/ (for transitions)	/ʃ/
MZf1a	29 (20)	40
MZf1b	38 (22)	38
MZf2a	37 (20)	31
MZf2b	44 (19)	37
MZm1a	40 (20)	38
MZm1b	37 (20)	34
MZm2a	28 (24)	33
MZm2b	25 (15)	29
DZf1a	29 (17)	34
DZf1b	34 (17)	28
DZf2a	30 (18)	28
DZf2b	31 (18)	35
DZm1a	12 (12)	27
DZm1b	21 (21)	36
MEAN	31.1 (18.8)	33.4
SD	8.29 (2.9)	4.23

5.2.2.2 Measuring COG and PEAK values

First, segments of produced /s/ and /ʃ/ were labeled and annotated as described in Section 3.4.1. Then, start-, end- and midpoints of the labeled intervals were calculated. COG values were measured at intervals of 30 ms around the midpoint of the segmented sibilant (sampling frequency = 22000 Hz, lower limit: 2000 Hz, upper limit: 6500 Hz for /ʃ/ and 8000 for /s/). The most prominent PEAK values were measured in the segmented sibilant between 2000 Hz and 8000 Hz for both sibilants. Mean COG and PEAK values were calculated for each speaker. ANOVAs for each sibilant with speaker as independent variable and the parameters COG and PEAK as dependent variables and post hoc Tukey tests were carried out to look for within-pair variability (cf. Tables C.3 and C.4 in the appendix for details).

5.2.2.3 Analysis of spectra

Since PEAK and COG are measured (average) frequencies and can only explain part of the characteristics of the acoustic outputs of sibilants, mean spectra were also analyzed. These spectra show the overall envelope of the frequency-amplitude distribution, averaged over all repetitions of the respective speakers. Pearson correlation coefficients were calculated for the envelope of the mean spectra between the two speakers of each twin pair[13]. Moreover, three DCT coefficients were calculated for each speaker and repetition of the two sibilants and mean DCT values were compared between the twins. Figure 35 gives an example of the mean spectra of /ʃ/ for the two speakers of the twin pair DZf2 (MG = blue, TG = red) and shows two of the three corresponding cosine waves that are used to measure the DCT coefficients. DCT1 (which is based on a half-cycle cosine wave, left graph of Figure 35) corresponds to the direction and the magnitude of tilt of the spectrum, i.e. the slope. DCT2 (which relates to a whole cycle cosine wave) is a measure of the spectrum's curvature or the degree of u-shape. Thus it is negatively correlated to kurtosis (which measures the peakedness of a spectrum). DCT3 (which corresponds to a one and a half cycle cosine wave, right graph of Figure 35)

[13] Note that such a procedure violates the assumption that samples are independent of each other. However, here, the correlation coefficients are only used as an exploratory tool to describe the spectra in a qualitative way. To quantify the comparisons the DCT-values (which are most appropriate to parameterize the spectra) are used. Note though, that the DCTs overall confirm the heights of the correlations.

gives an idea of the amplitude of the higher frequencies of the analyzed spectra. For a spectrum with a second peak – like that of speaker MG (blue) – the coefficient of DCT3 will be higher than for a spectrum with just one peak and a monotonous falling slope – like that of her sister TG (red).

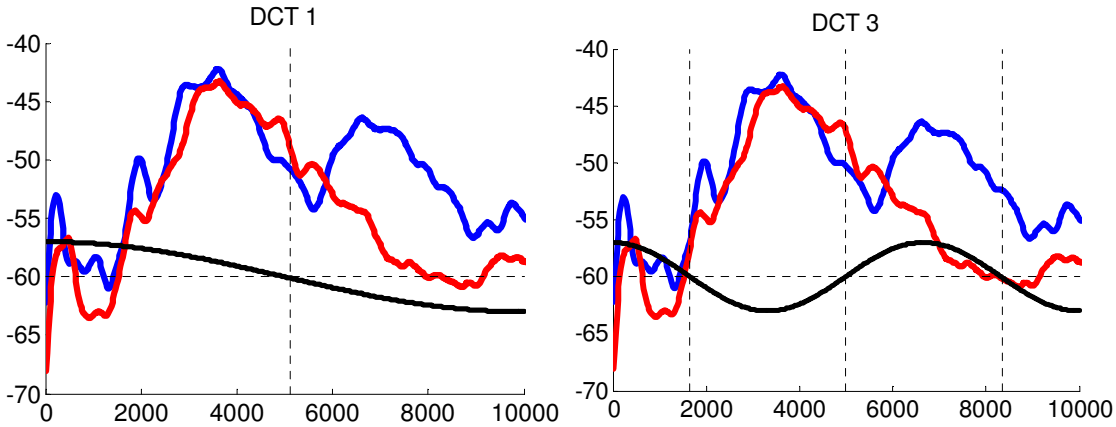


Figure 35: Mean spectra of /ʃ/ for the two speakers of DZf2 (MG = blue, TG = red) and cosine wave forms that correspond to DCT1 (left-hand graph) and DCT3 (right-hand graph) coefficients.

5.2.2.4 Analysis of sibilant-vowel TRANSITIONS

Apart from the acoustic target parameters COG and PEAK, and the calculated DCTs, the characteristics of the transitions between the sibilant and the following vowel were also investigated since it was hypothesized that coarticulation parameters and transitions are more speaker specific than targets (see H2). MATLAB and PRAAT scripts written by Martine Toda were used. Five formants were measured automatically over the whole utterance including the sibilant and the following schwa with the following adjustments: sampling interval 0.010 s, maximum frequency 5500 Hz, window length 0.025 s, pre-emphasis from 50 Hz. After that, formant plots of the utterance were made for a time window from 0.5 s before to 0.5 s after the end of the fricative. The end of the fricative was aligned for all repetitions of one speaker. In this way, the formant transitions between the sibilant and the following schwa for each speaker could be investigated graphically. To compare the transitions statistically the formants at the end of the fricative (formantTransition) and in the vowel target (formantTarget = 40 ms after the end of the sibilant) were measured. The differences between formantTransition

and formantTarget for F2 and F3 were calculated and compared between the speakers of each twin pair.

5.2.3 Results of the acoustic analysis of sibilant TARGETS

Since it is hypothesized that auditory targets are crucial for speech production also regarding the speaker-specific variability in the acoustics of sibilants, **H1b** (cf. Section 5.2.1) will be taken as our assumption.

H1b: Acoustic inter-speaker variability is due to differences in learned auditory targets. (Auditory targets (and NURTURE) are the most important factor determining the speaker-specific characteristics in the production of /s/ and /ʃ/. Physiology only plays a minor role.)

> MZ twins are *as similar* as DZ twins in the acoustic characteristics of their sibilants (because MZ and DZ pairs share the same degree of their social environment).

5.2.3.1 Inter-speaker variability in the acoustic TARGETS of /s/: COG and PEAK

COG and PEAK values were measured for each speaker. A one-way ANOVA was conducted with SPEAKER as independent factor, and a post hoc Tukey test served to look for significant differences within the twin pairs (cf. Tables C.3 and C.4). The following table gives information about the mean COG values, standard deviations (SD), number of items (n) and p-values regarding inter-speaker variability within the pairs. Only MZf1, DZf1 and DZm1 show no significant differences in their measured COG values. The significant difference in COG for MZf2 reflects the different articulatory strategies between the sibilings discussed in section 5.1.3.1. Note that due to the fact that the parameter COG is an average measure, we have to be aware of problems that can arise when the spectrum of the respective fricative consists of two peaks. Then, the COG value will mirror the frequency in the middle of these peaks, although the amplitude of this frequency can be lower than those of the neighboring frequencies. The high standard deviations of over 500 Hz of speaker SL of MZm1 (626 Hz), MG of DZf2 (586) and HF of MZf1 (507) could indicate a spectrum with two equally high peaks or a plateau, for which the measurement of the COG value is difficult and ambiguous. Thus, in a next step the parameter PEAK is analyzed.

Table 20: Mean values and standard deviations for COG for /s/ of each speaker; significant differences between twins in bold.

Twin pair	mean COG (Hz) Twin A – Twin B	SD (Hz) Twin A – Twin B	n Twin A – Twin B	p-value (adj.)
MZf1 (afhf)	5176 - 5457	342 - 507	29 - 38	0.280
MZf2 (gsrs)	4457 - 4939	414 - 202	39 - 44	**< .001**
MZm1 (slcl)	5045 - 4537	626 - 365	38 - 40	**< .001**
MZm2(mima)	3978 - 4466	247 - 416	33 - 38	**< .001**
DZf1 (srlr)	5096 - 4920	454 - 355	34 - 31	0.923
DZf2 (tgmg)	4154 - 4795	434 - 586	34 - 30	**< .001**
DZm1 (fmhm)	4782 - 5185	284 - 374	12 - 21	0.318

The differences in PEAK values between the speakers of one twin pair can be found in Table 21. The table gives information about the mean PEAK values, the variation in the measured PEAKs (SD) and the level of significance of differences within a twin pair. The standard deviations of the PEAK values of some speakers are very high and point to a bimodal distribution of the measured PEAKs. These speakers seem to show a spectrum with two or more PEAKs. Nevertheless, the statistical analysis takes the high standard deviations into account, and only one pair, MZm1, shows significant differences ($F(13, 433) = 11.56$, $p < .05$).

Table 21: Mean values and standard deviations for PEAK for /s/ of each speaker; significant differences between twins in bold.

Twin pair	Mean PEAK (Hz) Twin A – Twin B	SD (Hz) Twin A – Twin B	n Twin A – Twin B	p-value (adj.)
MZf1 (afhf)	5385 - 6065	597 - 1269	29 - 38	0.388
MZf2 (gsrs)	4533 - 4826	1016 - 372	37 -44	0.995
MZm1 (slcl)	5414 - 4564	1308 - 456	38 - 40	**0.042**
MZm2 (mima)	4090 - 3628	258 - 2318	27 - 38	0.919
DZf1 (srlr)	4885 - 5135	1305 - 252	34 - 31	0.999
DZf2 (tgmg)	4390 - 4258	694 - 1009	31 - 30	0.999
DZm1 (fmhm)	5299 - 4715	425 - 1510	21 - 12	0.970

Looking at the two parameters COG and PEAK, no influence of zygosity (and thus NATURE) can be found on the acoustic output of /s/. In correspondence with hypothesis H1b, learned auditory targets and not physiological parameters seem to be the major influence factor on speaker-specific characteristics in the production of /s/.

5.2.3.2 Inter-speaker variability in the acoustic TARGETS of /s/: Mean spectra

In the following graphs, each twin pair is shown in one figure and the two speakers of a pair are plotted in different colors (red and blue, respectively, according to the color in the articulation plots).

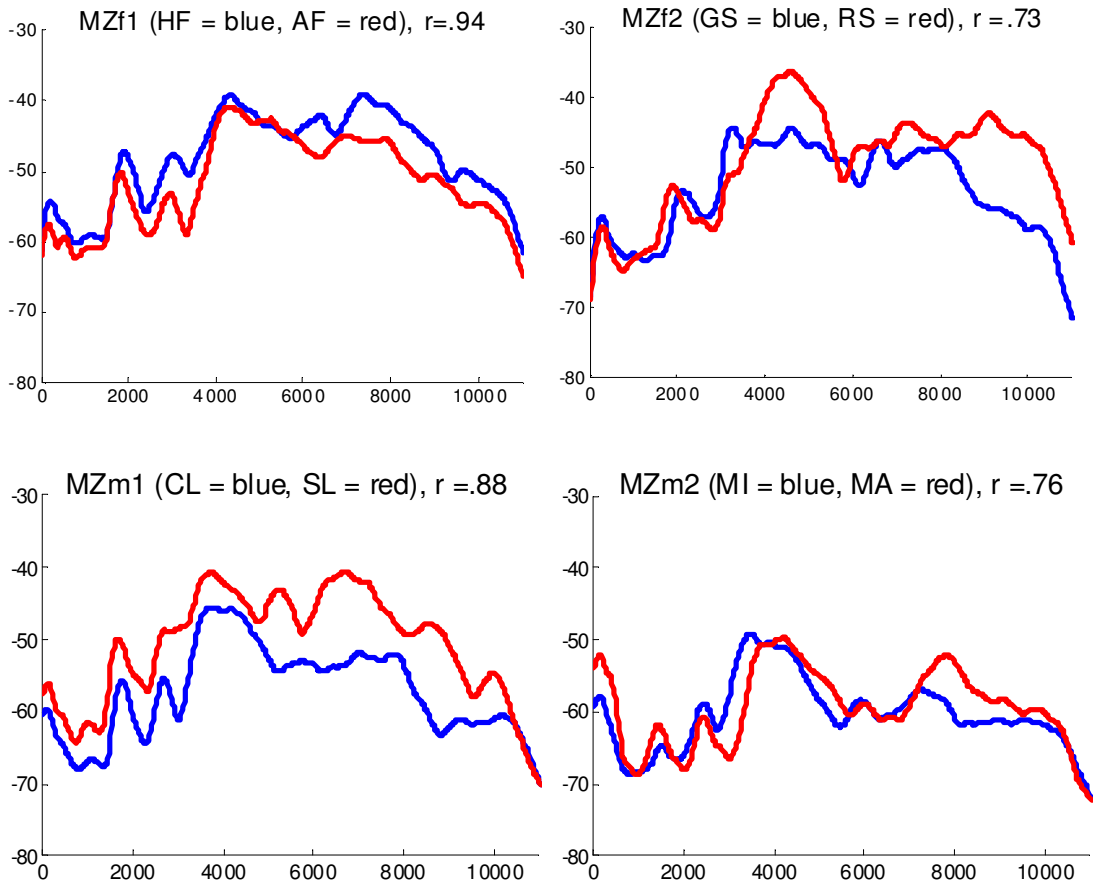


Figure 36: Mean spectra of /s/ for the MZ pairs; different speakers marked by different colors.

At first glance, the spectra of the twins look quite similar. Especially the envelopes of MZf1 seem to be nearly identical. The spectra of MZm2 also look quite alike but seem to be shifted slightly. The spectra of MZm1 are very similar (parallel) in the run of the curve, but differ in the amplitudes, which could be explained by different volumes in speaking or a different distance from the microphone. However, the latter reason can actually be ruled out since the same distance was used for all speakers. The envelopes of the other female MZ pair, MZf2, are nearly identical in the beginning but differ in the maximum amplitude. GS (blue) shows a plateau with equally distributed amplitudes between 3000 Hz and 8000 Hz, whereas RS (red) reveals a clear peak at around 5000 Hz. This could be seen previously in Table 20 and resulted in a significant difference for the COG. (Note that differences were already obvious in the articulation strategies of this pair.)

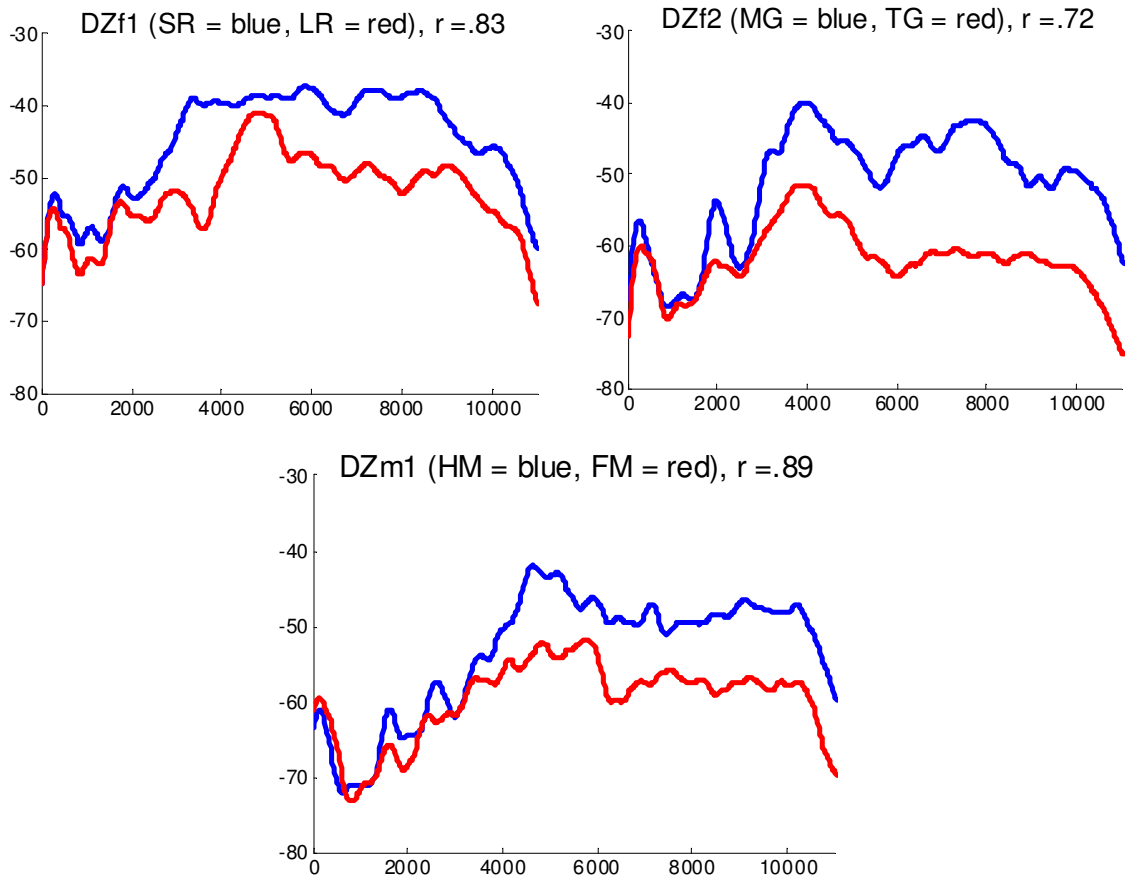


Figure 37: Mean spectra of /s/ fort the DZ twins; different speakers marked by different colors.

Both female DZ pairs (DZf1 and DZf2) reveal some differences in the shape of their spectra, especially DZf1. The spectrum of SR (blue) shows a regular amplitude of –40 dB between 3000 Hz and 8500 Hz. The spectrum of LR (red) on the other hand has a clear peak at 5000 Hz, but then reveals a similar plateau between 6000 and 9000 Hz at a slightly lower amplitude of –50 dB. The spectra of MG and TG start similarly but then differ in their steepness and amplitude. Both show a maximum amplitude at 4000 Hz; the spectrum of MG (blue) decreases after that until it reaches –60 dB and then stays even, while the spectrum of TG (red) shows another rise and peak at 8000 Hz, which resulted in a significant difference in COG but not in PEAK. The spectra of the male pair DZm1 show a similar rise at the beginning but differ in their maximum amplitude. At the higher frequencies the amplitudes again differ but both speakers reveal a plateau between 6500 Hz and 10500 Hz. To sum up, there is no clear difference in within-pair variability regarding the mean spectra of /s/ between MZ and DZ pairs. Correlations between the siblings in their mean spectral contour are very high (r = .88 - .94) for two (of the four) MZ and one (of the three) DZ pairs (MZf1, MZm1, DZm1) and moderate for the other four pairs (r = .72 - .83). Results therefore again point to H1b and learned auditory goals.

5.2.3.3 Inter-speaker variability in the acoustic TARGETS of /ʃ/: COG and PEAK

Again, statistical measurements were carried out to test for significant differences within the pairs, but this time for /ʃ/. Results are shown in Table 22 (details are given in Tables C.3 and C.4 in the appendix). DZf1 and DZm1 differ significantly in their COG values and show the greatest differences of all pairs: the mean COG values of the two speakers of each pair differ by more than 1000 Hz. Also, MZf2 reveals significance in the different COG values, even though their mean COG values vary only by about 250 Hz.

Table 22: Mean values and standard deviations for COG for /ʃ/ of each speaker; significant differences between twins in bold.

Twin pair	Mean COG (Hz)	SD (Hz)	n	p-value (adj.)
	Twin A – Twin B	Twin A – Twin B	Twin A – Twin B	
MZf1 (afhf)	3936 - 4005	161 - 254	40 - 38	0.997
MZf2 (gsrs)	3679 - 3929	252 - 157	39 - 39	**0.004**
MZm1 (slcl)	3957 - 3862	292 - 163	40 - 39	0.954
MZm2 (mima)	3523 - 3701	266 - 403	34 - 33	0.279
DZf1 (srlr)	4324 - 3497	315 - 421	39 - 42	**< 0.001**
DZf2 (tgmg)	3825 - 3743	276 - 278	38 - 28	0.994
DZm1 (fmhm)	3913 - 2926	216 - 151	28 - 40	**< 0.001**

In Table 23 it can be seen that the two speakers SR (of DZf1) and SL (of MZm1) reveal a very high standard deviation for the measured PEAK values and again two or more peaks must be assumed. Statistical tests revealed significant differences in PEAK for two of the DZ pairs but none of the MZ pairs.

Table 23: Mean values and standard deviations for PEAK for /ʃ/ of each speaker; significant differences between twins in bold.

Twin pair	Mean PEAK (Hz)	SD (Hz)	n	p-value (adj)
	Twin A – Twin B	Twin A – Twin B	Twin A – Twin B	
MZf1 (afhf)	3807 - 3593	449 - 516	40 - 37	0.985
MZf2 (gsrs)	3338 - 3740	301 - 105	39 – 39	0.363
MZm1 (slcl)	3492 - 3932	1333 - 509	40 - 39	0.209
MZm2 (mima)	3639 - 4200	713 - 379	22 - 28	0.201
DZf1 (srlr)	4213 - 2879	1407 - 666	39 -41	**< 0.001**
DZf2 (tgmg)	3841 - 3026	439 - 970	38 - 28	0.059
DZm1 (fmhm)	4244 - 2796	471 - 225	27 - 40	**< 0.001**

Regarding the hypothesis, an influence of zygosity and hence physiology (NATURE) on the acoustic parameters COG and PEAK of the sibilant /ʃ/ might be assumed since two of the three DZ twin pairs show significant differences in COG and PEAK but only one MZ pair in

the mean COG value. Hence, hypothesis H1a is corroborated and it seems that physiology plays a bigger role in the acoustics of /ʃ/ than of /s/. This reflects the results of the articulatory analysis, since a tendency could be observed towards more similar articulatory targets in MZ twins than in DZ twins in /ʃ/ but not in /s/. It is also supported by the fact that MZm2, the pair that had lived apart from each other for two years at the time of the recording, shows no difference in COG or PEAK for /ʃ/.

5.2.3.4 Inter-speaker variability in the acoustic TARGETS of /ʃ/: Mean spectra

We will now take a closer look at the mean spectra of all speakers. Figure 38 displays the spectra of the MZ pairs. Again, speakers of the same pair are plotted in the same graph with different colors. The first two graphs are also the most similar ones: the shape and steepness of the mean spectra of both female MZ pairs look surprisingly similar. GS and RS from the pair MZf2 reveal a steep rise of the envelope with a maximum amplitude of –30 dB at around 3900 Hz. Both speakers of the pair MZf1 show three little peaks over the course of an overall increase in amplitude of the spectra until –42 dB at around 4100 Hz and then a very smooth and even decrease. Pearson correlations are very high for both pairs (r = .90, .94).

The spectra of the male MZ pair MZm1 differ in shape, peak and decrease. Speaker CL (blue) reveals a maximum amplitude of –40 dB between 3500 and 4500 Hz. The spectrum of his brother SL (red) has two peaks (as was mentioned previously) at around 3000 Hz and again at around 6500 Hz. Because the frequencies with the highest amplitudes are averaged no difference was found in COG between these two speakers even though differences can be seen in their spectra: the measured COG values for both speakers are at 3900 Hz. Here, it becomes clear why the COG values may be an inadequate parameter in describing and differentiating between two spectra. Also, the analysis of the PEAK values failed to show significance: the variation (standard deviation) of SL (red) was too high because of the bimodal distribution and the averaged PEAKs of the brothers did not differ significantly. The mean spectra of MZm2 have similar shapes but seem to be shifted again: as was the case for /s/, MA (red) reveals a later peak than his brother. Nevertheless the shape of the spectra are similar and both male pairs still show quite high correlations of r = .83 and .82.

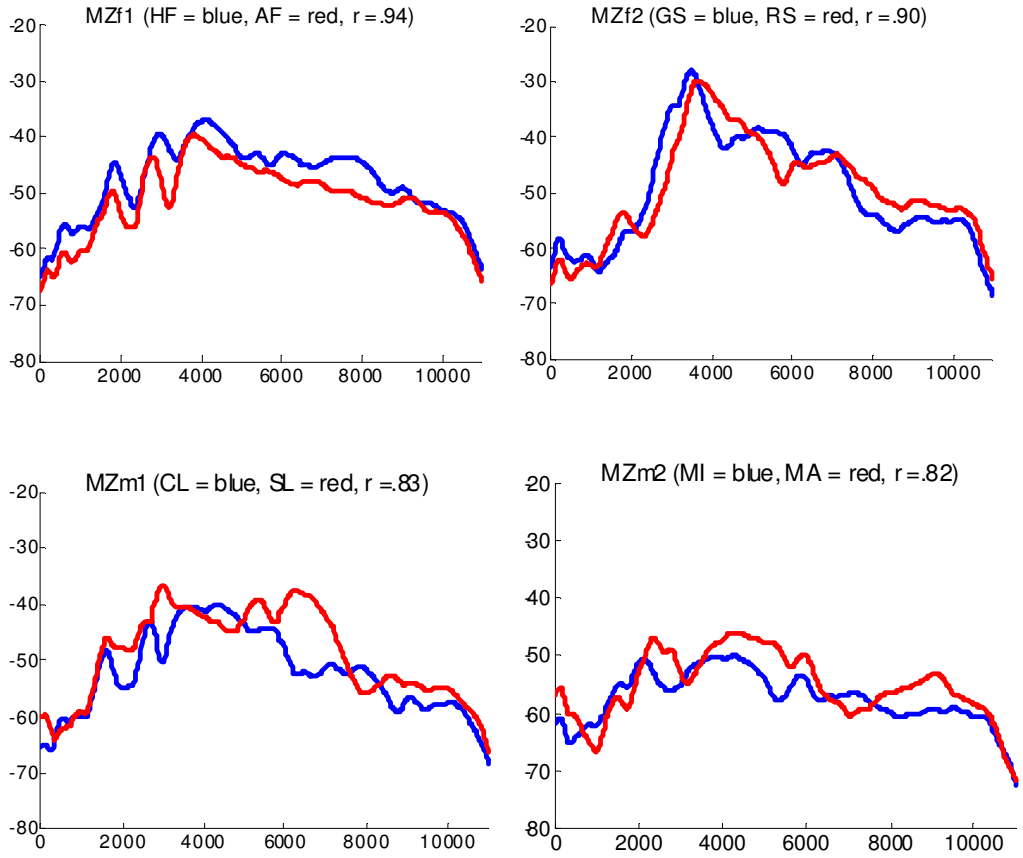


Figure 38: Mean spectra of /ʃ/ for the two female MZ pairs (above) and the two male MZ pairs (below); different speakers marked by different colors.

Fighure 39 shows the spectra of the DZ pairs. The speakers of DZf1 reveal quite similar mean spectra. Both show a rise of the envelope until 3000 Hz and after that a decrease, with a small second peak at around 5000 Hz for LR (red) and 6000 Hz for SR (blue). The pair DZf2 also shows a very similar rise at the beginning of the spectra until a maximum value of –40 dB at around 3800 Hz, but they differ then in the decline of the spectra, since TG (red) reveals a clear second peak at 7000 Hz but her sister does not. The male DZ pair DZm1 reveals a similar rise of the spectra but differences in the frequencies above 4000 Hz. Speaker HM (blue) shows a clear peak at 2500 Hz and then an even fall of the amplitude whereas his brother shows two peaks at 3000 and 5000 Hz. These differences already showed significance in the analysis of the COG and PEAK values.

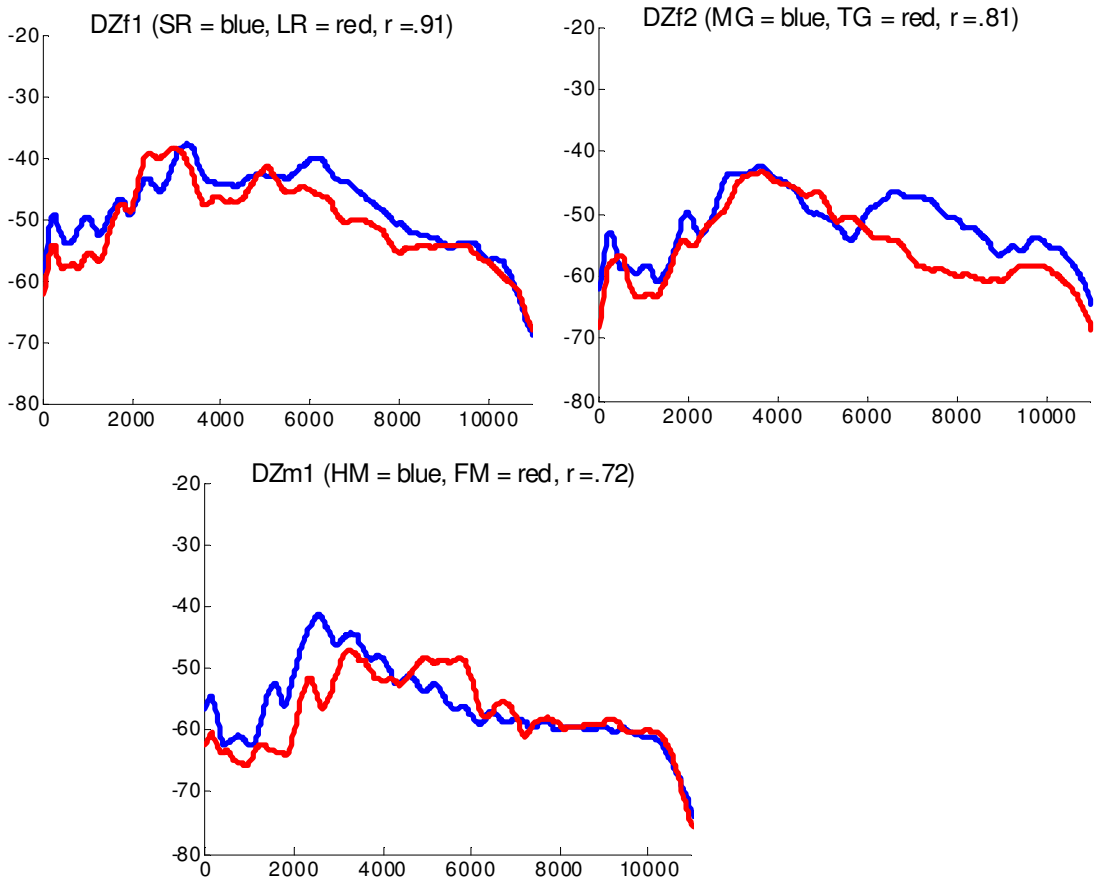


Figure 39: Mean spectra of /ʃ/ for the three DZ pairs; different speakers marked by different colors.

5.2.3.5 Inter-speaker variability in acoustic TARGETS of /s/ and /ʃ/: DCTs

DCT coefficients (DCT1, DCT2 and DCT3) were measured to parameterize the spectrum of each speaker and phoneme. Mean values and standard deviations for each speaker and phoneme were calculated and can be examined in the appendix (Tables C.6 and C.7). An ANOVA and a post hoc Tukey test revealed several significant differences in DCT values within the twin pairs (cf. Tables C.5 –C.7). For /s/, no difference between MZ and DZ twins was found since the MZ twins showed 6 out of 12 possible (4 pairs x 3 DCT values) significant differences, while the DZ twins showed 5 out of 9 significant differences (3 pairs x 3 DCT values). For /ʃ/ more differences were found for the DZ pairs: only 2 of 12 comparisons were significantly different for the MZ twins, but again 5 of 9 comparisons

showed significance in the DZ twins. Table 24 gives an overview of the significant differences found.

Table 24: Significant differences within twin pairs in three DCT coefficients for both sibilants (p < .01).

	DCT1	DCT2	DCT3
/s/	MZf2	MZf2, MZm1	MZf2, MZm1, MZm2
	DZf1, DZf2, DZm1	DZf1, DZf2	
/ʃ/	MZf2		MZm1
	DZf2, DZm1	DZm1	DZf1, DZf2

All three DCT coefficients mirror the shape of the sibilants' spectra and the acoustic output of the sibilants is best expressed when all information is taken together and not separated by each of the DCT values. Thus, two more measurements were made to define the acoustic difference between the two speakers of each twin pair. First, the mean values of all DCT coefficients were used to measure the Euclidean distance (ED) between the twins. Second, only DCT2 and DCT3 were taken into account to calculate the ED. Welch two-sample t-tests were conducted in R (version 2.9.0) with ZYGOSITY as independent factor and acoustic difference as dependent variable for each phoneme and the two measured EDs. Figure 40 visualizes the differences in ED between the twin types for both phonemes. As expected, no significant effect of zygosity was found for /s/ for either ED. However, when all three DCT values were taken into account, the DZ twins revealed obviously higher EDs than the MZ twins, although the results fail to show significance. Since the analysis so far has revealed a tendency for /ʃ/ to be more similar in MZ twins than in DZ twins, a greater influence of zygosity and thus physiology on /ʃ/ is assumed. Here again, results point to this assumption. The lower part of Figure 40 shows that the EDs for/ʃ/ differ between the twin types for both measurements: the DZ twins reveal higher acoustic differences than the MZ twins, and the ED that takes DCT2 and DCT3 into account differs significantly as a function of zygosity (mean DZ = 44.8, mean MZ = 21.1, t = 4.7916, df = 3.992, p-value < 0.01).

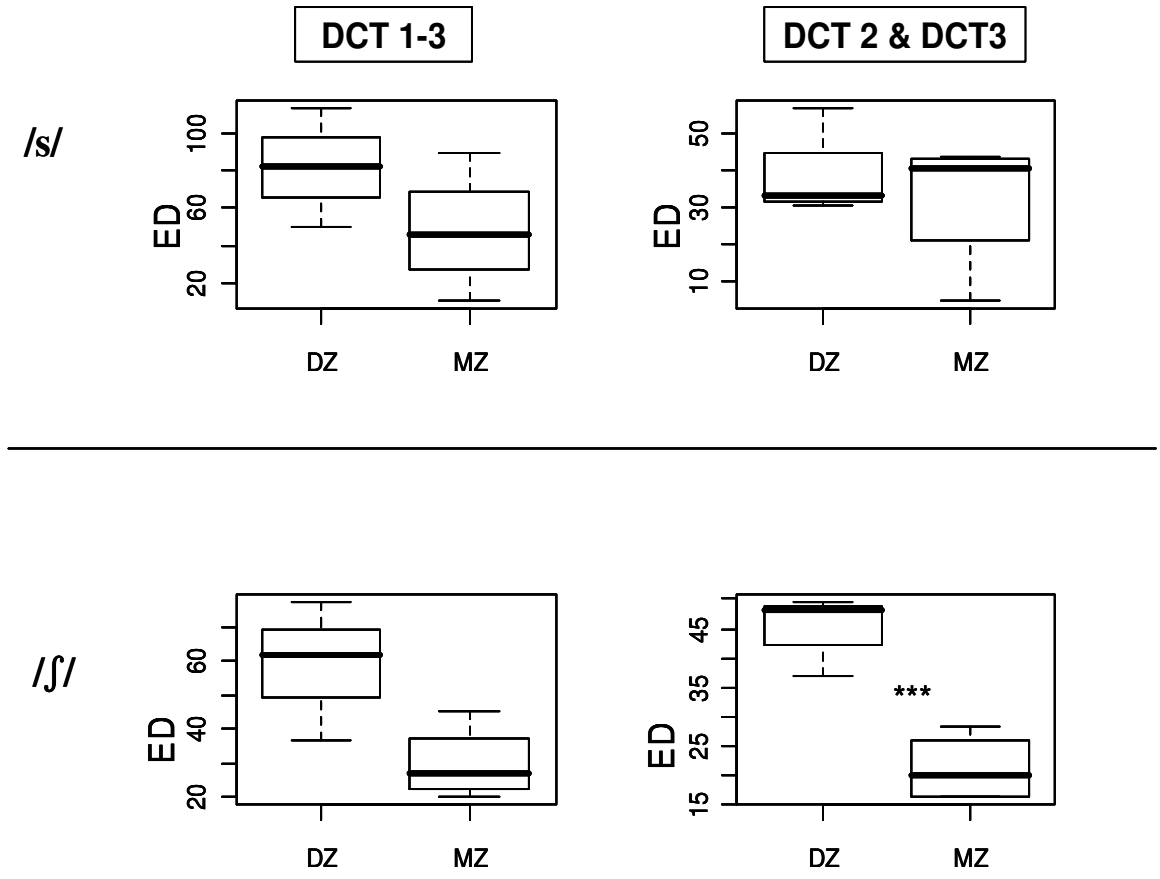


Figure 40: EDs in DCT coefficients within DZ and MZ twin pairs for /s/ (above) and /ʃ/ (below).

5.2.4 Summary of the acoustic analyses

The results of the acoustic analysis are summarized in the following table. The table shows for each twin pair the correlation coefficient r calculated between the mean spectra of the siblings and indicates significant differences in the measured COG and PEAK values for the two sibilants /s/ and /ʃ/. Moreover, the respective DCT coefficients that showed significant differences within the pairs are listed.

Table 25: Summary of the acoustic analyses with correlation coefficients (r), and information about significant (*) and non-significant (-) results in COG, PEAK and DCT coefficients.

Twin pair	/s/				/ʃ/			
	r	COG	PEAK	DCT	r	COG	PEAK	DCT
MZf1	.94	-	-	-	.94	-	-	-
MZf2	.73	*	-	1,2,3	.90	*	-	1
MZm1	.88	*	*	2,3	.82	-	-	-
MZm2	.76	*	-	3	.82	-	-	3
∅ (r)	.83				.87			
DZf1	.83	-	-	1,2	.91	*	*	3
DZf2	.72	*	-	1,2	.81	-	-	1,3
DZm1	.89	-	-	1	.72	*	*	1,2
∅ (r)	.81				.81			

No influence of zygosity (and thus shared physiology and NATURE) on acoustic similarities of /s/ can be assumed since 4/8 significant differences in COG and PEAK were found for the MZ pairs, but only 1/6 for the DZ pairs. In addition, the correlations of the mean spectra do not differ in their average height between the MZ and the DZ pairs, and significant differences in DCT coefficients were found for both twin types. In the acoustics of /ʃ/ more differences were found for the DZ pairs in all acoustic parameters. Furthermore, the average correlation (within pairs) is higher for the MZ than for the DZ pairs (r = .87 vs. r = .81), and a significant difference between MZ and DZ twins was found in terms of the acoustic difference measured by the ED based on DCT2 and DCT3 (see Section 5.2.3).

To sum up, hypothesis H1b (and hence the influence of auditory targets and NURTURE) was supported for the acoustics of /s/, but a tendency towards a greater influence of zygosity/shared physiology (NATURE) could be found for the acoustics of /ʃ/, and hence a corroboration of hypothesis H1a.

5.2.5 The acoustic realization of the /s/-/ʃ/ CONTRAST

With an approach similar to that used in the articulatory analysis it was investigated whether speaker- (or even twin-) specific behavior could be found regarding the acoustic realization of the /s/-/ʃ/ contrast. As mentioned in the introduction of this chapter, Newman et al. (2001) found inter-speaker variability in the acoustic distinction between /s/ and /ʃ/. The acoustic contrast was calculated as the average Euclidean distance (ED) between the two phonemes in a 2D space defined by mean COG and PEAK values. Figure 41 visualizes the calculated acoustic difference (ED) between /s/ and /ʃ/ for all 14 speakers. The twins are plotted next to each other and marked by the same color. Indeed, speakers differ in the amount of acoustic distinction between the two categories in terms of the investigated acoustic parameters. Some speakers reveal differences of less than 1000 Hz, while others show acoustic differences of more than 2500 Hz.

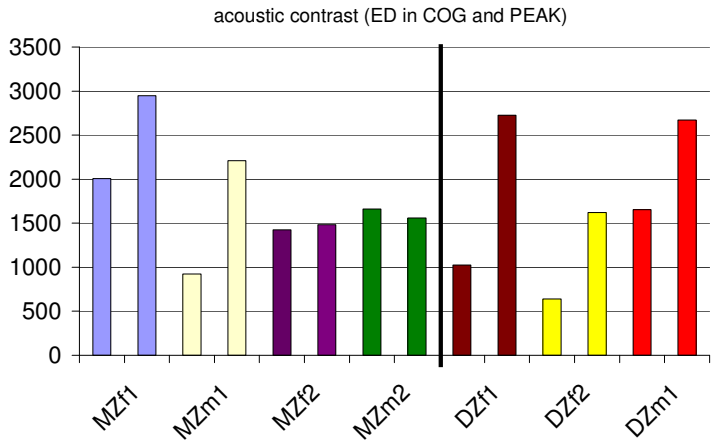


Figure 41: Euclidean distance (ED) between /s/ and /ʃ/ in a 2D space defined by average COG and PEAK values for each speaker.

From the bar plots it can be seen that intra-pair variability can be found for all DZ twin pairs and two MZ twin pairs. Thus, an influence of zygosity on the realization of the acoustic distinction between the two phonemes cannot be assumed from this analysis, since no clear difference between MZ and DZ twins is apparent. In a second step the acoustic contrast was measured in terms of the DCT coefficients. Taken together, the three DCTs give a good

parameterization of the whole spectral shape of the sibilants and thus the acoustic outputs. Thus, the ED between the phonemes in a 3D space defined by the three coefficients might be a more reliable parameter for distinguishing the overall acoustic output of the two sibilants than COG and PEAK. Figure 42 shows the calculated ED defined by the DCT coefficients for each speaker. Again, differences in the acoustic contrast can be seen for all DZ twins (especially for the male DZ twin pair) but also for one MZ pair (MZm2).

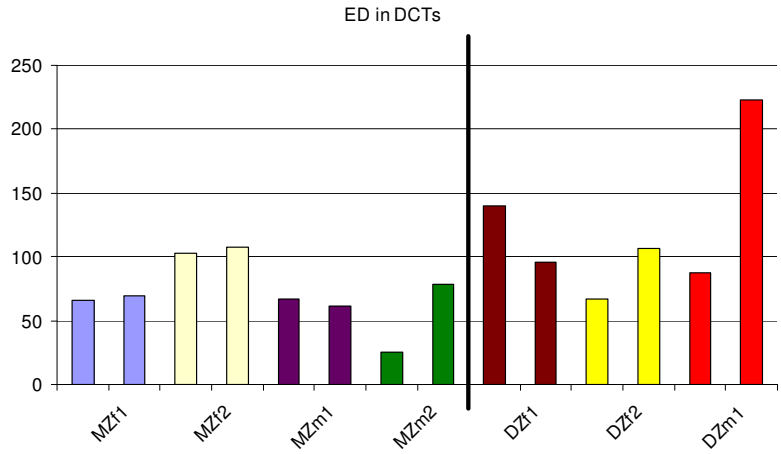


Figure 42: Euclidean distance (ED) between /s/ and /ʃ/ in a 3D space defined by DCT1, DCT2 and DCT3 for each speaker.

Thus, a tendency was found for zygosity (NATURE) to influence the acoustic realization of the phonemic contrast since MZ twins were more likely to show similar contrasts than DZ twins, but no distinct effect could be found, since MZm2 also revealed differences in the phonemic contrast defined by the DCTs and MZf1 and MZm1 showed differences in the ED defined by COG and PEAK. In addition, it is not clear what the reason might be for MZ twins' being more similar than DZ twins in their acoustic contrasts. On the one hand, shared physiology could have an impact and since MZ twins showed more similar realizations of the /s/-/ʃ/ contrast (in terms of horizontal and vertical differences in the tongue tip), this assumption is supported by the articulatory analysis. On the other hand, more similar auditory and/or somatosensory acuity is also a possible factor (Newman et al. 2001, Ghosh et al. 2010).

5.2.6 Sibilant-vowel TRANSITIONS

Transitions have been found to be speaker specific in the literature (Nolan 1983, Whiteside & Rixon 2003). Moreover, transitions are not seen as being learned but rather constitute a by-product of the necessary trajectory from one target to the next (c.f. Kühnert & Nolan 1999). Thus, hypothesis **H1a** (cf. Section 5.2.1) will be taken as our assumption.

H1a: Physiology has an influence on the acoustic transitions from the sibilant to the following vowel.

> MZ twins are *more similar* than DZ twins in the acoustic parameters of their TRANSITIONS.

5.2.6.1 Inter-speaker variability in TRANSITIONS between /s/ and a following vowel

To look for speaker-specific differences in the transitions between the sibilant and the following vowel the formants F1-F4 of the sequence /ʏsə/ from the target word /kʏsə/ were calculated for each speaker. Note that of the four carrier sentences only two were used for the analysis to reduce coarticulatory effects from differences in vowel contexts in the preceding word (cf. Table 19 in Section 5.2.2.1). Therefore, fewer renditions of /s/ were analyzed than of /ʃ/. Due to measurement errors of the formant analysis in PRAAT some files had to be excluded and the number of investigated items varies among the speakers (see also Table 19 in Section 5.2.2.1).

Figure 43 shows the resulting formant values of all repetitions for each speaker of the MZ twin pairs, plotted over the sequence /ʏsə/, starting 50 ms before the fricative and ending 50 ms after the fricative. Since the repetitions varied in their length due to intra-speaker variability in the speech rate, the sequences had to be normalized over time. The time point of the end of the fricative was taken as a reference value and set to 0 and all repetitions were aligned to this time point. This was done for each speaker separately. Each measured formant value of all repetitions of the sequence is marked by a red dot and the vertical black line in each plot refers to the time-normalized end of the fricative. Speakers of the same twin pair are plotted next to each other.

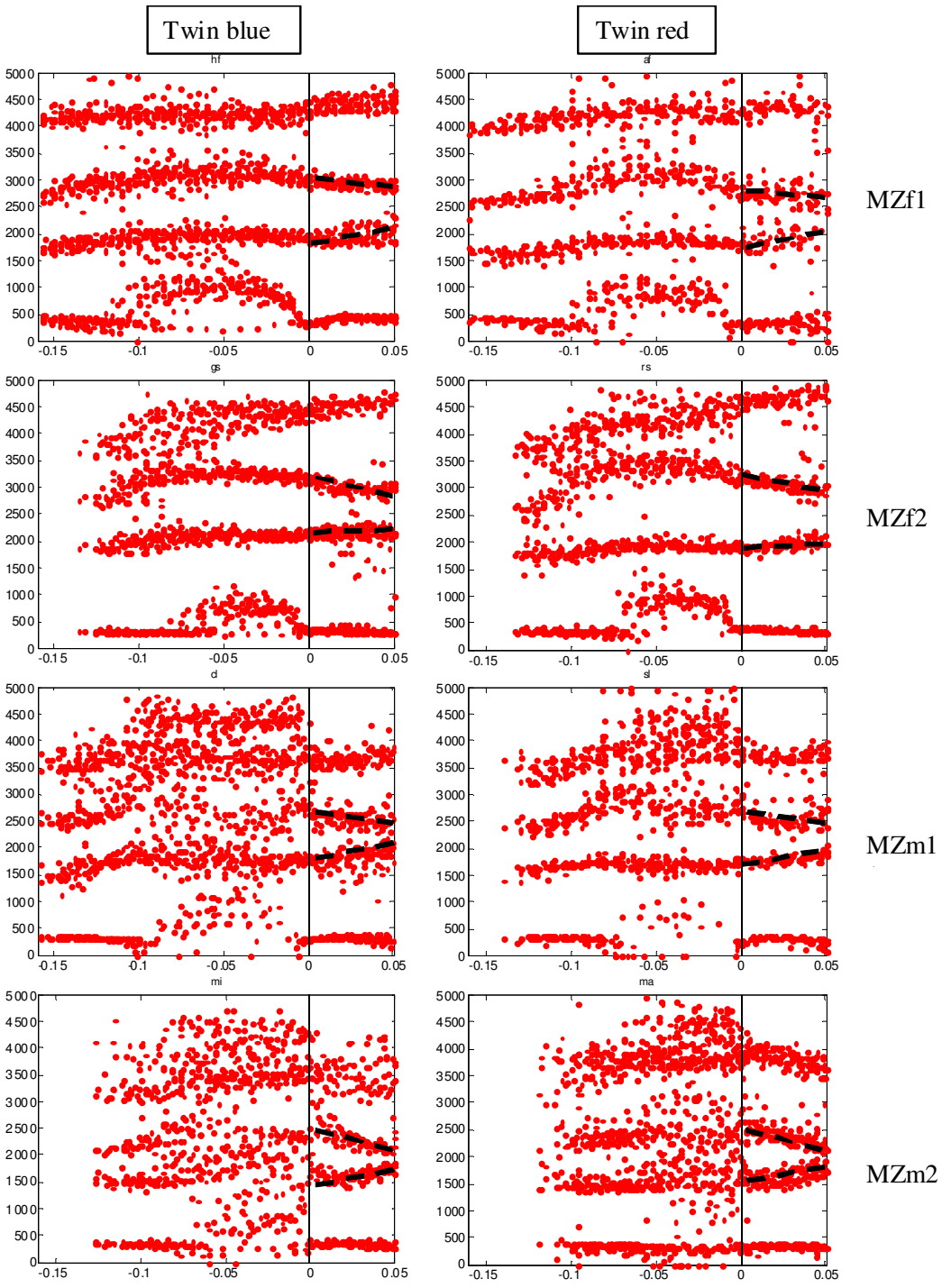


Figure 43: Transition plots of the sequence /ɪsə/ for all speakers of the MZ twin pairs; data are aligned at the end of the fricative marked by the black vertical line.

In the plots above the similarities within the MZ twin pairs in the form and shape of F1-F4 transitions over the whole sequence are apparent. Regarding all subjects, inter-speaker variability can be found in terms of the formant transitions between the end of the fricative and the following schwa. Differences can be seen especially in the F2 and F3 transitions in terms of their stability. Some speakers show clearly falling F3 transitions (both speakers of MZm2, and both speakers of MZf2), while some speakers reveal only slightly falling F3 transitions or nearly stable ones over time (speakers of the pairs MZf1 and MZm1). Interestingly these differences are only obvious between speakers who are not related, but not between speakers of the same twin pair. This supports hypothesis H1a and the influence of physiology and biomechanics (NATURE) on transitions.

The following figure shows the data for the DZ twin pairs. Here, differences in formant transitions between the sibilant and the following schwa can also be found for the two speakers of the same pair. These differences are most obvious for the male pair DZm1: twin blue (HM, left side) of this pair reveals a stable F3 transition whereas the transition of twin red (FM, right side) is clearly falling. Also, DZf1 reveals some differences: the transitions of twin blue (SR, left side) are falling or stable, while the transitions of her sister (LR, right side) are slightly rising.

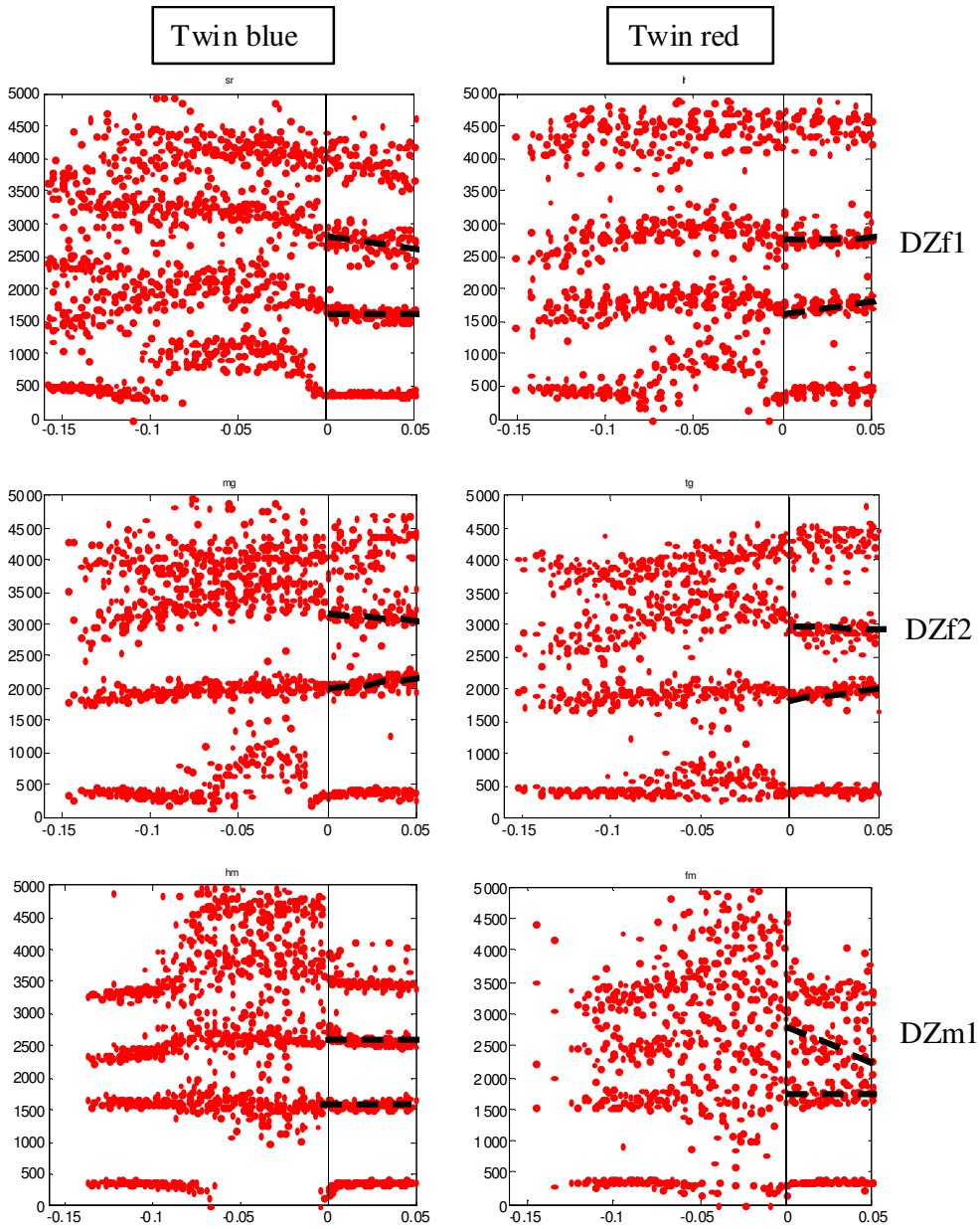


Figure 44: Formant transition plots of the sequence /ʏsa/ for all speakers of the DZ twin pairs; data are aligned at the end of the fricative marked by the black vertical line.

5.2.6.2 Inter-speaker variability in TRANSITIONS between /ʃ/ and a following vowel

Again, the formants F1-F4 for the sibilant and the following vowel were analyzed to compare the transitions within the twin pairs. Investigations were conducted on the sequence /ʃə/ from the target word /vaʃə/ for each speaker separately. All four carrier sentences could be used for the analysis since the preceding vowel context stayed constant (cf. Section 5.2.2.1). The number of investigated items that could be used for this analysis can be seen in Table 19. On average, 33.4 items could be examined for each speaker. Figure 45 shows the plots for the MZ twin pairs. Some differences in terms of the F2 and F3 transition slopes can be seen between speakers, but to a much lesser degree within twin pairs. Three of the four pairs show falling F2 transitions but the speakers of MZf1 reveal stable or even slightly rising F2 transitions. According to Fuchs & Toda (2008), falling F2 transitions (with a following non-front vowel) indicate palatalization, because the front cavity (with which F2 is affiliated) is long during the fricative and shortened towards the vowel. F3 transitions are nearly stable for all speakers, except for both speakers of MZm2, who show slightly falling transitions.

Figure 46 shows that all F2 transitions of the speakers of the DZ twins are falling. Nevertheless, some differences can be seen in the degree of the decline. Speaker blue (SR) from the pair DZf1 and speaker red (TG) from the pair DZf2 reveal the highest degree of falling F2. Also differences in F3 are obvious within the twin pairs: especially the female pairs differ in the slope of their F3 transition. Speaker red (LR) from DZf1 and speaker red (TG) from DZf2 show clearly falling F3 transitions, whereas their sisters have straight or even slightly rising F3 slopes. Although the male DZ pair shows a more stable transition of F3, some differences can be found regarding falling or rising formants. F3 transitions for speaker blue (HM) are slightly falling, and for his brother (FM) slightly rising.

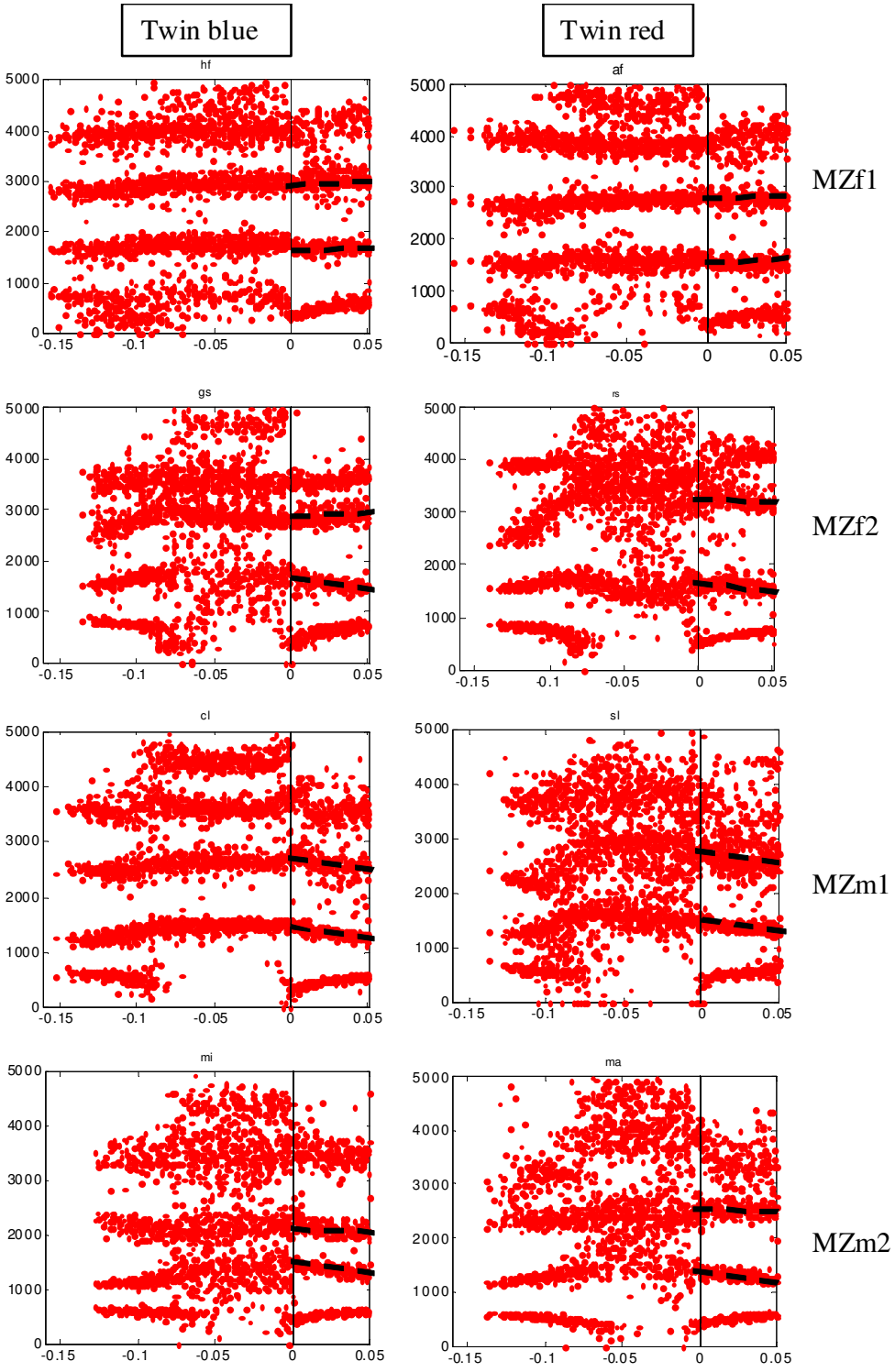


Figure 45: Formant transition plots of the sequence /aʃə/ for all speakers of the MZ pairs; data are aligned at the end of the fricative marked by the black vertical line.

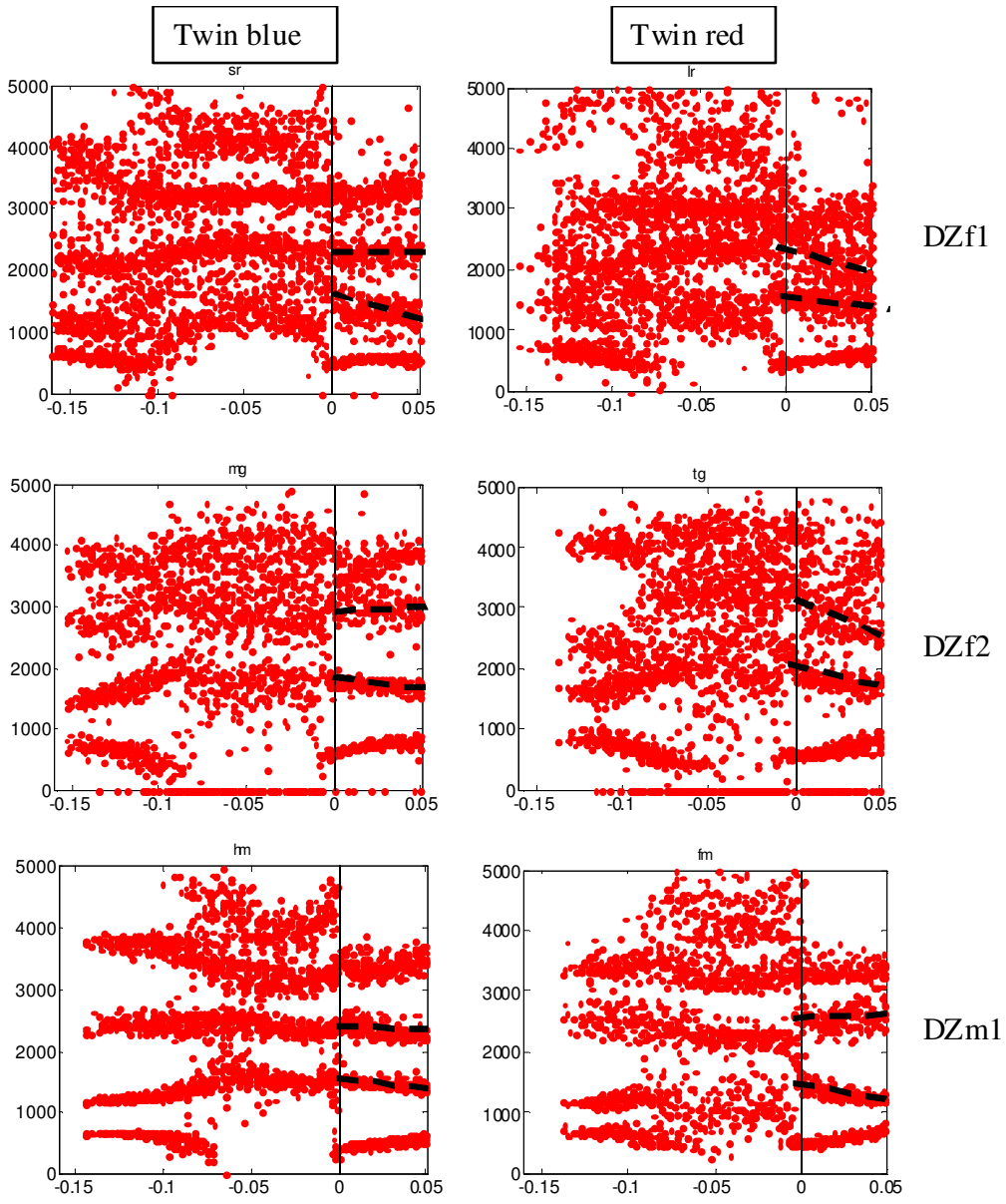


Figure 46: Formant transition plots of the sequence /aʃə/ for all speakers of the DZ pairs; data are aligned at the end of the fricative marked by the black vertical line.

5.2.6.3 Quantitative analysis of sibilant-vowel TRANSITIONS

To quantify the inter-speaker differences in formant transitions, in the next step statistical differences between F2 and F3 vowel targets and sibilant-vowel transitions for each pair were analyzed. To do this, the F2 and F3 targets of the vowel [ə] (Ftarget, 40 ms after the sibilant) and the formant value at the end of the sibilant (Ftrans) were measured. Then, the difference

between Ftrans and Ftarget was calculated to obtain information about the slope of the transition (falling, rising or stable transition). An ANOVA and a post hoc Tukey test were carried out to look for significant differences within the pairs for F2 and F3 targets (Ftarget) and for F2 and F3 transitions (Ftrans-Ftarget). No difference between MZ and DZ twins in inter-speaker variability in formant TARGETS could be found. Both twin types reveal differences. However, TRANSITIONS between the sibilant and the following schwa were more similar for the MZ than for the DZ twins. Like the formant plots suggested, no MZ pair showed statistically significant differences in their F2 or F3 transitions, neither for /s/ nor for /ʃ/, but significant differences were found for the DZ twins in 4 cases (cf. Tables C.8 and C.9 in the appendix for the statistics).

Table 26: Significant differences in vowel targets and transitions for all twin pairs.

		/s/		
Twin	Vowel target F2	Vowel target F3	Transition F2	Transition F3
MZf1				
MZf2				
MZm1		*		
MZm2				
DZf1			*	
DZf2		*		
DZm1	*			*
		/ʃ/		
Twin	Vowel target F2	Vowel target F3	Transition F2	Transition F3
MZf1	*	*		
MZf2		*		
MZm1				
MZm2		*		
DZf1		*		*
DZf2		*		*
DZm1	*	*		

5.2.7 Conclusion

To sum up, zygosity (and hence NATURE) seems to have no effect on acoustic TARGETS of /s/. However, a tendency towards more similar acoustic outputs in MZ twins than in DZ twins was found for /ʃ/, and this parallels the results of the articulatory analysis. In addition, the acoustic realization of the phoneme contrast shows no clear effect of zygosity, but here, too, a tendency was found towards MZ twins being more similar in their acoustic difference between the sibilants than DZ twins. Thus, concerning articulatory and acoustic TARGETS, results point to the fact that shared physiology (thus NATURE) is more important in /ʃ/ than in /s/, but overall it shows less impact on TARGETS than on TRANSITIONS, namely on sibilant-vowel transitions. These were found to be more similar for MZ twin pairs than for DZ twin pairs. This corroborates hypothesis H2 and the influence of individual biomechanical properties and restrictions on the acoustic output of trajectories between two targets. Learned auditory goals, the influence of shared social environment and hence NURTURE seem to play a minor (perhaps negligible) role regarding sibilant-vowel transitions.

5.3 Limitations and further research

One assumption that could be made from the articulatory analysis of the sibilants was that while no clear separation between MZ and DZ twins in their articulatory target positions was found, the shape and form of the tongue might be influenced by zygosity (NATURE): in general, the MZ twins revealed a more similar tongue configuration than the DZ twins. However, as noticed before, conclusions must be drawn very carefully, since EMA recordings do not provide the shape of the tongue during articulation but only give information on positional data of different *points* on the tongue; the tongue contour is only interpolated. An articulatory analysis using magnetic resonance imaging (MRI) or ultrasound could visualize the whole tongue contour during articulation and would therefore be much better for investigating the shape of the tongue. Although the present investigation can give some valuable impulse, further research with MRI or ultrasound could give deeper insights into the influence of physiology on speaker-specific tongue doming or stretching in different articulation strategies, especially in sibilants, where the shape of the edges of the tongue and the forming of a groove at the midsagittal line along the tongue blade are most interesting.

Another issue that deserves further investigation is the speaker-specific degree of lip protrusion in realizing the /s/-/ʃ/ contrast. Due to the recorded speech material, the phoneme /s/ only appeared in the context of a rounded vowel (the target word was /kʏsə/). Therefore, lip protrusion is still apparent in the production of /s/, independent of the speaker. The difference in the degree of lip protrusion in realizing the phoneme contrast could surely give some interesting information about speaker-specific articulation and the different motor control strategies, including the use of the lip cavity. Further research should involve /s/-/ʃ/ minimal pairs with comparable phonemic environments.

Since a link between speech production and speech perception is assumed, the importance of the ability to perceive different acoustic categories should not be neglected (cf. Newman et al. 2001, Perkell, Guenther et al. 2004, Perkell, Matthies et al. 2004, Perkell et al. 2008). The process of learning and the existence of auditory targets are considered to be crucial and similar acoustic outputs are explained by similar social environments. However, nothing can be produced on purpose if it is not perceived first. Therefore, further investigation should

take the individual's auditory acuity into account to learn more about the possible reasons for differences in acoustic outputs. This is especially relevant when it comes to the acoustic realization of the phoneme contrast. Even though a tendency was observed for MZ twins to be more similar in their acoustic contrast, no clear effect could be found. This might be due to the small size of the subject group, and therefore more research should be done on this topic. Moreover, the auditory and somatosensory acuity of the speakers should be investigated as well, to better understand the interplay and different impacts of shared physiology, and auditory and somatosensory acuity.

An interesting finding of this analysis was the differences in sibilant-vowel transitions between the speakers of DZ twin pairs in comparison to the very similar transitions within MZ twin pairs. The results of the present analysis support the findings of earlier studies that found coarticulatory effects to be more similar in MZ twins than in normal siblings or unrelated speakers (Nolan & Oh 1996, Whiteside & Rixon 2003). Following these findings, coarticulatory parameters seem to be more influenced by physiological characteristics than targets. However, studies from auditory perturbation experiments have shown that people adapt to changes in the auditory feedback of transitions (Cai et al. 2010). This in contrast points to the relevance of transitions in terms of auditory goals. Thus, further research is needed to examine more closely the role of physiology/NATURE on the one hand and the role of auditory goals/NURTURE on the other hand in transitions. More insight on this issue will be given in Chapter 7, where the auditory similarity of the stimulus /vaʃə/ in MZ and DZ twins is investigated. In addition, the production strategies corresponding to the different transition patterns in the sibilant-vowel sequences are a topic for future research.

In the following chapter, the articulation of /aka/-sequences and, in particular, the articulatory strategies used during the production of these sequences are at the center of attenation. Hence, the focus of the analysis changes from articulatory targets to articulatory GESTURES.

6 INTER-SPEAKER VARIABILITY IN ARTICULATORY GESTURES: /aka/

A great deal of articulatory variability can be found in the production of velar stops. Velar stops vary in their horizontal position depending on the neighboring vowels: they tend to be produced at a more anterior position if they are surrounded by front vowels and at a more posterior one if they are surrounded by back vowels (cf. Alfonso & Baer 1982, Parush et al. 1983, Geng et al. 2003). Variability is also found in the vertical position due to manner of articulation: fully voiced stops in intervocalic position often show no burst due to incomplete closure. An even more complex and also variable articulatory production pattern related to velar stops is the looping movement of the tongue back that is found in VCV sequences where C is a velar consonant: these sequences are not simply produced on a straight path from the vowel to the consonant target and back along another straight line to the next vowel target, but an elliptical movement of the tongue back – a loop – is found (Kent & Moll 1972, Mooshammer et al. 1995, Hoole et al. 1998, Löfqvist & Gracco 2002, Perrier et al. 2003, Brunner et al. 2011).

While studies agree on the fact that the loops exist, the explanation of this phenomenon is less clear. The shape of the elliptical movement seems to be dependent on the neighboring vowels, the direction of the loop might be affected by the preceding vowel (normally a forward loop is observed, although a preceding /i/ seems to trigger a backward loop), and the size of the loops might depend on the voicing status of the consonant and the degree of air pressure used. Furthermore, speaker-specific looping patterns in terms of size and shape have been found but have not been discussed much yet. In addition, the reasons for the observed movements still remain unclear. Several explanations for the looping pattern are offered in the literature, and there has been a great deal of discussion on a possible influence of aerodynamic forces (among others Houde 1967, Hoole et al. 1998). Moreover, different degrees of planning related to the trajectories have been suggested. Some consider the whole gesture to be planned within the concept of motor control principles (Löfqvist & Gracco 2002), while

others assume biomechanical properties and physical characteristics of the speech production system (i.e. NATURE) to be crucial (Perrier et al. 2003).

The first study reporting on articulatory looping patterns was Houde (1967). He found the loops in VCV sequences where V1 = V2 and showed that an explanation of these patterns based only on vowel-to-vowel coarticulation is not sufficient. Houde (1967) and Kent & Moll (1972) suggest that the air pressure behind the constriction may play a role; Coker (1976) and Houde (1967) assume that it might be used to sustain voicing through active cavity enlargement.[14] Ohala (1983) also suggests that the looping pattern could serve as an active cavity enlargement and a compensation for other factors that disfavor voicing in velars. Mooshammer et al. (1995) analyzed lingual movement during VCV sequences with varying vowel contexts and manners of articulation of two speakers of German by means of electromagnetic articulography. They, in contrast, found that the articulatory loops were larger for the sequence with the unvoiced stop consonant (/aka/) than for the sequence with the voiced consonant (/aga/). Thus, cavity enlargement cannot be the main reason for the emergence of loops and other reasons must exist. Hoole et al. (1998) conducted a study to analyze the possible influence of aerodynamic pressure forces by comparing the production of velar consonants during normal and ingressive speech. In both cases forward articulatory loops could be found, but of a smaller size in ingressive speech. Thus, aerodynamics seems to influence tongue movements, but other factors such as biomechanical properties of the tongue also tend to have a significant effect on the loops.

To investigate the influence of biomechanics (i.e. NATURE) Perrier et al. (2003) conducted a modeling study and hypothesized that biomechanics plays an important role in the forward looping movements of the tongue. The authors suggest that complex articulatory patterns do not necessarily need the existence of complex internal models since these patterns can at least partly be explained by the anatomical and physiological arrangements of the tongue muscles. To prove this, they simulated V1CV2 movements, where C is a velar consonant and V is [a], [i] or [u], by using a tongue model based on Payan & Perrier's (1997) 2D tongue model.

[14] Houde (1967) claims that "[t]he direction of the movement during closure is consistent with an increase in oral pressure, and as in the case of labial closures, a compliant element is required in the oral cavity during the voiced palatal stop in order to sustain voicing. The passive reaction of the tongue may provide that required compliance" (p. 133).

Basically, the model used consists of hard structures (like palate, teeth and bones) and a soft body that represents the tongue with seven muscles that can be modeled. Vowels and consonants are specified in terms of targets (in the case of the consonants the target is virtual and lies beyond the palate[15]). The resulting trajectories of the different VCV simulations were inspected. Results show that if V1 is [a] or [u], the trajectory forms a forward loop that starts before the consonant closure and continues to slide along the palate during the closure, while the direction of the path is backward for sequences where V1 is [i].[16] The amount of vertical movement during the consonantal closure turned out to be dependent on V1 and ranges from 5 mm for [a] over 3 mm for [u] to 2 mm for [i]. The authors assume that the curved trajectories consist of independent horizontal and vertical components, and biomechanical factors, such as the way the tongue muscles produce the velar closure and the interaction with the palate during the consonantal closure, are seen as being crucial in explaining the trajectories. A relation of the length of the sliding contact section (i.e. the horizontal movement along the palate) to the distance between the position of the tongue when it first touches the palate and the position of the consonant's virtual target is suggested. This is especially relevant for the current study as it may also be seen in the light of speaker-specific characteristics, since the amount of horizontal movement might be constrained by the (individual) palatal contour. In addition, the authors also point to the impact of speaker-specific properties, which might explain differences between their simulations and the findings of Mooshammer et al. (1995) regarding the size of the loops in V1kV2 sequences: "while the general orientation of the loop is the same for each speaker, the amplitude of the sliding movement during the closure depends on speaker-specific properties, at a control and at a physical level" (Perrier et al. 2003, p. 1594).

---

[15] A *virtual* target can never be reached. In the case of velar stop consonants this is based on the assumption that the consonant target is located beyond the palate. The virtual target hypothesis has been suggested by Löfqvist & Gracco for labial and lingual stops (1997, 2002) and has been supported by a comparative study of articulatiory data from German and simulation by Fuchs et al. (2001).

[16] Note that this finding is contrary to the results of Houde (1967), who found superimposed forward loops for V1 = [i]. Thus, a more variable pattern for V1 = [i] in the orientation and shape of the loops depending on speakers or languages might be assumed.

To summarize, from their findings Perrier et al. assume that the curvature of the trajectory is due to biomechanical properties of the tongue model (the passive tongue elasticity, the muscle arrangements in the tongue, the force generation mechanism), and no general optimization principle that plans the entire trajectory (a complex internal model) is necessary to explain the shape of the trajectory as proposed by Löfqvist & Gracco (2002).[17] Thus, the presence and general shape patterns of the loops are considered to result from tongue biomechanics and muscular anatomy, while the upper portion of the loop and the amount of horizontal sliding during the palatal contact might be influenced by speaker-specific palatal shapes (NATURE).

With these studies in mind, the central question that motivates this analysis on elliptical trajectories in twins' speech is concerned with speaker-specific variability in the production of the loops. Much variability is found in the production of loops depending on several factors as discussed above (surrounding vowels, air pressure, voicing status, virtual target specification), but the crucial factor that is analyzed in the following chapter is the *speaker* and, in particular, the speaker's individual physiological restrictions. Are the loops that are produced in the sequence /aka/ a natural consequence of the biomechanical properties of the speech articulators as proposed by Perrier et al. (2003), and thus do biology and individual differences in physiology (NATURE) influence the shape of the elliptical trajectories in VCV sequences? In other words and going a step further, the aim of the following analysis is to find out whether speaker-specific biomechanical and physiological characteristics influence the articulatory looping movements and are therefore more similar in physiologically identical MZ twins than in DZ twins or unrelated speakers.

[17] Löfqvist & Gracco (2002) suggest that the looping patterns arise from general motor control principles based on a cost minimization, where the whole trajectory of the tongue is planned. Physical factors (aerodynamics and biomechanics) are considered to play a minimal role.

6.1 Hypotheses

Based on the abovementioned studies and previous findings of the current study two assumptions are made: 1) articulatory gestures (and particularly looping patterns in VCV sequences) are influenced by physiological and biomechanical speaker-specific characteristics, and 2) MZ twins are more similar in their physiology than DZ twins and unrelated speakers. Therefore, it is hypothesized that:

H1: MZ twin pairs are more similar than DZ twin pairs and unrelated speakers in the articulatory movements of the tongue back during the sequence /aka/ (i.e. there is an influence of NATURE).

The alternative hypothesis is:

H2: MZ twin pairs, DZ twin pairs and unrelated speakers do not differ in the degree of inter-speaker variability in the articulatory movement of the tongue back during /aka/ (i.e. there is no influence of NATURE).

6.2 Method

6.2.1 Subjects

Since the twin pair MZm2 had to be excluded from the present analysis due to low correctness factors and high variability values for the tongue back coil (cf. Section 3.5.1) 10 speakers (5 twin pairs) took part in the further investigation: 3 MZ pairs (2 female, 1 male) and 2 DZ pairs (both female) (remember that no articulatory data could be recorded from DZm1). Again, some renditions of the speakers could not be used due to articulatory measurement errors (i.e. 4%), and therefore a different number of tokens was analyzed for each speaker. An overview of the investigated speakers and the number of analyzed items is given in Table 27.

Table 27: Number of analyzed items for all speakers.

Twin pair	Speaker	No. of analyzed items
MZf1	AF	10
	HF	10
MZf2	GS	10
	RS	10
MZm1	CL	10
	SL	9
DZf1	LR	11
	SR	9
DZf2	TG	7
	MG	10

6.2.2 Speech material

Looping patterns of the tongue were investigated in the sequence /aka/ in the word /k**aka**du:s/ (cf. Section 3.3). The movement of the tongue back was analyzed by inspecting the positional data of the tongue back coil over the duration of the sequence /aka/. Start and end of the gesture were determined with the help of the tangential velocity of the tongue back coil. The minimal tangential velocity and thus the lowest position of the tongue back determine the reaching of the first and second /a/-target (cf. Section 3.5.3).

To get a better impression of the analyzed data, Figure 47 shows the articulatory movement of the tongue back during the sequence /aka/ for three renditions of one speaker (SL, MZm1). The three different renditions of the speaker are visualized by different colors and the start of each gesture is marked by an asterisk. It can be seen that the tongue does not simply lift for the /k/ and go down again for the /a/, but performs a forward loop. The

tongue first lifts while moving backwards and then slides forward along the palate during the occlusion of the /k/ before it moves down and back again to the second /a/.

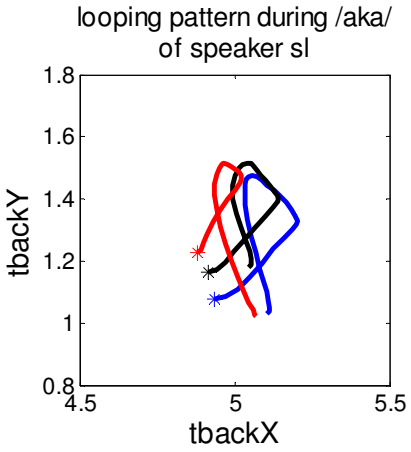


Figure 47: Articulatory movements of the tongue back in three /aka/ sequences for speaker SL of MZm1; different colors indicate three different renditions and the asterisks mark the beginning of the trajectory.

6.2.3 Data processing

For the present investigation the whole articulatory interval over the sequence /aka/ was used, since the focus lies on the realization of an articulatory gesture rather than on a specific target. In order to compare inter-speaker variability in an articulatory gesture the data had to be processed in several ways. First, the articulatory signals had to be aligned in time. This was done with the help of Functional Data Analysis (FDA, Ramsay & Silverman 1997, see below). Second, the aligned tokens had to be paired between speakers, resulting in multiple pairwise comparisons that consist of 1) tokens from two speakers of the same twin pair (MZ and DZ pairs), 2) tokens from two unrelated (sex matched) speakers and 3) different tokens from the same speaker. And third, a measurement of similarity in articulation that could be used for the statistical analysis (Euclidean distance between the aligned repetitions) was calculated for each pairwise comparison.

6.2.3.1 Functional Data Analysis

A problem that arises when comparing articulation and especially articulatory gestures between different speakers is that the articulatory data of each speaker and even of each repetition have different durations due to differences in speech rate. To compare time series (in this case, articulatory data points) that have different durations, the investigated tokens have to be aligned in time. The alignment of the bidimensional data consisting of the horizontal and the vertical movement of the tongue back during the looping gesture of /aka/ is done with the help of Functional Data Analysis (FDA). By this means time-aligned data can be obtained that can be compared more easily.

The technique of FDA to analyze time-varying signals was introduced by Ramsay & Silverman (1997), who use nonlinear time-warping to bring multiple signals into closer alignment with an average signal. Several studies have used FDA to study variability in speech: Lucero et al. (1997) used nonlinear temporal normalization to analyze the common pattern of lip movements and the variability in shape and timing. Lucero & Koenig (2000) investigated irregularities of voice signals, and Koenig et al. (2008) examined the intra-speaker variability in the fricative production of children vs. adults by this means. Lucero & Löfqvist (2005) studied articulatory variability in VCV sequences and found that it varies depending on the phonetic requirements of the consonant and the biomechanical characteristics of the articulatory structures involved. Variability in speech timing near a phrasal boundary was studied by Lee & Krivokapic (2006). All of these studies illustrate that FDA is a powerful analytic tool for studying aspects of variability in the speech production process including articulation and acoustics.

FDA has several advantages over non-normalized averaging, where the average signal is computed by taking the average point by point. As a result, the averaged signal no longer resembles the individual waveforms due to cumulative distortions with increasing time. In linearly normalized averaging the average is closer in shape to the original records but still reveals significant differences in the amplitude. Here, the major cause of distortion is phase variability: the interpolated signals can be slightly out of phase because of non-uniform timing changes (variation in the timing of landmarks) in the different trials. With the help of FDA, warping functions (transformations of time) for each repetition are created in such a way that

each aligned token (as a function of the transformed time) is close to the average. After a set of warping functions, a new average of the registered tokens is used to compute a new set of warping functions. This process is iterated until there is no significant difference between two consecutively calculated sets of warping functions or averages anymore (see Lucero et al. 1997 for a more detailed overview). Lucero et al. (1997) discuss the advantages of FDA and state that "(…) if sequential movements reflect a patterning process the only way to understand the process and identify the pattern is to faithfully reconstruct the pattern from the observations. Nonlinearly normalized averaging preserves the shape of the movement patterns while retaining the option to investigate variability in their production" (p. 1117).

Two kinds of variability can be investigated by means of FDA: variability in time and in amplitude. The warping functions indicate the differences between the timings of the unaligned waveforms and the computed average. Therefore, standard deviations across the warping functions are an index of phasing variability (how much the tokens differ in time). The standard deviation across the aligned tokens is an index of magnitude variability (how much the tokens differ in amplitude). In this study FDA was used following the method proposed by Lucero.[18] By using this method, plots of the raw articulatory tokens and of the aligned articulatory tokens were obtained and can be inspected. In addition, the mean amplitude variability (the standard deviation across the aligned tokens) for each pair in horizontal and vertical movement is discussed in the results. As an example, Figure 48 shows the original data and the time-aligned data of all repetitions of the two speakers of MZm1 for the horizontal and vertical movement of the tongue back.

The plots on the left side of Figure 48 show the original data for both speakers of MZm1 of the horizontal (upper part) and vertical (lower part) movement of the tongue back during the sequence /aka/. Differences in the lengths of the articulatory tokens can be seen. On the right side the time-aligned data (normalized time from 0 to 1) of both speakers are plotted and similar movement patterns can be observed.

[18] A MATLAB script by Jorge C. Lucero, last modified February 2010, was used for the alignment of the bidimensional records; the required Functional Data Analysis Toolbox by James Ramsay is available at http://www.psych.mcgill.ca/misc/fda/.

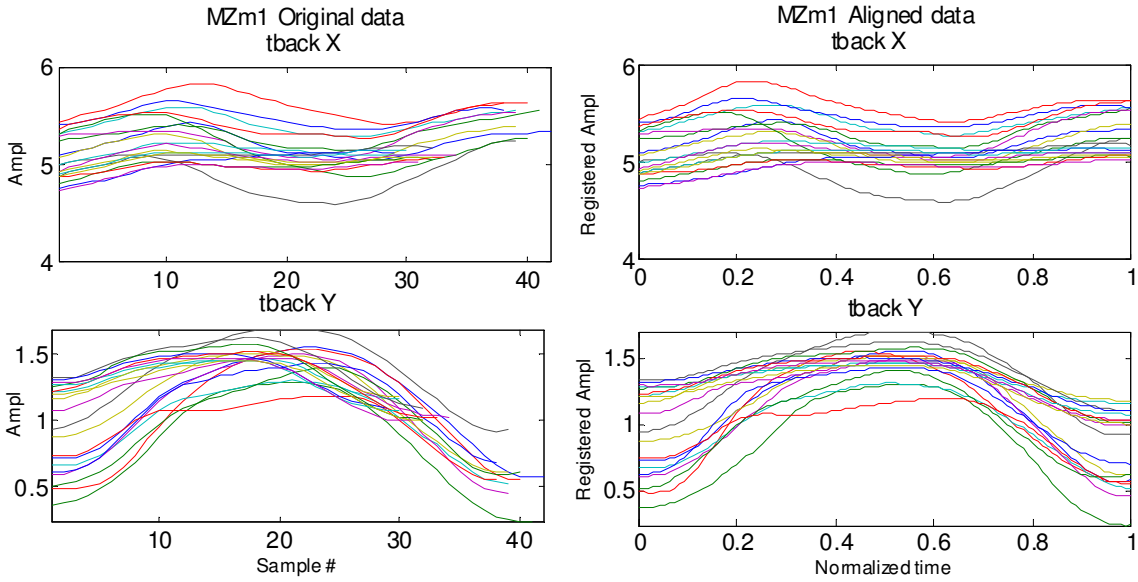


Figure 48: Original data (left) and aligned data (right) for the horizontal movement (tbackX, upper part) and vertical movement (tbackY, lower part) of the tongue back for all articulatory tokens of the two speakers of MZm1.

6.2.3.2 Multiple pairwise comparisons and Euclidean distances

The focus in this investigation lies on the comparison of articulatory gestures between related speakers (twin pairs) and unrelated speakers. Therefore, all possible pairwise combinations of the time-aligned tokens were constructed for each speaker pair separately. This was done a) for the 5 twin pairs, b) for all other possible (gender matched) pairwise comparisons, resulting in 24 different unrelated speaker pairs, and c) for the different renditions of each speaker. After that, the mean Euclidean distance (ED) between the aligned bivariate articulatory trajectories of the relevant region (i.e. vertical and horizontal movement of tongue back for the sequence /aka/) for each pairwise comparison was calculated: the ED was measured for each data point between the aligned tokens and summed up for every possible comparison. Figure 49 shows two renditions of speaker SL; the red lines visualize the measured EDs.[19] In this way a mean ED for each comparison could be obtained.

[19] Note that the figure only serves to visualize the procedure. Here, the two renditions have not yet been normalized; in reality the time normalized data (with the same number of data points) were used to calculate the ED.

The measured EDs between the articulatory tokens of the twins and the non-related pairs served as an index of the amount of inter-speaker variability as a function of the different pairs. In addition, intra-speaker variability could be obtained by looking at the EDs between the different repetitions of each speaker. The resulting EDs between the articulatory tokens for the four groups *DZ twins, MZ twins, unrelated speaker pairs* and *same speaker* were used for the statistical analysis. The measured mean distances and standard deviations for each possible speaker pair can be inspected in appendix D (Table D.1).

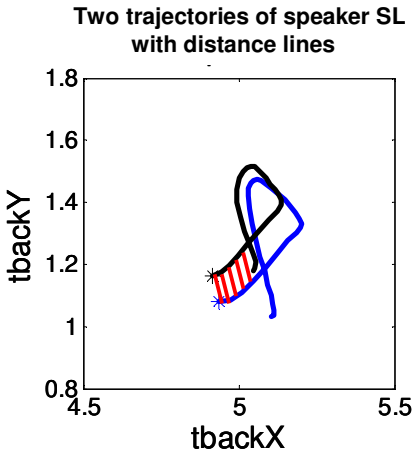


Figure 49: Two articulatory trajectories of the tongue back coil during the sequence /aka/ with visualized EDs between early datapoints; the start of the sequence is marked by an asterisk.

Note that, even though we are looking at positional data and the ED between positional data points (of bidimensional data), the shape of the trajectory is also indirectly considered and compared since the trajectory consists of various data points and each of these points is analyzed and compared between the speakers (see Figure 49). Differences in the shape of the trajectories should therefore result in a higher ED.

6.3 Results

A qualitative and quantitative analysis was conducted. First, the raw data of the looping movements of the tongue back during /aka/ is shown for each speaker and compared within the twin pairs qualitatively. After that, the results of the FDA are shown and figures of original and aligned data are presented, since the alignment is necessary for the following quantitative analysis. Here, Euclidean distances (ED) are calculated between the normalized articulatory tokens for each speaker pair. A statistical model is fitted to find significant differences in ED between same speaker pairs, MZ speaker pairs, DZ speaker pairs und unrelated speaker pairs.

6.3.1 Qualitative comparison of the looping patterns

As a first step, the shape of the loops is more closely inspected. The extent and direction of the upward movement, the length of the horizontal sliding movement at the palate, and the amplitude and shape of the downward movement can be investigated. Figure 50 gives a closer look at the looping trajectories of the tongue back during the sequence /aka/ for the MZ twins. Each twin pair is shown in a subplot and different speakers are indicated by different colors. For a better visualization of the general pattern of the loops only two (out of 10) renditions are shown for each speaker. At first glance, a great deal of variability in the size and shape of the loops can be seen *between* the pairs. Each pair reveals a different pattern, but *within* the pairs similarities are obvious. The speakers of MZm1 reveal different sizes of the loops, but the curvature and shape of the trajectories are very similar. The different sizes of the loops can be explained by differences in speech rate and the degree of clear pronunciation. Although subjects were told to speak normally and without paying special attention to their speech during the recordings, speaker CL (blue) sometimes used very precise speech. He speaks more slowly and articulates more clearly than his brother, thus this may explain the different looping sizes. The shape of the loops, on the other hand, is very similar between the brothers: the tongue goes up and backwards until it reaches a certain position (palatal contact), then it slides forward and up along the palate until it moves down and a little back again, forming a triangle. Hence, even though different speech rates and

degrees of casual articulation were used, the overall patterns of the loops are nearly identical between the brothers and point to an influence of shared physiology and biomechanics. Both speakers of MZf2 reveal a rounder and smoother trajectory than the triangle shaped loops of MZm1. They also start with an upward and backward movement, but no clear start and end point of the palatal contact can be seen. Differences between the speakers can be found regarding the downward movement to the second vowel: while RS (red) just lowers the tongue, GS (blue) also retracts it.

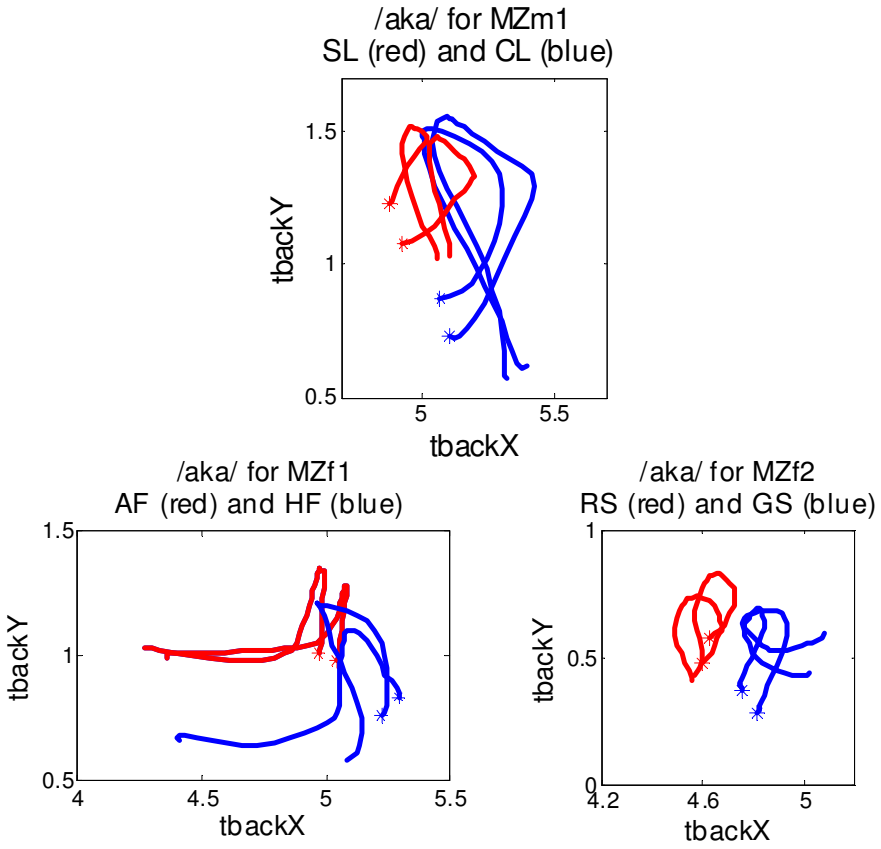


Figure 50: Looping trajectories of the tongue back during /aka/ for the monozygotic twin pairs (MZm1, MZf1, MZf2); each twin pair is shown in a separate subplot with two renditions of each speaker; different speakers are indicated by different colors; starting points of the trajectories are marked with an asterisk.

For the pair MZf1, it turned out to be difficult to find an articulatory landmark that determined the end point of the loop. The trajectories of both speakers of this pair look

different due to coarticulatory processes. As described in the method section, the minimum of the tangential velocity of the tongue back sensor was used to mark the beginning and end of the looping movement describing the /aka/ sequence. For this pair, the articulatory movement for the following /d/[20] (i.e. moving the tongue forward) could in some cases not be separated from the looping movement by using the minimum of the tangential velocity. This can be seen in the plots and explains the horizontal movement of the tongue after it reaches a vertical minimum for the /a/. In addition, MZf1 differs from the other pairs with respect to the upward movement from the first vowel to the stop: both speakers of MZf1 move upwards in a straight line or even with a slightly forward directed movement, while the other speakers started with an upward and backward movement. The length of the palatal contact seems to differ within the pair, in any case: speaker AF (red) lowers the tongue directly, while the angle of the loop of her sister HF (blue) is less steep. Despite this, the downward movement is straight for both speakers and less rounded than that of the speakers of MZf2.

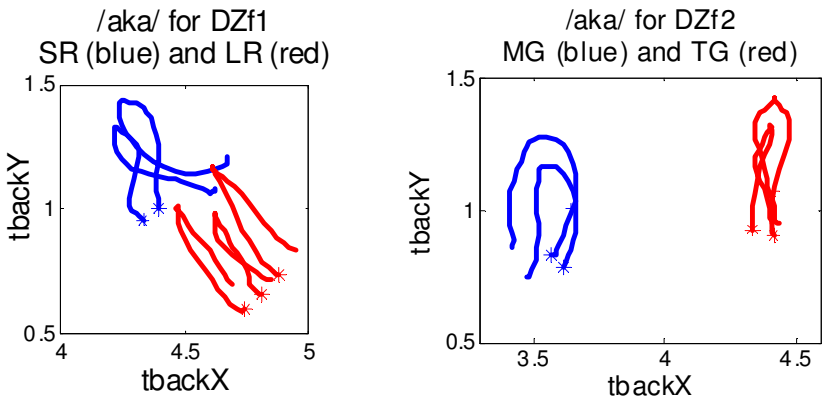


Figure 51: Looping trajectories of the tongue back during /aka/ for the dizygotic twin pairs (DZf1, DZf2); each twin pair is shown in a separate subplot with two (or three) renditions for each speaker; different speakers are indicated by different colors; starting points of trajectories are marked by an asterisk.

[20] Remember that the /aka/ sequences were part of the carrier word /kakadu:s/.

Figure 51 shows the looping trajectories of the dizygotic twin pairs. The speakers of the dizygotic twin pair DZf1 differ in their horizontal and vertical positions of the trajectory. The loop of speaker SR (blue) is situated more anterior and higher than that of her sister, which might point to differences in vocal tract or tongue size. In addition, the form of the trajectory differs: SR (blue) reveals a more rounded trajectory with a straight upwards tongue elevation, a slightly horizontal movement at the palate and a down and backwards movement of the tongue to the second /a/. LR (red), in contrast, shows a forward directed tongue elevation with a very steep angle at the turning point at the palate and a downward and backwards movement to the second vowel. No horizontal sliding at the palate can be seen. Moreover, LR (red) reveals a strikingly different looping pattern as compared to all other speakers and the loops discussed in the literature: in the sequence /aka/ a *forward* loop is expected, but LR shows at least for some trajectories a *backwards* directed movement (which is normally only expected when the first vowel is /i/). The second dizygotic twin pair DZf2 differs especially in the horizontal position of the trajectory. The loop of speaker MG (blue) is situated more anterior (and slightly lower) than that of her sister, which again points to differences in vocal tract or tongue size. This can also be seen in the next figure (lower part of Figure 52, right-hand): the aligned tokens of this twin pair differ strongly in tbackX, the horizontal position of the tongue back. Here, two different groups of articulatory tokens are apparent. But in addition to these positional differences, MG (blue) and TG (red) also vary in the amount of horizontal movement at the palate. Both speakers use straight upward and downward tongue movements, but MG reveals much more sliding movement at the palate than her sister. Additionally, the vowel targets before and after the velar seem to be closer in their horizontal position for TG than for her sister.

6.3.2 Alignment of data

In Figure 48 the original articulatory data and the resulting time-aligned tokens for MZm1 were shown. As indicated above, amplitude variability indexes for the time-aligned tokens were calculated and inspected more closely. The mean standard deviations of the amplitude variability functions over normalized time were measured for all twin pairs. MZm1 reveals a mean amplitude variability of 0.2 cm for the horizontal movement and 0.18 cm for the

vertical movement. In the following Figure 52 the time-aligned articulatory data for the female MZ pairs and DZ pairs are given.

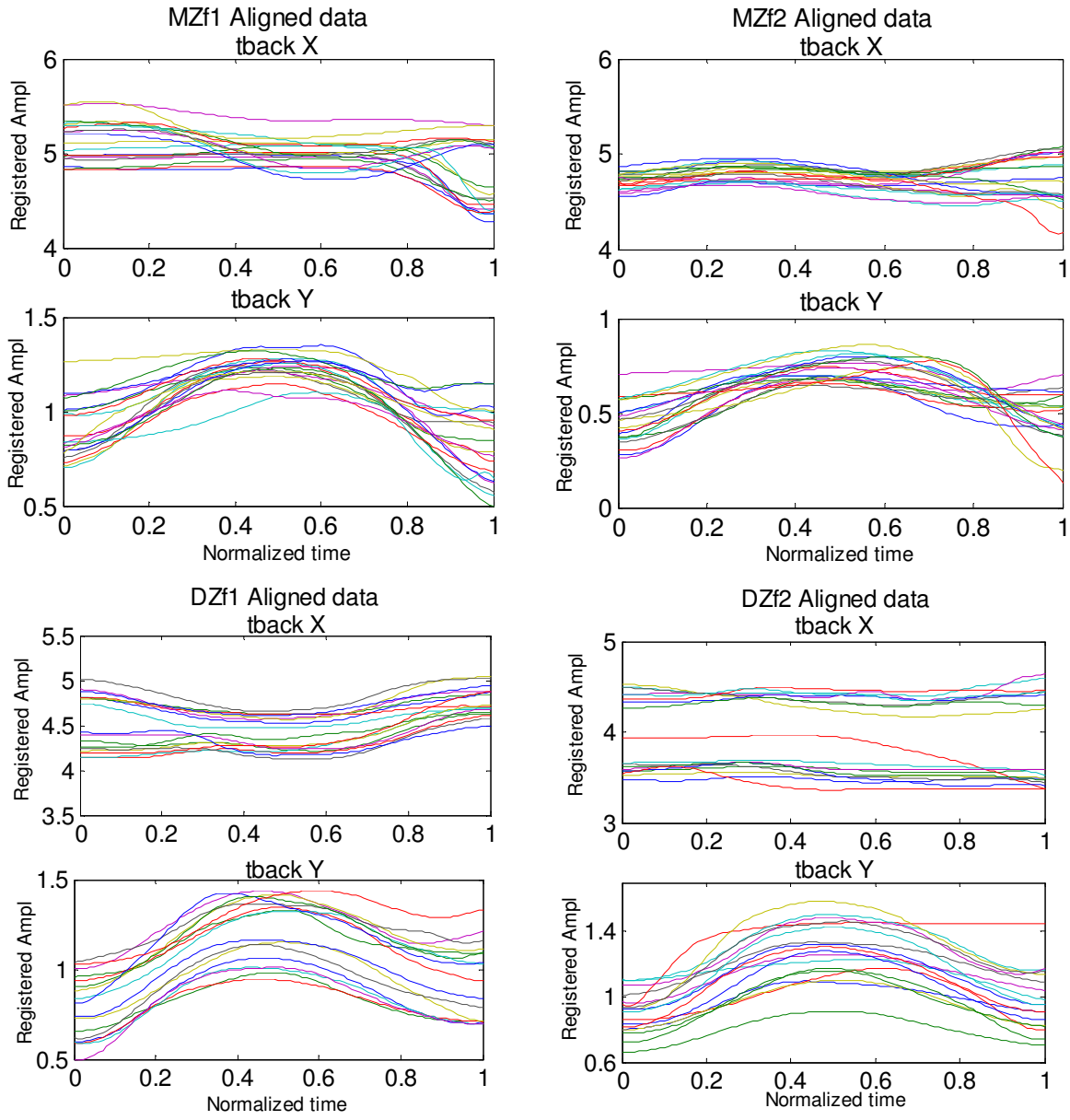


Figure 52: Time-aligned data for the MZ twin pairs (upper part) and the DZ twin pairs (lower part) for the horizontal movement (tbackX) and vertical movement (tbackY) of the tongue back during /aka/; y-axis: amplitude in cm, x-axis: normalized time from 0-1.

In general, it is apparent in the graphs that larger differences between the tokens of the DZ pairs exist, since the different renditions are distributed over a greater range. In numbers, this means that the time-aligned tokens of MZf1 reveal a mean amplitude variability of 0.18 cm

for the horizontal movement (tbackX) and only 0.09 cm for the vertical movement (tbackY). MZf2 shows even less amplitude variability: 0.11 cm for the horizontal and 0.07 cm for the vertical movement. In contrast, the DZ pairs reveal higher variability indexes: the horizontal movements have a standard deviation of 0.2 cm for DZf1 and even of 0.42 cm for DZf2. The vertical movement varies on average by 0.17 cm for DZf1 and 0.16 cm for DZf2. Thus, the amount of amplitude variability expressed in the mean standard deviation of the aligned tokens points to greater differences between DZ pairs. Even more noticeable, a visual inspection of the graphs reveals a subdivision into two speakers, since the tokens of both DZ pairs can clearly be separated into 2 groups (=speakers), especially for the horizontal movement of DZf2.

Furthermore, for the MZ twins, less variability is found for the vertical movement at the middle of the sequence, where the consonantal closure is expected, but not for the DZ twins. MZf1 and MZf2 reveal more amplitude variability at the beginning and end of the sequence than in the middle (Figure 52, upper part); DZf1 and DZf2 reveal a more stable amplitude variability for the vertical movement over the whole part of the sequence (Figure 52, lower part). This might be due to the similarity of the palates of the MZ twins; pointing to an influence of the factor palatal shape on the upper part of the looping trajectory (i.e. the time of the velar closure), while the start and end of the sequence (i.e. the vowels) are more variable and less influenced by physiological constraints.

6.3.3 Quantitative comparison of the looping patterns

To look for a significant effect of speaker pair (and in particular the pairs' physiological similarity) on the similarity of the looping pattern, statistical tests were conducted using R (version 2.8.1, R Development Core Team 2008) and included linear mixed models (Pinheiro & Bates 2000). A generalized linear mixed model was calculated with the measured articulatory *Euclidean distances (ED)* as dependent variable. Since the EDs were very small and the differences in the values of the EDs between the speaker pairs were not expected to be linear, the logarithmic values (and not the absolute values) of the calculated EDs were used. Mixed models incorporate fixed and random factors as independent variables. For the model the factor (speaker) GROUP was taken as a fixed factor, and four different levels from this

factor mirroring the different physiological similarities were built. Since 10 speakers took part in the analysis, 10 pairs form the level *same speaker* (SSp), 3 pairs represent the group of the monozygotic twins (MZ), 2 pairs the dizygotic twins (DZ), and 24 different sex matched pairs could be formed out of the 8 female speakers (see also Table 28).

Table 28: Overview of the fixed factor (speaker) GROUP with different levels and numbers of pairs.

Factor	Levels	Number of pairs
(speaker) GROUP	Same Speaker (SSp)	10
	MZ	3
	DZ	2
	Unrelated Speaker (UN)	24

Since it has been hypothesized that biology has an impact on the shape of the looping patterns, the levels of the factor GROUP were ordered according to the speakers' genetic and thus physiological similarity. It is hypothesized that EDs grow in the following way: (a) same speaker < (b) MZ pairs < (c) DZ pairs < (d) unrelated speakers. Thus, the factor GROUP was considered as an ordered factor (SSp < MZ < DZ < unrelated) and expressed through a successive difference contrast (Venables & Ripley 2003).[21] In addition to the fixed factor GROUP, the different *speakers* that form the speaker pairs are included in the model as random factors (named SPEAK1 and SPEAK2). Thus, the resulting model is:

MODEL: lmer(log(distance) ~ group + (1|speak1) + (1|speak2))

As a first step and for a visual inspection of the data, boxplots were made mirroring the ED (on a logarithmic scale: log(distance) separated by the four different groups (see Figure 53). The median of the distribution in each group is visualized by vertical lines in the boxes, the boxes comprise 50% of the data, the whiskers extend to the most extreme data point which is

[21] The relevant command in R that was used to test for significant differences between all levels of the factor group is: contrasts(data$group)<-contr.sdif(levels(data$group)).

no more than 1.5 times the interquartile range from the box, and outliers are marked with open dots. The figure clearly shows the expected differences between the groups in the amount of ED, with same speaker pairs showing the lowest values. Interestingly, no difference in mean values is obvious between DZ twins and unrelated speakers.

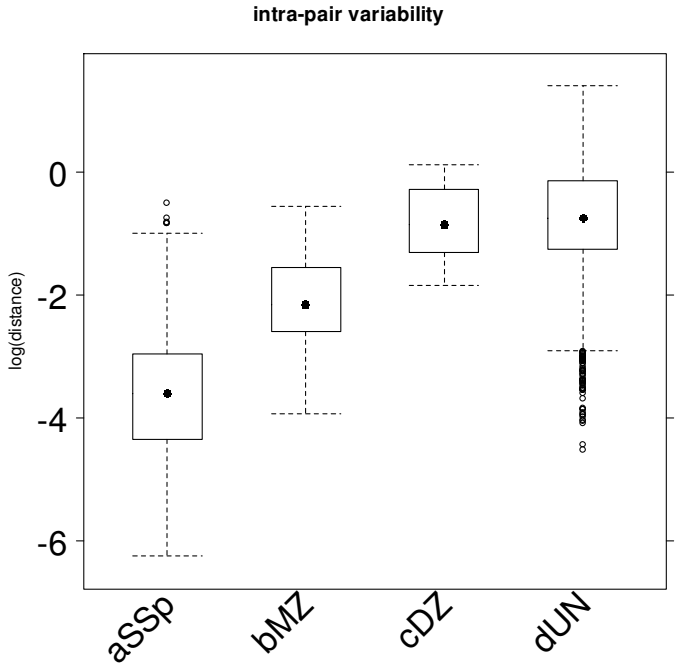


Figure 53: Boxplots of logarithmic Euclidean distances (log(distance)) separated into the four groups same speaker (aSSp), monozygotic twins (bMZ), dizygotic twins (cDZ) and unrelated speakers (dUN). The median of the distribution in each group is visualized by vertical lines in the boxes, the boxes comprise 50% of the data, the whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box, and outliers are marked with open dots.

To answer the question as to whether there is a significant effect of physiology on the pattern of the looping movement and which of the differences between the groups are significant, the above described model was run. A detailed overview of the model and the random effects is given in the appendix (Table D.2). Since the four levels of the fixed factor GROUP were put in the expected order, the model looks for significant contrasts between the neighboring levels and gives t-values, which can be seen in Table 29. As was expected from the visual inspection of the data in Figure 53, the t-values for the contrasting levels MZ vs. SSp and DZ vs. MZ are very high (22.5 and 15.5, respectively) and reveal significance ($p < 0.001$).

However, although it would not have been expected from the data in Figure 53, the contrast between the levels UN and DZ also reaches significance, albeit marginally ($p < 0.05$).

Table 29: Results for the fixed effects with t- and p-values.

	Estimate	Std. Error	t-value	p-value (pMCMC)[22]
(Intercept)	-1.95078	0.29458	-6.622	0.0001
Group MZ-SSp	1.46444	0.06491	22.562	0.0001
Group DZ-MZ	1.21376	0.07832	15.498	0.0001
Group UN-DZ	0.12582	0.05686	2.213	0.03

This somewhat surprising result might be explained in the following way. The model may be an inadequate description of the data due to the coding of the speakers' identities as levels of the random factors SPEAK1 and SPEAK2. That is, the occurrence of a given subject in the two positions is not balanced across the groups and the results observed in Table 29 may depend on the particular distribution of the speakers into the two positions. To discard this hypothesis, a simulation was run 100 times with different codings for the subjects (c.f. Baayen 2008) In each of the 100 simulations, for each pair of speakers, it was decided at random which speaker had to be coded as speaker1 and which as speaker2. Then the model was run over the 100 data sets and the values of the fixed effect coefficients (first column of Table 29) were stored. Table 30 shows the 95% confidence intervals for each of the coefficients computed over the results of the 100 runs. It can be observed that the only confidence interval which contains 0 (i.e. a switch from negative to positive values) is the one addressing the comparison between UN and DZ (last row of Table 30), suggesting that the contrast between these two levels is not significant. Thus, the EDs between the articulatory tokens are significantly different between SSp and MZ, and between MZ and DZ, but not between UN and DZ as displayed in Figure 53.

[22] P-values were calculated using an MCMC sampling (Markov chain Monte Carlo, see Baayen 2008). The relevant function that was used in R is pvals.fnc() which carries out an MCMC sampling (with 10000 samples by default) and also gives the p-values based on the t-statistic.

Table 30: Confidence intervals for the coefficients computed over the 100 model runs.

	2.5%	50%	97.5%
X.Intercept	-2.1227	-1.85807	-1.6358
Group MZ-SSp	0.9988	1.47971	1.8689
Group DZ-MZ	0.5129	1.44695	2.2186
Group UN-DZ	-0.6419	0.08940	0.7445

6.4 Summary and conclusion

To sum up, in general a great deal of variation could be found between the different speakers in the analyzed looping trajectories, but interestingly much more inter-speaker variation was apparent within the DZ twin pairs than within the MZ twin pairs, corroborating H1: physiology (NATURE) has an impact on articulatory gestures.

Already during the FDA and the alignment of the tokens within the speaker pairs it could be observed that the DZ twins reveal larger differences than the MZ twins. More interestingly, this effect seemed to be more obvious in the middle of the sequence, thus during the velar closure (see Figure 52). This finding was expanded in the qualitative analysis of the gestures: the MZ twins reveal very similar looping patterns in terms of the position and the general shape of the loop and the direction of the upward movement (straight or back and up). The DZ twins show striking differences in the horizontal and vertical position of the trajectory, the shape of the loop, the direction of the upward movement and the amount of horizontal sliding movement at the palate. These findings were also supported by the quantitative analysis investigating the amount of ED between the articulatory tokens: the fewest differences between the trajectories were found within a speaker, the most differences between unrelated speakers. But most interestingly, MZ and DZ twin pairs differ significantly in their amount of articulatory inter-speaker variability in terms of the looping trajectories.

These results support the assumption of an influence of biomechanics and physiology on the looping pattern in VCV sequences as suggested by Perrier et al. (2003). The geometry of the vocal tract and the size of the tongue influence the general spatial position of the trajectory in

the vocal tract; the palatal shape and biomechanical characteristics of the tongue muscles have an impact on the shape of the loop, the amount of horizontal sliding at the palate, and the directional patterns of the trajectory.

6.5 Limitations and further research

Regarding the quantitative analysis it should again be noted that the measured Euclidean distances do not mirror the movement pattern in a direct way but can only give some information about the comparison of the positional data. The difference in inter-speaker variability between MZ and DZ pairs mainly arises because of differences in the (horizontal) position of the trajectory. This might point to a crucial influence of the size of the palate and tongue. However, since the EDs are measured between each of the points that form the trajectory, the shape of the looping movement is indirectly considered, too. Moreover, the results of the qualitative analysis strongly support the idea of a more similar trajectory shape within the MZ twins in terms of the size and form of the loop.

This assumption could be investigated by defining significant points that determine the shape of the trajectory (like the start of the closing gesture, the start and the end of lingual-palatal contact, and the end of the opening gesture). Further research could then investigate the EDs between these points and describe the form of the loop in a quantitative way. A difficulty concerning this method is the exact determination of the points, in particular, the lingual-palatal contact. However, the tangential velocity of the tongue back coil could help to locate the relevant turning points of the loop.

The following chapter comprises the last result section of the present study. Here, the focus changes again, from articulation to acoustic parameters but also to auditory cues. Thus, in addition to articulation and acoustics, the perception of twins will be examined. In detail, the perceived auditory similarity in twins and unrelated speakers will be analyzed and discussed.

7 PERCEIVED AUDITORY SIMILARITY AND ACOUSTIC CORRELATES

7.1 Perceived auditory similarity

The analyses discussed above focused on *articulatory* and *acoustic* similarities and differences in twin pairs. In this way some light could be shed on explaining speaker-specific variability – which plays a crucial role in speech – in regard to the influencing factors NATURE and NURTURE. However, to understand the principal mechanisms in speech not only the production of speech (with its underlying articulatory strategies and the resulting acoustic outputs) but also the *perception* of speech has to be addressed. Therefore, this section will focus on perceived auditory similarity between different speakers (i.e. between speakers of MZ pairs, speakers of DZ pairs and unrelated speakers). Here, the influencing factors NATURE and NURTURE will be discussed in regard to their potential impact on *perceived* speaker identity.

7.1.1 Introduction

Perceived auditory similarity is a crucial topic in automatic speaker recognition as well as in forensic speaker identification. Here, the testimony of earwitnesses or descriptions of voices are important issues and several studies have investigated the ability of listeners to recognize and discriminate between speakers by their voices (e.g. Schiller & Köster 1996). The similarity of voices has been addressed by comparing unrelated speakers but also related speakers, such as siblings (Feiser 2009) and twins (Nolan & Oh 1996), the subject group under investigation in the present study. In general it can be assumed that, excluding differences in linguistic background and dialectal influence, related speakers, who share some amount of physiological characteristics due to genetic similarity, also have more similar sounding voice characteristics than unrelated speakers. Even though no strong effect of zygosity on acoustic similarity has

been found in the present study (DZ twins being most of time as similar as MZ twins), one hypothesis regarding perceived similarity could be that MZ twins are more difficult to distinguish than DZ twins. In previous studies, which have already been discussed in detail in Section 2.2, it has been found that MZ twins have very similar voice characteristics leading to perceived similarity (Whiteside & Rixon 2000, 2003, Decoster et al. 2001). Interestingly, Johnson & Azara (2000) found in their study that the perceptual difference between DZ twins was not larger than between MZ twins. Nevertheless, this result is limited by the fact that only one DZ pair participated in their perception experiment. Thus, no study seems to have compared DZ and MZ twins regarding their perceived similarity in more detail.

It is clear that the ability to discriminate between voices is dependent not only on the blood relationship of the compared speakers but also on several additional factors, such as the quality and length of the stimuli and the degree of auditory similarity between the compared voices, as well as on wheather the listeners are linguistically naïve or have a phonetic education and if they are familiar with the speakers (for an overview see Rose 2002). One interesting question addressed here is the required length of the stimuli that are compared. Is it possible to discriminate speakers by listening to a short word of only 0.35 s on average? The abovementioned perception study by Decoster et al. (2001) (see Section 2.2.2.1) revealed that listeners were able to discriminate twins while listening to a sentence in about 80% of the cases but showed a near to chance ability of about 60% when a sustained /a/ of only 2.5 s was used as stimulus. What happens if the stimulus is even shorter in length than that but contains more information, because a bisyllabic word is used in contrast to just one phoneme?

The participating twins in the present study were asked about their perceived auditory similarity and all of the MZ pairs reported that people have difficulties in identifying the siblings on the phone. One of the MZ twin pairs even stated that they themselves have problems distinguishing their voices when listening to them on a tape recorder. This finding has also been reported previously in the literature (Gedda et al. 1960, Cornut 1971). Of the three DZ pairs only one pair stated that people have problems distinguishing them on the phone. This (subjective) difference in perceived similarity is investigated in the following perception test and may be explained by acoustic characteristics that 1) were not accounted

for by the acoustic analysis carried out so far, or 2) might not be measurable in any straightforward sense even though they are crucial for the perceived auditory similarity.

To test for the variability in perceptual similarity within MZ twin pairs, DZ twin pairs and unrelated speakers and to search for possible acoustic correlates relevant for the perceived auditory similarity an AX discrimination test was carried out.

7.1.2 Hypotheses

Based on the abovementioned studies regarding perceived auditory similarities the following hypotheses are investigated:

H1) Listeners are able to distinguish speakers by listening to a short word (0.35 s on average).

H2) Unrelated speakers are easier to distinguish than related speakers (twin pairs).

H3) Speakers of DZ twin pairs are easier to distinguish than speakers of MZ twin pairs.

7.1.3 Method

7.1.3.1 Subjects

To restrict the overall length of the perception test, only the data of the female twin pairs was selected. Two of the pairs are dizygotic (DZf1 = LR & SR, DZf2 = MG & TG) and two pairs are monozygotic (MZf1 = AF & HF, MZf2 = GS & RS).

An AX (same-different) perception test was conducted and 30 native German listeners (13 male and 17 female) were asked to judge whether two stimuli they listened to belong to the same speaker or different speakers. Two male listeners had to be excluded due to their extremely low identification scores (their ratings were outside the normal range of the other listeners: over 40% of their answers were wrong even for unrelated pairs in contrast to 15% for the other listeners). This results in a total of 28 listeners (11 male, 17 female), who were on average 29.6 years old (SD = 6.4, Min =21, Max =45).

7.1.3.2 Stimuli

The stimuli consisted of different tokens of the word /vaʃə/ 'wash' 1st person sg., which were extracted from the sentence "*Ich wasche Haku/Hag(u,i,a) im Garten*" (I wash Haku/Hag(u,i,a) in the garden). The labeling and segmentation of /vaʃə/ was done manually with PRAAT and oriented on the oscillogram of the word (see Section 3.4.1 in the method chapter for details). For each of the 8 speakers 6 renditions of this word were selected and served as stimuli, resulting in a total of 48 different stimuli. To control for differences in intensity between the speakers or even between two renditions of the same speaker, the stimuli were normalized for intensity with the help of the software *Adobe Audition:* the highest amplitude/peak of each signal was normalized to 0db (which refers to 100%), and all other amplitude values of this signal were adjusted accordingly, so that the relation of different amplitudes in each signal was kept the same. As a result the perceived stimuli did not differ in their overall intensity and differences between the stimuli cannot have resulted from differences in produced loudness.

7.1.3.3 Perception test

For the preparation of the perception test the stimuli had to be permuted resulting in pairs of two stimuli. Each stimuli-pair was presented in both possible orders (AX and XA) to avoid auditory memory effects. The four resulting *speaker groups* of different stimuli-pairs are shown in Table 31.

Table 31: Speaker groups of the perception test; in group 3 (same speaker) different renditions of each speaker were paired.

Speaker group	Stimulus pairs	Correct answer
1	monozygotic twins (MZ)	different
2	dizygotic twins (DZ)	different
3	same speaker (SSp)	same
4	unrelated speakers (unrelated)	different

These pairwise comparisons result in a too large number of possible pairs, which had to be reduced due to time restrictions of the experiment. This was done in the following way: first, 4 *listener groups* (with 7, 6, 6 and 9 listeners) rated each of the possible speaker pairs but not all renditions of each speaker. In detail, this means that the 6 stimuli from each speaker (A, B, C, D, E, F) were distributed over 4 groups, each group contained 3 renditions. Each group consisted of different renditions and each stimulus was rated twice by two different listener groups resulting in the 4 stimuli groups (ABC, CDE, DEF, FAB). The distribution of the renditions was done for each speaker. Still, this led to a very large amount of data, especially of the stimulus pairs for the unrelated speaker group (4). This (control) group is a measure of the overall reliability of the listeners but also helps to locate similar sounding unrelated speakers. To reduce the number of stimuli-pairs within this group, not 3 but only 2 renditions were taken to build the stimuli-pairs (i.e. each of the six renditions was taken only once and not twice for the unrelated speaker pairs). In this way, each listener had to rate 432 different AX stimuli-pairs twice, resulting in 864 AX pairs. The presented audio signals consisted of 72 MZ, 72 DZ, 144 same speaker and 576 unrelated speaker pairs. This distribution is visualized in Figure 54. As an example, the 72 pairs for the MZ group result from 2 speakers x 3 renditions = 18 possible AX pairs x 2 MZ pairs = 36 x 2 repetitions = 72 stimuli-pairs.

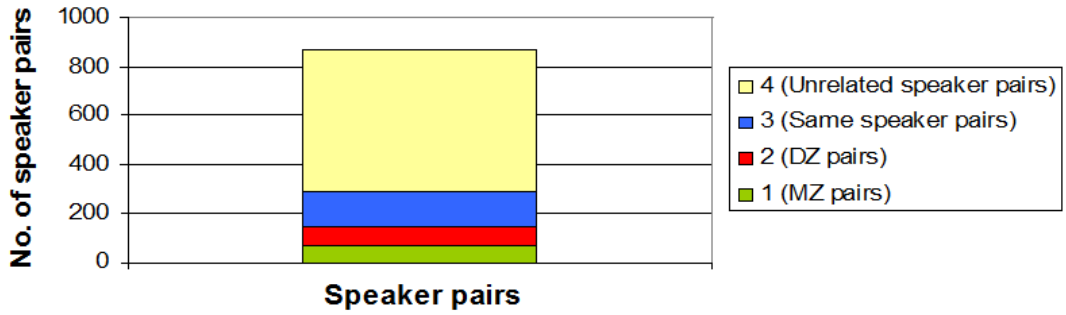


Figure 54: Distribution of the presented stimuli groups.

The perception test was run in PRAAT (via the embedded script *ExpMFC*) and took 40 minutes without breaks, but the listeners were given the chance to take a break whenever they wanted to. Subjects listened to each presented AX pair only once over *Sennheiser HD 595* headphones in a randomized order and were asked to click on a screen either the button "*same speaker*" or "*different speaker*" directly after listening to the stimuli.

After the test, the ratings of all listeners were collected and imported into an Excel data sheet for further examination. Each response was coded with *0* if it was correct and with *1* if it was false.

7.1.4 Probability of correct speaker discrimination

The judgments of the listeners were checked for their correctness and an overall reliability score was calculated for each listener. For the abovementioned speaker groups 1-3 the correct answer would have been "*different speaker*" and for group 4 "*same speaker*" (cf. Table 31). Figure 55 shows one subplot for each listener. Each subplot shows the percentage of correct (left bar) and false answers (right bar) for all responses.

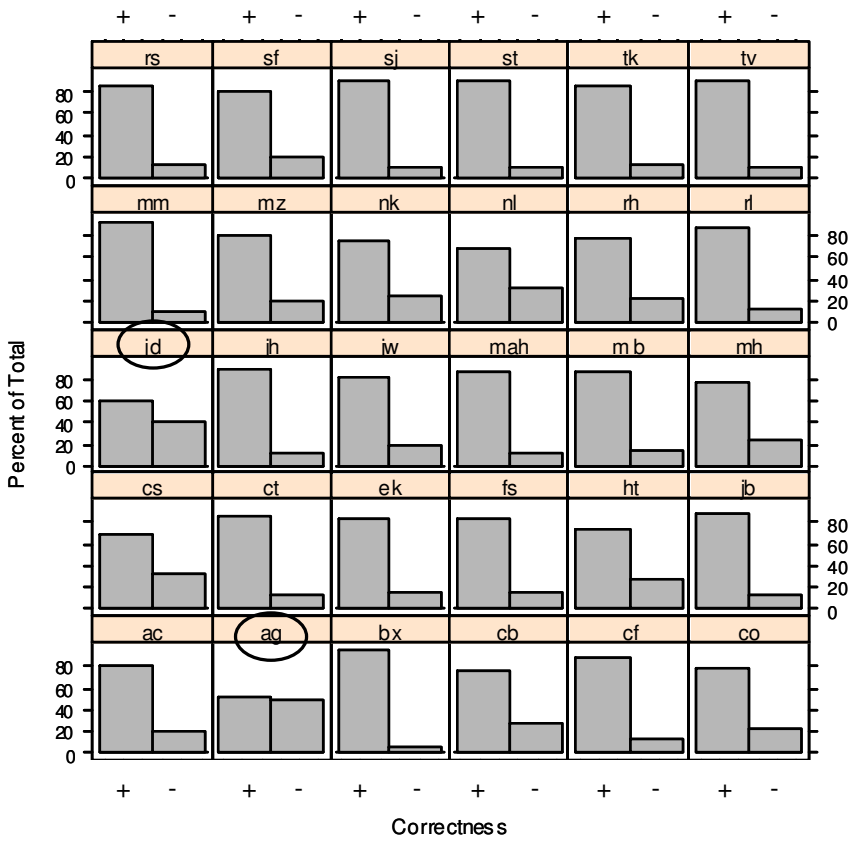


Figure 55: Overall scores (in percent) for each listener for correct (left bar) and false answers (right bar); the two excluded listeners (jd and ag) are indicated by black ellipses.

Two listeners stand out in their low reliability scores. AG and JD reveal 48% and 41% overall false ratings. Even for the unrelated speaker group their reliability scores do not reach over 60%. Because of their atypical answers they were excluded from further analysis. The remaining 28 listeners vary in their overall correctness scores from 67% to 95%. On average the listeners were able to differentiate same and different speakers in 82.8% of the cases, supporting H1 which states that listeners are able to distinguish speakers by listening to only one word. The following Figure 56 visualizes the cumulative correctness scores separated by speaker groups. Group 3 (*same speaker*) and group 4 (*unrelated speakers*) reach very high correctness scores of around 90%. However, for the twin groups 1 and 2 the percent of correct identification scores reached less than 50%. This result supports H2 and the difficulty of distinguishing related speakers. The listeners were not able to differentiate speakers of MZ and DZ twin pairs. The figure also gives the exact percentage of correct answers for each group. There is a weak tendency for MZ pairs to be the most difficult group to differentiate, but it is less than expected from Hypothesis 3 (46% correct answers for the MZ pairs vs. 48% correct answers for the DZ pairs).

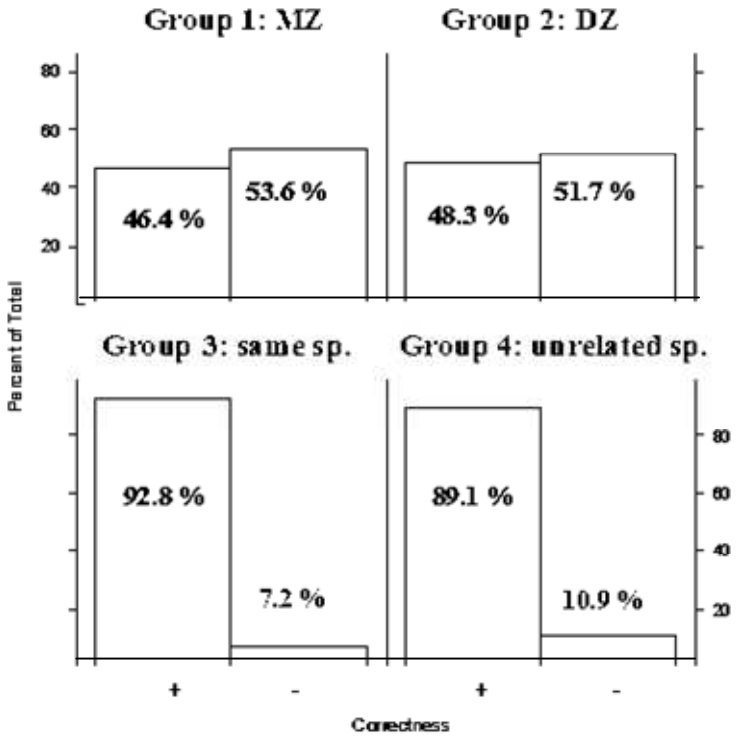


Figure 56: Overall scores (in percent) for the 4 different speaker groups of correct (left bar) and false answers (right bar).

7.1.5 Statistical analysis

7.1.5.1 Influence of the different speaker groups on perceived similarity

Statistical tests were conducted using R (version 2.8.1). To look for a significant effect of the 4 different groups on their correctness/error scores a generalized linear mixed model was calculated. Since the correctness ratings were binomial distributed (0 = correct answer, 1 = false answer), the logistic link function with family = *binomial* was selected to calculate the model. Remember that mixed models incorporate fixed and random factors. For our model the factor GROUP is taken as a fixed factor with the different speaker groups as fixed levels. Group 3 (*same speaker*) was excluded from the analysis since it only served as a control group. For the remaining groups the answer "*different speaker*" was the correct answer. (In this way, it could also be controlled for possible response tendencies of the listeners.) According to hypotheses 2 and 3, the factor GROUP (which now excludes the factor level *same speaker*) was ordered according to the expected error scores in the following way: *MZ - DZ - unrelated*, with *MZ* revealing the highest error score, *DZ* an intermediate error score and *unrelated pairs* the lowest error score. Table 32 gives an overview of the fixed factor, its levels and the order of the levels according to their expected error score.

Table 32: Overview of the fixed factor GROUP with different levels and order of levels.

Factor	Levels	Order according to expected error score
GROUP	MZ, DZ, unrelated speakers	MZ > DZ > unrelated

In addition to the fixed factor GROUP, the random factors LISTENER, (SPEAKER) PAIR and STIMULUS are included in the model. To determine the structure and relative importance of the random effects, different (lmer-) models with increasing complexity were compared in a stepwise fashion by means of a chi square test[23] (see Table E.1 in the

[23] The different models tested were: test0: correctness ~ group + (1 | subject); test1: correctness ~ group + (1 | subject)+(1|pair); test2: correctness ~ group + (1 | subject) + (1|pair) + (1|stimulus).

appendix). It turned out that the model with GROUP as a fixed factor and all three random factors LISTENER, (SPEAKER) PAIR and STIMULUS fits the data best:

Model: correctness ~ group + (1 | listener) + (1 | pair) + (1 | stimulus)

To test for significant different error scores between *MZ* and *DZ,* and *DZ* and *unrelated pairs* the contrasts between the different levels of the factor GROUP were calculated (as in Section 6 on /aka/). Table 33 displays the effects of the fixed factor. The summary statistics and the effects of the random factors can be seen in Table E.2 in the appendix. As expected, the difference between *unrelated pairs* and *DZ pairs* is highly significant ($p < 0.001$) (and hence also between *unrelated pairs* and *MZ pairs*), but no significant difference was found between the *MZ* and the *DZ* pairs (see Table 33, Group DZ-MZ, Group UN-DZ). Therefore, Hypothesis 3 and a higher perceived similarity in MZ than in DZ twins could not be confirmed.

Table 33: Overview of the fixed effect group (with the levels: MZ, DZ and UN(related) pairs).

	Estimate	Std. Error	z-value	Pr(>|z|)
(Intercept)	-0.8416	0.3035	-2.773	0.00555 **
Group DZ-MZ	-0.1641	0.7468	-0.220	0.82607
Group UN-DZ	-2.8178	0.5506	-5.117	3.10e-07 ***
Signif. codes: 0.0001 '***' 0.001 '**'				

7.1.5.2 Influence of the different speaker pairs on perceived similarity

Since the results do not indicate that zygosity is a strong factor, it is worth looking into the ratings for each speaker pair in more detail. Figure 57 shows the percentage of correct answers in relation to all possible speaker pairs. Each subplot shows the correct and false identification scores for the different speaker pairs (correct = left bar, false = right bar). The speaker pairs are named after the initials of the two speakers involved in this stimuli-pair: the first two letters denote one speaker, the last two letters denote the other speaker. For example, the combination RSRS indicates that both stimuli come from the same speaker (namely RS). Great differences can be seen in the correctness scores between the pairs. The 4 twin pairs, which are marked by black ellipses in the figure, clearly stand out in their high

number of false ratings (right bar). Interestingly, two of the four pairs have higher error scores, and thus seem to be more similar than the other two. The MZ pair GSRS and the DZ pair LRSR have the highest percent of false answers. The speakers GS and RS were falsely rated as same speakers in 68% of the cases and the speakers LR and SR in 62% of the cases (in contrast to 41% for the DZ pair MGTG and 39% for the MZ pair AFHF). This is in line with the statistical analysis discussed above and the result that the factor zygosity (the difference between Group DZ- MZ) did not show a significant effect on perceived speaker similarity.

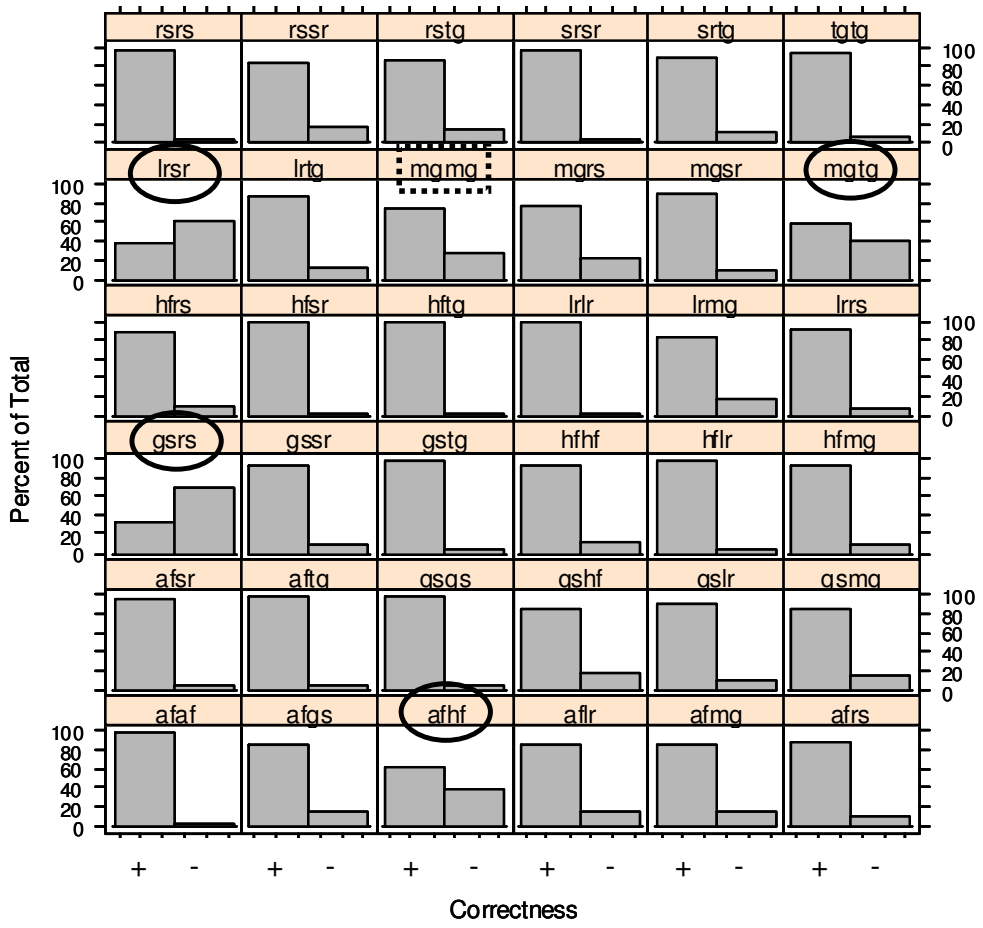


Figure 57: Overall percentage of ratings separated by speaker pair; left bar = correct answer, right bar = false answer; twin pairs are marked by circles; the dotted rectangle indicates the speaker (MG) with a high number of false ratings when compared with herself.

Another interesting finding that can be seen in Figure 57 is the fact that the speaker MG (marked by a dotted line) receives a strikingly high probability of false responses when listeners are asked to rate two renditions by this speaker. Thus, in 27% of the cases, two stimuli produced by MG were rated as coming from different speakers, even though the false answer probability for the group *same speaker* is only 7.2% on average (see Figure 57). This finding shows that this speaker realizes several repetitions of the same word with a huge variability.

In the following section, an acoustic analysis is conducted of certain parameters of the segment /vaʃə/. The aim of the analysis is to find acoustic correlates of perceived speaker similarity and hence possible reasons for the variation in perceived similarity of the twin pairs (with one MZ pair being the most and the other MZ pair being the least similar) and the strikingly high number of false ratings for speaker MG.

7.2 Finding acoustic correlates: An acoustic analysis of the rated stimuli

7.2.1 Introduction

Given that differences in the perceived similarities of MZ and DZ speakers were found, an acoustic analysis of the stimuli was carried out to look for acoustic correlates that might be responsible for the differences in ratings. Oriented on earlier literature discussed in the twin chapter (Chapter 2), several acoustic parameters were investigated to explain the results of the perception experiment. In contrast to the acoustic analysis in the previous chapters, here the focus lies on source-based parameters and characteristics related to perceived voice quality, since these parameters are considered to be crucial in perceived auditory similarity and are related to physiology (NATURE). Note, though, that the investigated parameters can only be seen as a reference and cannot cover all the information provided by the acoustic speech signal that might be used by the listener to discriminate between the speakers. The parameters were chosen based on a) their possible impact on perceived speaker similarity, and b) their relation to physiology and genetics.

Some studies conducting perception experiments and speaker discrimination tasks also investigated different acoustic parameters in regard to their influence on perceived speaker similarity. The most investigated parameter is *fundamental frequency* (F0) and several studies have found it to be crucial (Decoster et al. 2001, Sørensen acc.). Therefore, one hypothesis is that speaker pairs that show a very similar F0 should be mixed up more often than pairs with larger differences in F0. Since it has been shown in previous studies that F0 is influenced by organic and physiological constraints, it is also hypothesized that MZ twins are more similar in their fundamental frequency than DZ twins due to their greater physiological similarity (Przybyla et al. 1992, Debruyne et al. 2002) (see Hypothesis 1a below). In contrast, in the previously reviewed study of Debruyne et al. (2002) (see Chapter 2) it was found that the *variation in F0* seems to be less physiologically determined since MZ and DZ twins revealed the same amount of similarity in their study. Therefore, no difference in variation in F0 is hypothesized between MZ and DZ pairs. In what way the differences in F0-variation might

have an effect on perceived similarity is not known, thus this will be investigated in the following acoustic analysis (see Hypothesis 2 below).

In the comprehensive twin study of van Lierde et al. (2005) different voice quality characteristics (e.g. perceptual voice characteristics, maximum phonation time, vocal performances, overall vocal quality by means of the Dysphonia Severity Index) were investigated and found to be very similar within the MZ twins, thus a strong relation to organic parameters can be assumed. Interestingly, the two parameters jitter (micro-perturbations in frequency) and shimmer (micro-perturbations in amplitude) revealed no similarity within the twins. These parameters might be correlated to environmental factors such as state of health or to situational contexts like anxiety or tension, thus they may be independent of physiological constraints; therefore, they will also be examined in the following analysis. Shimmer is also known to be a physical correlate of perceived "hoarseness" (Lieberman 1963, Deal & Emanuel 1978), which might be a crucial parameter in discriminating voices. In addition to shimmer, the harmonics-to-noise ratio (HNR), which relates the harmonic level of a signal to its noise level, also correlates with perceived hoarseness and breathiness. Johnson & Azara (2000) mention in their perception study the factor breathiness as a possible auditory cue in perceived similarity. Therefore, the voice quality parameters shimmer, jitter and HNR are investigated (see Hypothesis 3 below).

7.2.2 Hypotheses

The following hypotheses are derived from the studies discussed above. H1 (a and b) addresses the parameter fundamental frequency (F0), H2 the variation in fundamental frequency, and H3 the voice quality characteristics jitter, shimmer and HNR.

H1a) Mean fundamental frequency (F0) should be more similar in MZ than in DZ twins due to physiological constraints.

H1b) Mean fundamental frequency (F0) should be more similar in speaker pairs with high error rates (and thus high perceived speaker similarity).

H2) Variation in fundamental frequency differs equally in MZ and DZ twins, since no physiological constraints are assumed.

H3) Voice quality parameters like jitter, shimmer and HNR (harmonics-to-noise ratio) are crucial in perceived speaker similarity and are influenced by environmental factors, thus they should be most similar in the twins that were confused most often (MZf2 and DZf1).

7.2.3 Acoustic analyses

The acoustic parameters were analyzed in PRAAT. Altogether 6 acoustic parameters were chosen due to their importance for perceived speaker identity (see above). The parameters 1 (A-C) are F0-related; the parameters 2 (A-C) are measures of voice quality and correlates of perceived hoarseness.

(1) Mean F0 and F0-variation

A	Mean fundamental frequency (Mean_F0)
B	Variation (normalized standard deviation) in the mean fundamental frequency (Std_norm)
C	Slope of fundamental frequency over the whole stimulus (F0_contour)

(2) Voice quality

A	Jitter
B	Shimmer
C	Harmonics-to-Noise Ratio

7.2.3.1 Mean F0 and F0-variation

Mean and standard deviation of the fundamental frequency (F0) were calculated over 6 repetitions of each speaker of the sequence /vaʃə/. The following standard adjustments in PRAAT were used for all speakers: positive time step of 0.01, minimum of 50 Hz and

maximum of 400 Hz. The standard deviation is dependent on the fundamental frequency. For a male speaker with an average F0 of 120 Hz, a standard deviation of +/-20 Hz would be perceived as 'normal' average F0-variation. For a female speaker with a higher F0, a higher standard deviation is also expected. To control for this difference in perceived F0-variation a normalized variation coefficient that is independent of the F0 was calculated for an objective comparison (Künzel 1987). The normalized variation (Std_norm) was given as a percentage of the mean F0. The following formula with s = standard deviation in Hz and x = fundamental frequency in Hz was used:

$$Std\_norm = \frac{100 \cdot s}{x}$$

Mean fundamental frequency in Hz (**Mean_F0 = A**) and mean normalized standard deviation in percent (**Std_norm = B**) were calculated over all repetitions for each speaker. Additionally, the slope of the fundamental frequency over the whole sequence /vaʃə/ was calculated and the resulting **F0-contours (C)** for each repetition of all speakers were plotted and compared. To do this, time-normalized F0-values were extracted from the signal in the following way: for each signal the same number of time points (in this case = 10) was defined and the corresponding F0-values were extracted (this time the minimum was set to 30 Hz, the maximum to 400 Hz). Hence, 10 values of F0 (at 10%, 20%, 30% etc. of the time) for each signal were measured and an F0-contour with the same length for each signal (interpolated over the 10 measured values) could be estimated and plotted. No periodic signal (and thus no F0) is assumed to be present in the voiceless sibilant in the sequence /vaʃə/, thus no calculation of F0 was conducted for the period of the sibilant, and the F0-values at the time points 6, 7 and 8 (corresponding to 60, 70 and 80% of the elapsed time of the speech signal) were excluded. A dotted line in the F0-contours discussed in the result section marks the interpolated part.

7.2.3.2 Voice quality

In addition, the following voice quality parameters were investigated: j*itter, shimmer* and *harmonics-to-noise ratio* (HNR). All of them were measured over the labeled vowel /a/ in /vaʃə/.

Note that a comparison of jitter and also HNR values is only reliable for measurements of the same vowel.

(A) Jitter

Jitter is an acoustic measurement of how much a given period differs from the period that immediately follows it and accounts for minimal frequency perturbations (see Baken & Orlikoff 2000). Several jitter measurements exist; here, an F0-adjusted measurement is taken. The measurement of *relative average perturbation* (RAP) reduces the effect of relatively slow changes in F0 and estimates the lengths that a period *should* have according to the adjacent cycles without jitter influence (= three-point period perturbation quotient). The difference between the real period values and their (corrected) estimates is the degree of jitter. PRAAT also gives several jitter measurements; here the **PPQ5,** which is the five-point period perturbation quotient, was chosen. It is similar to the RAP but takes the four closest neighbors into account (RAP only looks at the adjacent cycles). Thus, PPQ5 gives an average absolute difference between a period and the average of it and its four closest neighbors, divided by the average period. Regarding our perception experiment it has to be kept in mind that jitter or frequency perturbation is a physical correlate of perceived "hoarseness" (Lieberman 1963, Deal & Emanuel 1978).

(B) Shimmer

Shimmer or amplitude perturbation quantifies the short-term instability of the vocal signal and is based on the peak amplitude of each phonatory cycle (Baken & Orlikoff 2000). It seems to be at least as important as jitter in its contribution to the perception of hoarseness (Wendahl 1966a, b; Takahashi & Koike 1975). The Amplitude Perturbation Quotient (APQ, Takahashi & Koike 1975[24]) compensates for long-term changes, and can be interpreted analogously to RAP for the amplitude variation: APQ gives an average absolute difference between the amplitude of a period and the average of the amplitude of its neighbors, divided by the average amplitude. Davis (1979) found it to be optimal for 4 neighbors. Thus, equivalent to PPQ5, **APQ5** was chosen as the measurement for shimmer.

[24] Takahashi & Koike (1975) used an 11-point average that takes into account the amplitudes of 10 neighboring periods.

(C) HNR

The harmonics-to-noise ratio (**HNR**) gives a measurement of the perceived hoarseness and aspiration and is expressed in dB. It describes and quantifies the features that appear in the spectrogram of the hoarse voice and relates the harmonic level of a signal to its noise level. Thus, it is the mean amplitude of the average wave divided by the mean amplitude of the isolated noise components for the train of waves (Baken & Orlikoff 2000). It was established and improved by Yumoto et al. (1982) and Yumoto et al. (1984). Yumoto et al. (1982) found a mean HNR for normal subjects of 11.9 dB (SD = 2.32; range 7.0 to 17.0) in a sustained [a]. They also found a correlation of 0.809 between the HNR and psychophysical scaling of hoarseness (Yumoto et al. 1984). PRAAT uses an algorithm that performs an acoustic periodicity detection on the basis of a forward cross-correlation analysis with a time step of 0.01 s, a minimum pitch of 75 Hz, a silence threshold of 0.1, and one period per window (these default values are also used in this analysis).

To summarize, 6 acoustic measurements were obtained and average values were calculated for each speaker: mean fundamental frequency (Mean_F0), variation in fundamental frequency (Std_norm and F0_contour), jitter (PPQ5), shimmer (APQ5) and hoarseness (HNR). Inter-speaker variability (ISV) of these parameters will be investigated especially within each twin pair.

7.2.4 Relation between perceived similarity and voice quality in twins' speech

7.2.4.1 Mean F0 and F0-variation

The upper part of Figure 58 (58a) shows the mean fundamental frequency (Mean_F0) and its standard deviation measured in the word /vaʃə/ for each of the 8 speakers. The speakers of the DZ pairs are on the left side of the black vertical line, and the MZ pairs are on the right side. First of all, it is obvious that the DZ twins have higher fundamental frequencies than the MZ twins. This might be explained by their younger age – both DZ twin pairs are 20 years old, whereas the MZ pairs are 34 (MZf1) and 26 (MZf2) years old. However, it is more likely due to normal inter-speaker variation between individuals. Furthermore, it can be seen that

the difference between the speakers within one twin pair is much higher for the DZ pairs than for the MZ pairs.

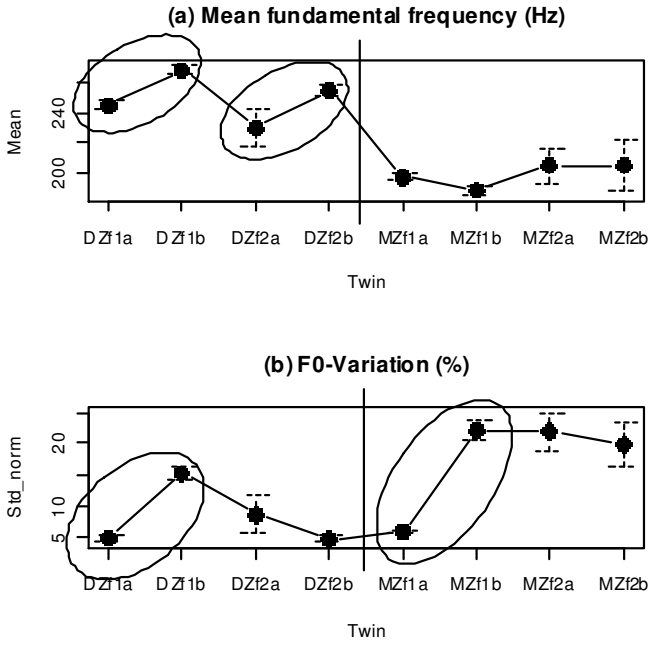


Figure 58 (a,b): Plots of mean fundamental frequency in Hz (a) and mean normalized variation in fundamental frequency in percent (b) split by speaker; MZ and DZ pairs are separated by the vertical line in the middle of the figures; pairs with the greatest differences are marked with circles.

Table 34 gives the corresponding values for mean fundamental frequency (measured over the 6 repetitions of each speaker), its standard deviation and the amount of ISV (difference in Mean_F0 in Hz between the speakers) within the pairs. Both DZ pairs reveal differences of over 20 Hz in their mean fundamental frequency, whereas the MZ pairs show either no variation at all (MZf2) or less than 10 Hz (MZf1). A Wilcoxon rank sum test[25] was conducted to look for significant differences within the pairs; only DZf1 revealed a significant difference with $p < 0.01$ (DZf2 failed to reach significance with $p = 0.064$, which might be due to the high intra-speaker variance of DZf2a). Although the difference between the mean F0 for the

[25] An F-test was conducted to test for equality in variances of the speakers within one pair and revealed significant differences. Therefore, a Wilcoxon rank sum test was conducted instead of a normal t-test.

speakers of MZf1 is only 9 Hz, it turned out to show significance (p = 0.041). MZf2 reveals no significant difference. Nevertheless, the magnitude of the differences in mean F0 within the pairs (DZ over 20 HZ, MZ less than 10 HZ) points to an influence of zygosity on this parameter, meaning that MZ twins are more similar in F0 than DZ twins. The factor biology and hence physiological constraints seem to be crucial in regard to F0 and H1 is thus supported. With respect to the results of the perception experiment, and thus the perceptual similarity of the twins, the higher differences in Mean_F0 for the DZ twins do not seem to be relevant, since DZf2 and MZf1 were the speaker pairs that were mixed up most often.

Table 34: Mean fundamental frequency (Mean_F0), standard deviation (SD), and difference in Mean_F0 (ISV) within the twin pairs in Hz. Significant differences are marked with asterisks (= p < .05, ** = p < .001).*

Twin pair	Twin red	Mean_F0	SD	Twin blue	Mean_F0	SD	ISV_F0
MZf1	AF	197.9	4.22	HF	188.8	6.98	9.1 (*)
MZf2	RS	205.1	37.91	GS	205.1	26.01	0.0
DZf1	LR	245.2	5.56	SR	268.6	6.84	23.4 (**)
DZf2	TG	254.9	8.57	MG	230.4	28.96	24.5

The lower part of Figure 58 (58b) shows the mean variation in the fundamental frequency in the sequence /vaʃə/. Some speakers reveal quite high variations of more than 20% (40 Hz), some show less than 10% (20 Hz) variation in their fundamental frequency. The greatest amount of inter-speaker variation within a pair was shown by DZf1 and MZf1.

To get a better impression of the slope of the fundamental frequency during the sequence /vaʃə/, the F0-contour of each repetition for all speakers is plotted in Figure 59. Since no periodic signal (and thus no F0) is assumed to be present in the voiceless sibilant /ʃ/, no calculation of F0 was conducted for this part of the signal and a dotted line in the graphs shows the interpolated contour. The two speakers of the same twin pair are plotted in one graph and separated by the colors red and blue (parallel to their colors in the previous analyses).

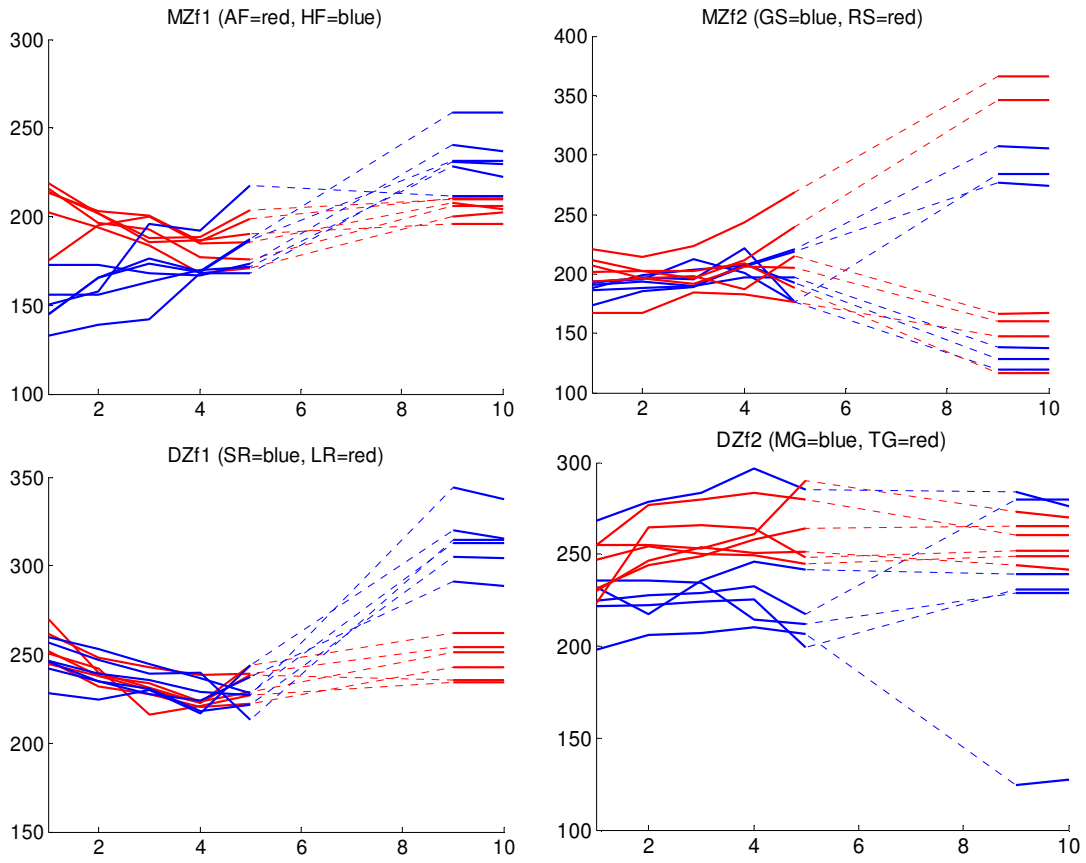


Figure 59: Time-normalized F0-contours interpolated over measured F0-values at 10 time points (x-axis) during the sequence /vaʃə/ for each repetition and all speakers. Speakers of the same pair are plotted in one graph and indicated by different colors; the dotted lines in the graphs reflect the interpolated contour for the voiceless sibilant /ʃ/ for the time points 6-8.

The MZ pairs are plotted in the upper part of the figure, the DZ pairs in the lower part. MZf2 reveals he most interesting F0-contours: some of the repetitions rise at the end of the sequence and some fall. But this is the case for both speakers. They have in common that they reveal a high amount of F0-variation and that this variation may be due to a falling or rising F0 at the end of the sequence (cf. Figure 58b: high mean values of F0-variation and high standard deviations for MZf2a and MZf2b). The other MZ pair (MZf1) shows slightly different F0-contours. HF (blue) has a rising F0-slope whereas her sister AF (red) does not show much variation at all. This verifies the difference in mean F0-variation (of about 30 Hz) plotted in Figure 58b. The same inter-speaker variability can be found in DZf1. Here, SR (blue) shows a rising F0-contour and her sister LR (red) reveals a more monotonous pronunciation of the word with a relatively stable F0. This again confirms the obvious

difference in F0-variation from Figure 58b. The pair DZf2 is interesting with respect to one repetition of MG (blue) that stands out from the overall pattern: the F0-contour is very stable for both speakers and nearly all repetitions. But one repetition of MG (blue) shows a strongly falling F0-slope at the end of the sequence. Looking again at Figure 57, it is conspicuous that MG had the most false ratings within the group *same speaker*. This seems to be at least partly due to the huge difference in F0-contour between one repetition and the rest of the stimuli. Here, the importance of F0-variation within one word for the perceived speaker similarity is apparent. However, with respect to the results of the perception experiment and the most similar speakers of the pairs MZf2 and DZf1, the differences in F0-contour do not seem to be crucial, since DZf1 reveals differences in this parameter. In addition, no effect of zygosity could be found, since the DZ twins do not show a higher amount of F0-variation than the MZ twins. Thus, the findings corroborate H2 and the minor role of biological identity regarding intonation and variation in F0 (in one word) in (normal) speakers.

7.2.4.2 Voice quality

In the Table 35 below the measured mean values for shimmer (APQ5), jitter (PPQ5) and HNR are shown for each speaker separately. Welch two-sample t-tests were conducted within all twin pairs to find significant differences. Bold numbers correspond to significant differences between speakers of a twin pair. Especially the pair MZf1 deserves attention since it shows high inter-speaker variation in all voice quality measures. The greatest difference appears to be in the HNR measurement, which is associated with perceived hoarseness. Remember that Yumoto et al. (1982) found a mean HNR for normal subjects of 11.9 dB, and the lower the HNR the higher the perceived hoarseness. HF shows a low mean HNR value of 9.83 dB and differs from the value of her sister by 6 dB. Also, the jitter value of HF is twice as high as that of her sister AF. HF reveals the highest PPQ5 and APQ5 values of all speakers. In addition to the low HNR values this supports the impression of a hoarse and breathy voice. This could explain the results of the perception test: MZf1 was the twin pair that was mixed up least, hence, voice quality measures mirroring perceived hoarseness (high jitter and shimmer, low HNR) seem to help in distinguishing similar sounding voices like the voices of twins.

Table 35: Mean values and standard deviations for APQ5, PPQ5 and HNR; significant differences in bold ($p < .05$ for APQ5 and PPQ5, $p < .01$ for HNR).

Twin pair	Twin (n=6)	APQ5	SD	PPQ5	SD	HNR	SD
MZf1	AF	**0.0351**	0.0122	**0.0060**	0.0021	**15.84**	1.770
	HF	**0.0555**	0.0147	**0.0125**	0.0049	**9.83**	3.186
MZf2	RS	0.0470	0.0166	0.0061	0.0029	12.00	3.370
	GS	0.0365	0.0087	0.0044	0.0013	13.62	0.733
DZf1	LR	0.0320	0.0100	0.0069	0.0027	**15.37**	1.316
	SR	0.0269	0.0027	0.0041	0.0007	**18.30**	1.525
DZf2	TG	0.0256	0.0132	0.0041	0.0024	14.44	3.010
	MG	0.0359	0.0190	0.0039	0.0012	14.01	4.650

The plots in Figure 60 visualize the measured voice quality values given in Table 35, and here, too, the high within-pair variability for MZf1 is noticeable (values for speaker HF are marked by a circle). Additionally, it becomes clear that no effect of zygosity can be assumed regarding the investigated parameters since the pairs on the left side of the black line (DZ pairs) do not show more inter-speaker variability than the pairs on right side of the line (MZ pairs). These results suggest that the influence of biology and shared physiology in healthy speakers is rather negligible, whereas different environmental influences can affect voice quality. Note that this finding is restricted by the (physiological) fact that all speakers were female and around the same age. One possible cause for a hoarse voice quality is the habit of smoking, but since none of the speakers were heavy smokers, this cannot be the reason. However, speaker HF (= MZf1b) is a teacher and her voice seems to reflect the extensive use of speech during her workday in a hoarse voice quality.

DZf2 is a second twin pair that was distinguished correctly more often than it was confused in the perception test. In looking at the voice quality parameters, no great differences between the two speakers are apparent for HNR, jitter or shimmer. There may be other audible differences in their voices that are not accounted for by the acoustic measurements in this investigation. However, the great difference in mean fundamental frequency seems to be crucial and especially the strongly differing stimulus of MG with an extremely falling F0-contour at the end of the sequence. Another stimulus of MG stands out because of a very low HNR value (5.4 dB). These two stimuli were rated most of the time as coming from a

different speaker when compared to her sister TG but also when compared to herself. This might explain the difference in perceived speaker identity and the relatively high rating of "*different speaker*" in the perception test for the group *same speaker* for MG and the *twin group DZf2* (with MG and TG). When the two conspicuous stimuli are excluded from the data, the resulting error score for DZf2 increases from 41% to 51% and the probability of a false rating of two stimuli from MG as coming from different speakers decreases from 27% to 18%. Regarding the influence of voice quality parameters in perceived speaker identity, H3 can be partially supported. In distinguishing the pair MZf1 the differences in the investigated parameters *jitter*, *shimmer* and *HNR* are crucial and helpful, but other characteristics of an acoustic stimulus (like extreme values in F0 and F0-variation) can be more important.

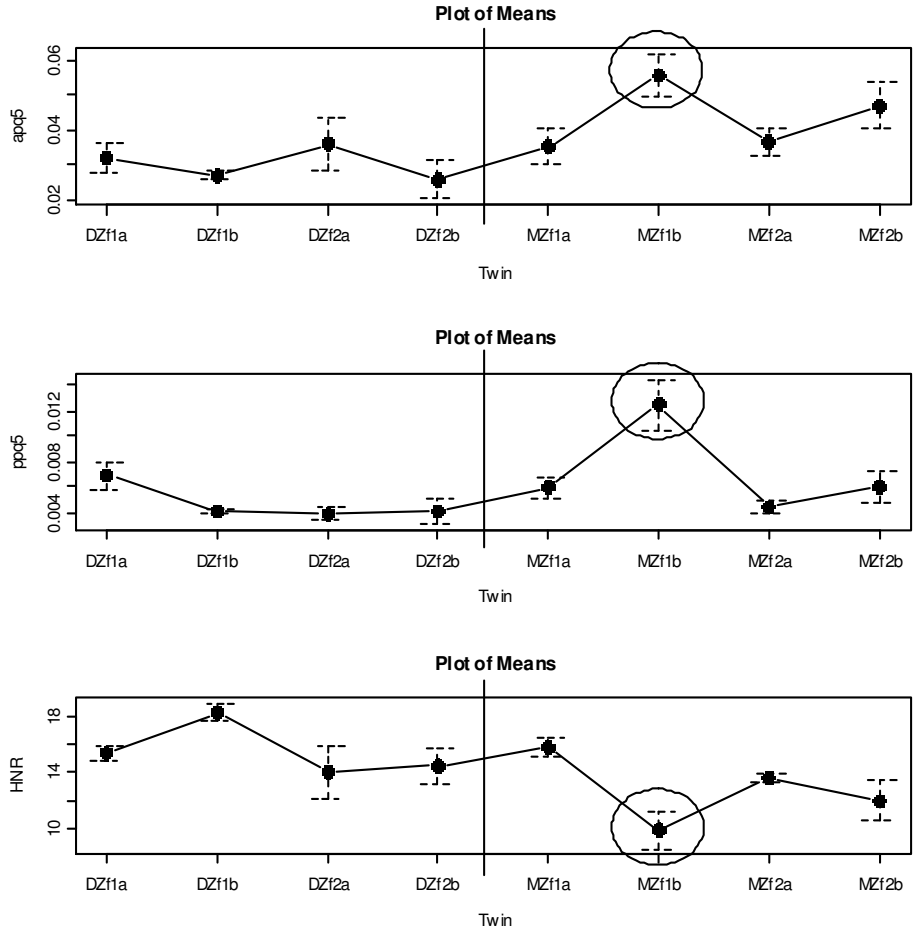


Figure 60: Plots of mean values for APQ5, PPQ5 and HNR separated by speaker; circles indicate the speaker with striking voice quality parameters (i.e. high jitter and shimmer, low HNR); MZ and DZ pairs are separated by the vertical line in the middle of the figure.

7.2.5 Relation between perceived similarity and voice quality in unrelated speakers

Here, a closer look is taken at the differences in perceived similarity of *unrelated* speakers and the acoustic characteristics correlated with these differences, thus testing H1a, which assumes speakers with similar F0s are mixed up more often.

In the following table the measured differences in the analyzed acoustic parameters between all possible speaker pairs (without group *same*) are given. The values show the differences between the calculated averages for each speaker in mean F0, F0-variation (here again the normalized variation coefficients are compared), APQ5, PPQ5, and HNR. In addition, the degree of perceived similarity that is expressed in the error score of the perception test is shown in the last column. If the error score is high, the perceived similarity is high, and speakers of this pair were difficult to differentiate and often confused. The first 4 rows show the twin groups and reveal the highest error scores as discussed above. The remaining 24 unrelated pairs differ in their error scores (from 0.01 for HFTG to 0.24 for MGRS).

Table 36: Differences in acoustic parameters and error scores in the perception experiment (PE) of all speaker pairs ordered from highest error score to lowest; differences in APQ5 and PPQ5 have been multiplied by 1000 due to very small values.

Group	Pair	Δ F0	ΔSD_norm	Δ APQ5 (x 1000)	Δ PPQ5 (x 1000)	Δ HNR	Error score in PE
MZf2	GSRS	0.00	1.29	10.48	1.56	1.62	0.68
DZf1	LRSR	23.35	10.35	5.07	2.79	2.94	0.62
DZf2	MGTG	24.59	3.26	10.32	0.19	0.43	0.41
MZf1	AFHF	9.06	16.23	20.43	6.49	6.01	0.39
UN1	MGRS	25.30	12.60	11.01	2.15	2.01	0.24
UN2	LRMG	14.85	3.06	3.98	3.01	1.35	0.18
UN3	RSSR	63.49	5.30	20.06	1.93	6.31	0.16
UN4	AFMG	32.54	2.09	0.85	2.12	1.83	0.16
UN5	GSHF	16.31	0.25	19.04	8.02	3.79	0.16

UN6	AFLR	47.38	0.97	3.12	0.88	0.47	0.16
UN7	GSMG	25.29	13.89	0.53	0.59	0.39	0.15
UN8	AFGS	7.25	15.98	1.39	1.53	2.22	0.15
UN9	RSTG	49.89	15.86	21.32	1.96	2.44	0.14
UN10	LRTG	9.75	0.20	6.34	2.81	0.92	0.13
UN11	SRTG	13.61	10.56	1.27	0.03	3.86	0.11
UN12	HFRS	16.31	1.54	8.57	6.46	2.17	0.11
UN13	AFRS	7.24	14.69	11.86	0.03	3.84	0.11
UN14	GSLR	40.14	16.94	4.51	2.41	1.74	0.10
UN15	HFMG	41.60	14.14	19.58	8.61	4.19	0.10
UN16	MGSR	38.20	7.30	9.05	0.22	4.29	0.09
UN17	LRRS	40.14	15.66	14.98	0.85	3.37	0.09
UN18	GSSR	63.49	6.59	9.58	0.37	4.68	0.08
UN19	AFSR	70.73	9.38	8.19	1.90	2.47	0.06
UN20	GSTG	49.88	17.15	10.85	0.40	0.82	0.05
UN21	AFTG	57.13	1.17	9.46	1.93	1.40	0.04
UN22	HFLR	56.45	17.20	23.55	5.61	5.54	0.03
UN23	HFSR	79.80	6.85	28.63	8.39	8.48	0.03
UN24	HFTG	66.19	17.40	29.89	8.42	4.61	0.01

To look for a possible influence of the different acoustic parameters on perceived similarity, Pearson correlations were calculated between the acoustic measurements and the error scores (remember: a high error score reflects a high perceived similarity). If the difference in an acoustic parameter is high, the error score (and hence the perceived similarity) should be low, given that this parameter has an influence on the perceived speaker identity. Figure 61 shows the relation between differences in mean F0 and error scores for all pairs. The highest error scores, which are indicated by the green color, are those of the four twin pairs.

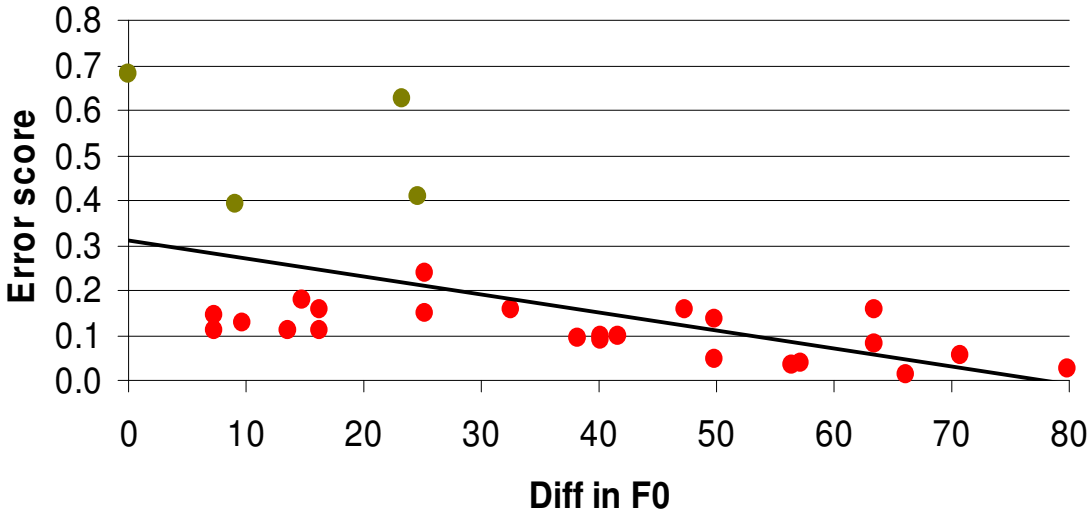


Figure 61: Scatterplot with trend line of error scores (y-axis) and differences in F0 (x-axis) for twin pairs (green) and unrelated speakers (red).

As seen in the acoustic analysis above different pair-specific parameters are important when siblings with generally very similar sounding voices are compared. Small differences in only one of the mentioned acoustic parameters can be crucial. For the more global investigation of the relation between perceived similarity and acoustic parameters in speakers in general only the data of the *unrelated pairs* (UN1-UN24) was taken for the correlations. This was done to disregard this time the above discussed special case of extremely similar sounding voices and extremely similar physiology as is the case with twins. Table 37 shows the results of the calculated correlations (Pearson) with the respective r, R^2 and p-values for the data of the unrelated pairs. It can be seen that correlations for all acoustic parameters are negative, revealing the expected negative effect of acoustic differences on perceived similarity. However, only F0 ($t = -3.6$, $df = 22$) and APQ5 ($t = -2.4$, $df = 22$) show a significant effect with p-values smaller than .05. HNR marginally fails to show significance with $p = 0.0544$ ($t = -2.0$, $df = 22$). The strongest impact on perceived similarity turns out to be F0, with $r = -.61$ and $R^2 = 0.37$, meaning that over one-third of the variance in error scores can be explained by the difference in F0, corroborating H1a.

Table 37: Correlations between error scores and acoustic parameters of unrelated pairs (without twins); r and p-values in bold if significant (p < 0.05).

Pearson correlations between error score AND acoustic parameters ONLY UNRELATED PAIRS					
	Diff_F0	Diff_Std_norm	Diff_APQ5	Diff_PPQ5	Diff_HNR
r	**-0.6110**	-0.2713	**-0.4617**	-0.2731	-0.3975
R^2	0.37	0.07	0.21	0.07	0.16
p	0.0015	0.1997	0.0231	0.1967	0.0544

Interestingly, F0 was not the major factor when distinguishing the voices of twins, since the MZ twins were more similar in their F0 with respect to the magnitude of difference in mean F0 than the DZ twins, and the DZ twin pair with significant differences in F0 (DZf2) was mixed up more often than one of the MZ pairs (MZf1). Hence, F0 seems to be crucial when comparing unrelated speakers but not when comparing similar sounding voices like those of siblings. Here, other parameters, like voice quality parameters, become significant. Differences in F0-variation seem to play a minor role when distinguishing speakers based on only one-word stimuli.

7.3 Summary and conclusion

Two of the three hypotheses regarding perceived speaker similarity were confirmed: 1) listeners need only very little input in order to differentiate voices; they were able to distinguish different speakers by listening to only one short word (with an overall correctness score of 82.8%), and 2) unrelated speakers are much easier to keep apart than twins (correctness score of 90% vs. 47%) even if the listeners have very little evidence to go by. The third hypothesis could not be corroborated. Thus, speakers of DZ twin pairs are not easier to distinguish than speakers of MZ twin pairs. This is in line with the results of the acoustic analysis of vowels and sibilants in the previous chapters. No strong influence of zygosity on inter-speaker variability in acoustic targets was found, pointing to auditory goals and the influence of shared environment. This mirrors the results of the perception test with similar sounding voices for MZ and DZ twins. An interesting finding is that the differences in formant transitions discussed above (see Section 5.2.6) seem to be irrelevant for the auditory similarity of speakers. The DZ pairs that showed differences in the formant transitions of /ʃə/ in the word /vaʃə/ that was used for the perception experiment were not easier to distinguish than the MZ twins with no differences in formant transitions. Thus, auditory cues do not seem to be crucial in sibilant-vowel transitions, but physiological constraints that are mirrored in acoustic parameters are (cf. Chapter 8 for further discussion). The acoustics of /ʃ/ within the stimulus word /vaʃə/ have already been discussed in Section 5.2. MZf1 was the twin pair with the most similar parameters (no significant differences in COG, PEAK, or the three DCT values and a very similar spectral shape of the mean long-term average spectra). Yet MZf1 was the twin pair that was most often distinguished correctly, thus parameters other than the acoustics of the sibilant seem to be more important in terms of perceived auditory similarity.

With respect to acoustic correlates that might be responsible for the differences in perceived similarity, it turned out that voice quality parameters like jitter, shimmer and HNR are crucial in perceived speaker similarity and very helpful in distinguishing otherwise very similar sounding voices like those of MZ pairs (cf. MZf1 and H3). With respect to the influence of physiology, H1b and H2 were supported, with the results revealing that mean fundamental frequency (F0) is more similar in MZ than in DZ twins while variation in fundamental

frequency differs equally in MZ and DZ twins. Regarding unrelated speakers, the assumption of a correlation between similarity in F0 and perceived speaker similarity (H1a) could be confirmed. In addition to the influence of F0 on perceived speaker similarity, shimmer also turned out to show a significant correlation.

7.4 Limitations and further research

Note that all results should be seen in the light of a perception experiment comparing stimuli that consist of only one word. Listeners had to concentrate on small salient indicators that triggered the decision of *same* or *different* speaker very spontaneously and directly, since they could listen to the short stimuli just once. It may be that in a longer stimulus the variation in F0 becomes more important and other acoustic cues might come to the fore. In addition, the stimuli were recorded for an articulatory and acoustic analysis, and not directly obtained with regard to a perception test. Therefore, they might not perfectly fit the purpose of investigating perceived speaker similarity and related acoustic correlates. Also, the speakers were only females and of the same age, thus no effect of gender or age could be investigated. Although several acoustic parameters were investigated, the examination could not address all the auditory information transferred by the signal to the listener. Further research is needed to investigate other acoustic parameters (such as durational aspects) and analyze their impact on perceived similarity with a larger (more twins) and more heterogeneous speaker sample. In addition, it would be informative to ask the listeners in a standardized way about the auditory cues they use to their discriminate the voices (such as melody, high/low voice, breathy sounding voice, etc.).

8 SUMMARY AND DISCUSSION

The work presented in this dissertation set out to evaluate articulatory, acoustic and perceptual similarities and differences among monozygotic and dizygotic twins. In this regard, this dissertation contributes to the discussion of the impact of NATURE versus NURTURE on the production and perception of speech. The aim of the present study was to evaluate parameters that differ in the amount of inter-speaker variability between MZ twins (who have identical genes and shared physiology) and DZ twins (who share only about 50% of their genes and differ in their physiology). It is assumed that both groups of twins shared their social environment and that their speech acquisition process was oriented towards the same auditory goals. Much effort was put into the inspection and verification of two crucial assumptions: 1) MZ and DZ twins differ in their amount of shared genes and physiology (NATURE), and 2) MZ and DZ twins do not differ in their amount of shared social environment (NURTURE).

The first assumption was supported by several analyses. For all participants of the study, pictures of the tongue were taken, silicone dental-palatal casts were made, body weight and size were measured, and a genetic test was carried out. Several metrics used to compare the different physiological and genetic features revealed nearly identical parameters for MZ twins but measurable differences for DZ twins. This was taken as evidence that tongue size and palate shape – crucial morphological properties for speech production – are genetically determined.

The examination of the second assumption was somewhat more complex. All twin pairs grew up and lived together for at least 18 years. In essence, this means that each pair was socialized together and each pair acquired and received language input from identical sources, went to the same schools, shared friends and were engaged in the same hobbies and activities during childhood and adolescence. In addition, all of the DZ twin pairs and the female MZ pairs were either still living together or saw each other every day at the time when the speech recordings were made. However, the two male MZ pairs differed in their amount of currently

shared environment. While at the time of recording the first twin pair lived in the same city and saw each other twice a month, the twins of the second pair had lived in different countries for 2 years; they saw each other only three to four times a year but kept in contact through e-mails and telephone calls at least once a month.

Hence, the MZ and DZ twins do not differ in the amount of shared social environment over the time-span from childhood to adolescence, but a pair-specific amount of currently shared environment shows up (with the male pairs revealing less time spent together). However, neither of the male MZ pairs stands out in terms of higher inter-speaker variability in the acoustics of vowels and sibilants than the other pairs. Therefore it is suggested that the early time of speech acquisition is more important than the current status of shared environment in influencing acoustic TARGETS, although this strongly depends on the duration of the separation and the peer group the twin shares time with. A potential influence of different environments over a period of years is not completely rejected and was found for example for a female twin pair that had been living apart for 45 years (Ryalls et al. 2004). These environmental effects come about through respective individualized social networks or different communities of practice in which these twins now engage. Language acquisition in the sense of adopting speech patterns of others within the same or across different social networks unfolds over a lifetime. Harrington was able to show that even the Queen's English had become less RP-like over the course of the previous 50 years (Harrington 2000, 2005, 2007). However, the evidence presented in this study points to a significant relevance of the length of time of the separation; a period of two years with periodical visits is not critical.

For the present analysis the two abovementioned assumptions led to the following statements:

- If a parameter differs more in DZ than in MZ twin pairs, the effect of NATURE must have a larger impact than the effect of NURTURE.

- If MZ and DZ twins show the same amount of inter-speaker variability, physiology (NATURE) does not play an important role while shared social environment, shared speech acquisition, and auditory goals (NURTURE) are important.

A second issue that arose during the cours of this study was that the potential influence of NATURE or NURTURE might not be equally well reflected in all speech parameters. NURTURE might be intensified in parameters which correspond to certain linguistic units, e.g. to prominent syllables, whereas NATURE might be more "visible" in dynamic aspects of speech production, where a large degree of inter-speaker variability is found.

Three factors turned out to be relevant for the influence of NATURE and/or NURTURE:

a) phoneme class (vowel vs. consonants),

b) lexical stress (stressed vs. unstressed),

c) degree of coarticulation (i.e. the nature of the analyzed item: target vs. transition).

The first factor involves the *phoneme class*. It may be surprising that vowels and consonants should differ with respect to the influence of NATURE and NURTURE. However, the literature is full of evidence that for vowels learned auditory goals (NURTURE) are crucial, whereas for consonants (here sibilants and stops) it is the tongue's movement within the surrounding vocal tract boundaries (NATURE) that is decisive. Moreover, these vocal tract boundaries, in particular the palate, can provide tactile feedback information and therefore contribute to the somatosensory goals of speech production (see Section 8.2.1).

For the second factor, *lexical stress*, it was supposed, in line with the literature, that syllables with lexical stress are more important regarding the communicative function of the speech signal than unstressed syllables. Syllables containing lexical stress are crucially dependent on learned auditory goals whereas in unstressed syllables the speaker's individual physiology may be more relevant (see Section 8.2.2).

The third factor deals with *coarticulation*. While it was hypothesized that TARGETS are learned entities that are influenced by and oriented towards shared auditory goals, the TRANSITIONS between the TARGETS are not controlled in the auditory domain. They are the consequence of the articulatory movements from one target to the next (coarticulation). In this sense, TRANSITIONS might mirror the individual properties of a speaker's vocal tract physiology and the biomechanical properties of the speaker's tongue muscles (see Section 8.2.3).

8.1 Summary of the results

The investigations of the present study cover several aspects of the speech production process with a focus on speaker-specific behavior and differences among related speakers. Articulatory, acoustic and perceptual analyses were carried out. In this section an overview of the results is given, subdivided into the following topics:

1) Vowel TARGETS, with the additional influence factors *lexical stress* and *consonant context* (a preceding velar stop or liquid) (Chapter 4)

2) Sibilant TARGETS, realizing sibilant phoneme CONTRASTS, and sibilant-vowel TRANSITIONS (Chapter 5)

3) Looping movements (GESTURES) of the tongue during /aka/ sequences (Chapter 6)

4) Perceived auditory similarity and acoustic correlates in a selected word (Chapter 7)

Regarding 1) The analysis of the vowels /a/, /i:/ and /u:/ in Chapter 4 revealed several interesting findings. First, zygosity only plays a minor role when it comes to the realization of stressed vowel TARGETS. This is true for articulation and acoustics. While tongue shapes display a greater similarity in MZ than in DZ twins, the investigated vowels did not differ in the same way in terms of target tongue positions and formant patterns (F1-F4). This strongly points to the role of auditory goals in speech production and the importance of NURTURE and shared social environment over NATURE. In fact, the formant analysis of stressed /i:/ revealed the possibility that MZ twins can differ even more than DZ twins in some cases. This suggests learned fine phonetic detail which is possible beyond what is predictable from the vocal tract physiology.

Nevertheless, several factors were found that contribute to and intensify the impact of identical physiology, i.e. the factors *lexical stress* and *consonant context* of the produced vowel. While /i:/ in a stressed position could not contribute to the distinction between MZ and DZ twins, /i/ in an unstressed position was found to be more similar in articulation and acoustics in MZ than in DZ twins. Similar tendencies were found for /i:/ following a velar stop vs. following a liquid. Here, too, biology and shared anatomy influence the similarity in acoustics

and articulation. This was interpreted in terms of a stronger degree of coarticulation in the /gi:/ sequence in contrast to the /li:/ sequence, and thus a stronger impact of NATURE on the coarticulatory patterns with the velar stop involved.

An additional finding was that the dimensions of the vowel space formed by F1 and F2 of /a/, /i:/ and /u:/ were more similar in MZ than in DZ twins. This might point to an influence of NATURE on the overall shape of the vowel space formed by the corner vowels, although the acoustics of the particular vowels does not seem to be influenced by physiological restrictions.

Regarding 2) In Chapter 5 an investigation of acoustic and articulatory inter-speaker variability in sibilants was presented. The analysis of the sibilant /s/ leads in the same direction as that of the vowels: no effect of zygosity was found regarding articulatory positions or acoustic parameters in phoneme TARGETS. However, a small influence of shared physiology on the production of /ʃ/ could be observed, since the MZ twins were more similar in their articulation and acoustics than the DZ twins. Further interesting results were revealed by the analysis of the articulatory realization of the phoneme CONTRAST: while no strong effect of zygosity on the acoustic difference between /s/ and /ʃ/ could be found, MZ and DZ twins could be differentiated by their degree of inter-speaker variability in terms of the articulatory distance between the sibilants (i.e. the relative amount of horizontal/vertical distance between the target positions).

In research on inter-speaker variability the focus has recently changed from investigating phonemic targets to investigating phonemic contrasts. Perkell, Matthies et al. (2004) and Ghosh et al. (2010) emphasize examining phonemic contrasts instead of phonemic targets and explain speaker-specific behavior in line with the relation between speech production and perception. More specifically, they found that speakers with a lower perceptual acuity of a phonemic contrast also tend to produce this contrast less distinctively in comparison to speakers with a higher perceptual acuity. Ghosh et al. (2010) extended this work and measured speakers' somatosensory acuity in addition to perceptual acuity and acoustic distance for the /s/-/ʃ/ contrast. They reported a positive correlation between speakers' acoustic realization of the contrast and their respective auditory and somatosensory acuity. From the results of the current study it is concluded that in addition to speaker-specific

auditory and somatosensory acuity, individual morphological differences in palatal shape influence the articulatory realization of the /s/-/ʃ/ contrast in German.

The results of this study are in line with Toda (2006), who reports two different strategies for the realization of the contrast in French: the *tongue placement strategy* and the *tongue adjustment strategy*. In the former case, subjects only retract the tongue horizontally without an elevation whereas in the latter case the tongue is elevated and follows the palate contour. Both of these speaker-specific strategies could be related to individual palate shape. For speakers with low palates, a pure retraction of the tongue tip from /s/ to /ʃ/ may already result in an appropriate production, since the tongue touches the palate at the lateral margins. Moreover, speakers with a flat palate may have a high articulatory acuity since they are very constrained in their articulatory variability (Brunner et al. 2009). In contrast, speakers with a relatively high and domed palate also need to elevate their tongue in order to produce tongue grooving and the relevant constriction for /ʃ/. Moreover, one could even suppose that depending on the palate shape, a certain degree of somatosensory acuity is learned (high acuity for speakers with a flat palate and low acuity for speakers with a domed palate), but this hypothesis remains to be tested.

Furthermore, a parameter that could distinguish MZ from DZ twins was also found in the acoustic domain: namely sibilant-vowel TRANSITIONS. The slopes of the transitions were very similar in MZ twins, while they revealed significant differences in DZ twins. Hence, individual biomechanical properties (NATURE) influence the trajectories between two targets. This is supported by Kühnert & Nolan (1999) and Rose (2002), who see coarticulatory behavior as essentially idiosyncratic. Moreover, the current results additionally emphasize the influence of individual morphology on coarticulatory patterns. This conforms to the results of Nolan & Oh (1996), who found individual differences in coarticulatory behavior in unrelated speakers but not in MZ twins.

Regarding 3) One of the strongest parameters distinguishing MZ from DZ twins was the looping trajectories of the tongue back during the sequence /aka/ (discussed in Chapter 6). The articulatory investigation revealed very similar positions, shapes and directions of the loops in all MZ twins, while the DZ twins revealed striking differences in the horizontal and vertical positions of the trajectory, the shape of the loop, the direction of the upward

movement and the amount of horizontal sliding movement at the palate. The statistical analysis of the Euclidean distances (EDs) between the aligned trajectories could provide further support for the assumed influence of physiology and NATURE on articulatory gestures involving a velar stop, since DZ twins were as similar as unrelated speakers in the measured articulatory parameters, whereas MZ twins differed significantly from DZ twins.

In research there has been much discussion about an explanation of the observed loops (Houde 1967, Mooshammer et al. 1995, Hoole et al. 1998, Löfqvist & Gracco 2002, Perrier et al. 2003). While Löfqvist & Gracco (2002) explain the shape of the trajectory by a general optimization principle that plans the entire trajectory (and involves a complex internal model), Perrier et al. (2003) suggest that the curvature of the trajectory is due to biomechanical properties of the tongue by using a biomechanical tongue model (the passive tongue elasticity, the muscle arrangements within the tongue, the force generation mechanism). Due to the particular subject group in the current investigation this study can make a significant contribution to the discussion: since the MZ twins were found to be more similar than the DZ twins and the DZ twins did not differ from the unrelated speaker pairs, the impact of NATURE and biomechanics is assumed to be crucial.

Regarding 4) In Chapter 7 the perceived auditory similarity of MZ twins, DZ twins and unrelated speakers was inspected and several possible acoustic correlates were investigated. It turned out that unrelated speakers are easier to distinguish than MZ and DZ twins. However, there is no difference in auditory similarity between the two twin types. Of the four investigated female twin pairs (2 DZ and 2 MZ pairs) the most similar sounding pair and the most often correctly distinguished pair were both monozygotic. Voice quality parameters were found to be helpful in distinguishing similar sounding voices, while the differences in sibilant-vowel TRANSITIONS reported above did not play a role. Differences in auditory similarity between the unrelated speaker pairs revealed a crucial role of F0.

The evidence presented in this perception experiment strongly points to a nurture-based approach to language acquisition and supports earlier findings of similar inter-speaker variability in acoustic TARGETS of MZ and DZ twins. However, the fact that the speakers of one MZ pair (who not only shared the same social environment during their upbringing

but also share an identical physiology) were distinguished above chance provides evidence of variation that reflects current environmental effects.

The following table gives a summary of the parameters that were more similar in MZ twins than in DZ twins. Here, an influence of NATURE on inter-speaker variability is assumed.

Table 38: Acoustic and articulatory parameters that differ in the amount of inter-speaker variability between MZ and DZ twins.

Phoneme		Parameter
Vowels	Formants and tongue positions of unstressed /i/ and /i:/following a velar consonant	Target
	F1/F2 dimensions of vowel spaces	Vowel space
Sibilants	Acoustic parameters and articulatory tongue position of /ʃ/	Target
	Articulatory realization of phoneme contrast /s/-/ʃ/	Phoneme contrast
	Vowel-sibilant transitions	Transition
/aka/	Looping trajectories	Gesture

Thus, regarding the research question as to which of the factors NATURE and NURTURE is the most important factor in inter-speaker variability, the conclusion is that this is strongly dependent on the parameters investigated in the speech signal. Both NATURE and NURTURE contribute to the amount of speaker-specific variability found in twins' speech: while in some cases a distinction in the amount of inter-speaker variability could be found between MZ and DZ twins (cf. Table 38), in other cases it could not.

However, several factors were identified as either strengthening the influence of NATURE or supporting the significance of auditory goals affected by NURTURE. As mentioned above these factors are a) the phoneme category, b) the lexical stress condition and c) whether the analyzed parameter is dynamic (TRANSITION or GESTURE) or static (TARGET). The further discussion will concentrate on these issues.

8.2 Enhancing the influence of NATURE on inter-speaker variability

8.2.1 The role of the phoneme category: Vowels vs. consonants

One research question that was asked within this investigation was the following:

> Is there a difference in the influence of NATURE on inter-speaker variability depending on the **phoneme category**?

It seems that indeed there is a difference in the relative impact of physiology and NATURE depending on the phoneme category: consonants and in particular sibilants (here especially /ʃ/) and stops (i.e. /k/ within the sequence /aka/) turned out to be more affected by biological constraints than vowels.

Several approaches may help to understand the difference between vowels and consonants regarding the impact of NATURE.

First, vowels are primarily based on auditory goals and shaped by the learned phoneme inventory. This is in line with several perturbation studies (Perkell et al. 2007, Brunner 2009, Cai et al. 2010) and the results of the current study. Moreover, vowels show only a limited linguo-palatal contact and are therefore less constrained by individual vocal tract boundaries, apart from high vowels. In contrast, most consonants, and in particular stops and fricatives, show a larger degree of tongue movement at the palate (Stone 1995, Honda et al. 2002, Fuchs et al. 2006). Stone even proposed that the palate not only constrains the articulatory movement, but also allows certain tongue shapes (as for sibilants) that would not be possible otherwise. In this sense, the individual vocal tract boundary (e.g. the shape of the palate) is crucial for the production of these sounds, leading to the conclusion that the goals of consonantal production are primarily somatosensory. Again, this is in line with various perturbation studies (Honda et al. 2002, Honda & Murano 2003, Brunner 2009).

Thus, a potential explanation for the differences in the relative importance of NATURE depending on the phoneme class might be an articulatory one, i.e. the difference in producing these sounds. For /ʃ/ the tongue has to be situated at the anterior palatal region; moreover, the tongue has to be slightly pushed against the palate and a lateral linguo-palatal contact exists. The anterior palatal constriction has to be quite narrow and the tongue has to follow the form of the palate. For /k/ normally a full closure between tongue and palate has to be formed, thus the linguo-palatal contact is crucial. In terms of the virtual target hypothesis (see Chapter 6), the movement of the tongue is stopped by the palate, since the target is seen 'above' the palate. Hence, the shape of the palate is much more important for the production of stops as well as for sibilants as compared to vowels (except for perhaps high front vowels, cf. Fuchs et al. 2006). The role of somatosensory feedback has been discussed for example in Honda et al. (2002), Honda & Murano (2003), Brunner (2009) and Ghosh et al. (2010). These studies have shown that somatosensory feedback plays a major role in sibilants and only a minor one in vowels. The results of the current study provide further support for the assumption that tactile or somatosensory feedback is most important in sounds with linguo-palatal contact. Moreover it is suggested that the somatosensory feedback is influenced by the individual vocal tract morphology. Consequently, physiological restrictions are more salient in sounds with more linguo-palatal contact and the influence of NATURE is stronger.

One might ask whether for the planning of a CV syllable it is necessary to switch from the somatosensory to the auditory control space, if consonants and vowels are defined in a different way. However, this may not be necessary since during speech acquisition a mapping between the auditory and the somatosensory control space is learned (Guenther 1995, Guenther et al. 1998, Guenther et al. 2006) and when learning is finished both control spaces can be used. Moreover, they may even provide, to some extent, redundant information. The one or the other control space may only be preferred when sudden perturbations are applied to the motor system.

Second, the differences found between vowels and consonants may be reflected in the special status that vowels are given in some models. Öhman (1966), for instance, suggests that VCV sequences consist of a basic vowel cycle, while consonants are considered as perturbations that are superimposed on the continuous production of vowels. Similarly, Fowler (1983)

states that at least in stressed syllables vowels are produced continuously while consonants are coordinated with them.

Furthermore, a general phenomenon is that there is a difference in the articulatory pattern (i.e. the velocity of the gesture) between vowel gestures and consonant gestures: opening gestures towards a vowel are slower in the articulatory movement than closing gestures towards a consonant (Gracco 1988). This is in line with the finding of the current study that transitions and gestures, i.e. more dynamic parameters of the speech signal, are more affected by vocal tract constraints than more static parameters (targets). This will be discussed in more detail in Section 8.2.3.

As mentioned above, the phoneme class is not the only influencing factor found to play a role in the relative importance of NATURE in inter-speaker variability. In addition, an interaction of phoneme class and stress is hypothesized. Unstressed syllables are assumed to show a stronger effect of the impact of NATURE than stressed syllables, as is discussed in the following section.

8.2.2 The role of lexical stress: Stressed vs. unstressed syllables

A second matter that arose during this investigation is related to the impact of communicative demands in speech. The following question was asked:

> Is there a difference in the influence of NATURE on inter-speaker variability depending on **stress** (vowel in a stressed syllable vs. unstressed syllable)?

In the analysis of the vowels it was found that the factor *lexical stress* could intensify the role of shared physiology. In particular, unstressed /i/ was more likely to differ in DZ twins than in MZ twins, while no difference between the twin types was found in the stressed condition.

As mentioned above, this should be seen in the light of communicative demands. In a free-stress language like German, stress is a useful cue in word recognition (for an overview, see Cutler 2005). Stressed syllables are more important for the delivery of the information carried in the signal. Statistical studies on lexicons have investigated the informative value of a portion of a word regarding its discrimination from other words in this lexicon. Here, the

important role of lexical stress has been shown: stressed syllables were found to be more informative than unstressed syllables (cf. Huttenlocher 1984, Altman & Carter 1989).

Furthermore, several studies on spontaneous speech have revealed that lexical stress can affect speech perception. For example, stressed syllables are recognized earlier than unstressed syllables (McAllistor 1991) and word-initial target phonemes are identified faster in stressed than in unstressed syllables (Mehta & Cutler 1988). Van Bergem (1993) assumes that stressed syllables may serve as anchor points in word recognition. Hence, prosodic information is used in the process of speech perception. Moreover, studies in speech pathology give further evidence that lexical stress can affect speech perception. In particular, speech distortions are more likely to be detected in stressed syllables than in unstressed syllables (Bond & Garnes 1980, Cole & Jakimik 1980).

All of these studies have in common that unstressed syllables are seen to be less salient in speech perception, and less informative for the carried linguistic information. In addition, it is known that a relation between stressed and unstressed syllables and hyper- and hypoarticulation exists (de Jong et al. 1993, de Jong 1995, 1998). Both duration and spectral quality change in unstressed syllables (van Bergem 1993). Van Bergem concludes that "vowels from stressed syllables are generally more clearly pronounced and hence closer to their target form than the vowels from unstressed syllables" (van Bergem 1993, p. 21). These phenomena are known within the concept of 'target undershoot' (see *H&H theory*, Lindblom 1983, 1990: output-oriented and hyperarticulated stressed syllables vs. system-oriented and reduced/hypoarticulated unstressed syllables). Mooshammer & Geng (2008) investigated not only acoustic but also articulatory manifestations of vowel reductions in German. Results regarding articulatory dimensions and formant patterns point to a greater degree of coarticulation with the consonant context in unstressed vowels than in stressed vowels. Consequently, it may be hypothesized that intra- but also inter-speaker variability in general should be greater in unstressed than in stressed syllables.

However, since a difference between MZ and DZ twins in the degree of inter-speaker variability depending on the stress condition was found in the current study (i.e. MZ and DZ twin pairs differ in the amount of inter-speaker variability in unstressed but not in stressed syllables), the abovementioned explanation does not seem to be sufficient. It is rather

suggested that the stressed syllables are oriented towards auditory goals, and the unstressed syllables are influenced to a greater degree by the individual's physiology and less by the auditory target; thus unstressed syllables differ least in physiologically quasi-identical speakers like MZ twins.

Hence, an interesting follow-up study would be to look into the influence of the factor *stress* on inter-speaker variability more deeply and more straightforwardly than the present investigation could do. Moreover, not only *lexical stress* but *focus* could be an attractive issue to look at since it has been found that words under focus are pitch accented and hyperarticulated as compared to words in a non-focused position, which are shorter in duration and reduced (see van Bergem 1993 for Dutch, or Hermes et al. 2008 for German).

A potential experiment with target words in non-focused or even post-focus positions (in contrast to words in a focused position) could examine the possible intensifying impact of physiology on speech. The following sentences serve as an example of two carrier sentences with the target word <*kocht*> (3rd p. sg. 'to cook') in a post-focus (a) and focus position (b):

a) Kocht MANU heute? Nein, MONA **kocht** heute.

('Is MANU cooking today? No, MONA is cooking today')

b) FEIERT Mona heute? Nein, Mona **KOCHT** heute.

('Is Mona CELEBRATING today? No, Mona is COOKING today')

Here, the predictions would be that NATURE has a stronger influence on the target word (e.g. on the formants of the vowel /ɔ/ or the production of the velar stop /k/) in the post-focus position (a) than in the focus position (b). In regard to the subject group of twins, it is hypothesized that MZ and DZ twins do not differ in the amount of inter-speaker variability in (b) but they do in (a). Here DZ twins are assumed to show more differences than MZ twins in the particular investigated parameter.

8.2.3 The role of coarticulation: Static (TARGET) vs. dynamic (TRANSITION) patterns

A third issue discussed in the present study concerned the separation of the speech signal into TARGETS and TRANSITIONS. The following research question was asked at the beginning of the thesis:

> Is there a difference in the influence of NATURE on inter-speaker variability depending on the particular characteristic of the analyzed parameter: **target (static) vs. transition (dynamic)**?

Indeed, during the analyses of the vowels, sibilants and VkV gestures it turned out that there is a difference. When it comes to the question as of distinguishing MZ and DZ twins regarding inter-speaker variability, acoustic TRANSITIONS and articulatory GESTURES become more important than TARGETS. The analyses conducted of vowels, sibilants and VkV gestures revealed that no strong effect of zygosity and thus biology and NATURE exists when articulatory and auditory TARGETS are concerned. However, remarkable results in terms of a distinction between MZ and DZ twins in inter-speaker variability could be found in a) articulatory GESTURES (VkV) and b) acoustic TRANSITIONS (/sə/, /ʃə/).

Hence, physiological constraints that are reflected in measurable acoustic parameters influenced the sibilant-vowel transitions. Furthermore, the differences in the transitions did not affect the perceived auditory similarity of the twins: the perception test revealed that the DZ pairs were not easier to distinguish than the MZ pairs (even though the differing transitions were part of the stimuli). Taken together this means that the differing physiological similarity (expressed in measurable acoustic parameters) did not affect the auditory similarity. This is supported by studies using automatic speaker identification systems, where it was found that computer systems were more successful in distinguishing twins than listeners (Rosenberg 1973, Künzel 2010). Acoustic signal parameters might be detected by the system even though they are not audible to the listener.

Nevertheless, it has to be noted that transitions can indeed carry linguistic information and consequently are of course auditorily salient. This has been found, for example, for Mandarin, where speakers reacted to perturbed formant transitions in triphthongs (Cai et al. 2010). That dynamic spectral information is crucial in speech perception has been shown by Strange et al. 1983 in their experiments regarding the "silent center approach." Here, initial and final transitions in /b/vowel/b/ sequences were shown to be sufficient for the identification of the (deleted) vowel. Thus, transitions are important in speech perception – however, the abovementioned research has been done mainly on the identification of vowels.

However, the results of the current analysis suggest that the investigated transitions are not crucial in terms of perceived speaker identity and the auditory cues listeners use to distinguish speakers.

Both parameters (the /aka/ gesture and the sibilant-vowel transition) have something in common: both are dynamic and not static like tongue positions or formant patterns measured in the (steady-state) middle part of a vowel. This distinction could be a crucial one where the potential impact of NATURE and physiology on inter-speaker variability is concerned. Thus, MZ twins are assumed to show fewer differences than DZ twins in dynamic speech patterns (like TRANSITIONS and GESTURES) but not necessarily in static ones (like TARGETS). These findings are in line with the suggestions from Nolan et al. (2006), Kühnert & Nolan (1999) and Rose (2002), who propose that TARGETS are linguistically determined and influenced by the learned and shared language system, while TRANSITIONS and coarticulatory strategies are organically determined and idiosyncratic.

However, a difference in inter-speaker variability between MZ and DZ twins was also found in a few static speech parameters, but especially in the case where the relative distance between two targets was considered. In particular, it was found that zygosity and hence physiological similarity play a role in the articulatory strategy to realize the phoneme contrast between /s/ and /ʃ/ in German. The amount of distance and especially the relation of the vertical to the horizontal distance between the two target tongue positions were more similar in MZ than in DZ twins. Thus, the precise realization of the phoneme contrast is influenced by the individual palatal shape and hence NATURE.

Furthermore, the results regarding the difference between TRANSITIONS and TARGETS can contribute to issues related to speech motor control. Here, an open question is whether the motor goals specify only the final target position or whether they constrain the entire movement trajectory. In the first case, the trajectory between the targets emerges from neuromuscular processes. In the latter case, not only the target but also the trajectory is controlled and results from complex neural processes that are shaped by motor goals (for an overview, see Grimme et al. 2011).

Thus the question arises as to whether targets or trajectories are the relevant motor goals in speech production (and perception). Several studies have been conducted favoring the one or the other possibility. In acoustic theories (see for instance Fant 1960 or Stevens 1972), it has been suggested that vowels are characterized by steady state spectral characteristics. In articulatory theories, the place and manner of articulation are seen as characterizing the vowel or consonant target (Browman & Goldstein 1986, Guenther 1995). In contrast, other studies have revealed the significance of trajectories/transitions in speech perception (see Strange et al. 1983 for vowels in American English, or Cai et al. 2010 for triphthongs in Mandarin Chinese). In general, it is assumed that in speech perception both steady state characteristics and transitions provide information that is crucial for the identification of sounds. However, the role of the control mechanism underlying speech production in terms of the distinction between targets and transitions is less clear. Here, the present investigation may give some impulses since targets and transitions differed in the influence of NATURE (physiology). Thus, from the results of the current twin study it is suggested that TARGETS are controlled but TRANSITIONS are not, since TARGETS have been found to be less influenced by physiology than TRANSITIONS.

To conclude, this investigation has examined the role of NATURE and NURTURE in inter-speaker variability in the acoustics, articulation and perception of speech. It is limited in its explanatory power by the restricted number of subjects and the particular speech material that was chosen. Nevertheless, it could shed some light on the NATURE-NURTURE debate by analyzing the speech of MZ and DZ twins, who differ in their amount of similar physiology. Results point to the overall importance of NURTURE, shared social environment and the crucial role of auditory goals in speech production. Nevertheless, NATURE revealed its significance in several aspects: Somatosensory feedback plays a bigger role in consonants than

in vowels, and thus individual physiology was found to shape articulation more in sibilants and stops than in vowels. Moreover, the articulatory realization of the phoneme contrast /s/-/ʃ/ turned out to be essentially dependent on physiology, namely palatal shape. In addition, the articulatory gesture of the tongue back during /aka/, which forms a loop, is strongly affected by biological restrictions. In acoustics, sibilant-schwa transitions (which are not auditorily salient as shown in the perception test) were found to be more similar in identical twins; thus, here NATURE showed its influence here. On the other hand, NURTURE turned out to be more important in auditorily salient parameters that carry most of the communicative information transimitted by the speech signal: stressed syllables and vowels did not differ in terms of inter-speaker variability between MZ and DZ twins.

Thus, it is concluded that the influence of NURTURE is strong in *static* speech TARGETS that go hand in hand with auditory goals and communicative demands (like conveying lexical information), while the impact of NATURE is greatest in auditorily less salient and linguistically less relevant patterns of speech like unstressed syllables and *dynamic* GESTURES and TRANSITIONS.

REFERENCES

Aguilar, L., Downey, G., Krauss, R., Pardo, J. & Bolger. (under review) Too Much Too Soon: Rejection sensitivity and speech accommodation in dyadic interaction. Journal of Personality & Social Psychology.

Alfonso, P. & Baer, T. (1982) Dynamics of vowel articulation. Language and Speech 25, 151-173.

Altman, G. & Carter, D. M. (1989) Lexical stress and lexical discriminability: Stressed syllables are more informative, but why? Computer Speech and Language 3, 265-275.

Baayen, R. (2008) Analyzing Linguistic Data: A Practical Introduction to Statistics Using R. Cambridge: Cambridge University Press.

Baken, R. & Orlikoff, R. (2000) Clinical Measurement of Speech and Voice. San Diego, CA: Singular Pub Group.

Bandura, A. (1977) Social Learning Theory. Englewood Cliffs, N.J.: Prentice-Hall.

Bauer, H. R. & Kent, R. D. (1987) Acoustic analyses of infant fricative and trill vocalizations. Journal of the Acoustical Society of America 81, 505-511.

Beck, J. M. (1999) Organic variation of the vocal apparatus. *in* W. J. Hardcastle; J. Laver & F. E. Gibbon (eds.), The Handbook of Phonetic Sciences, Oxford and Cambridge: Blackwell Publishers, 256-297.

Bond, Z. S. & Garnes, S. (1980) Misperceptions of fluent speech. *in* R. Cole (ed.), Perception and Production of Fluent Speech, Hillsdale, NJ: Lawrence Erlbaum, 115-132.

Boersma, P. & Weenink, D. (2009) Praat: Doing phonetics by computer (Version 5.1) [Computer program]. Retrieved from http://www.praat.org/.

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

Bowen, C. (2002) Beyond lisping: Code switching and gay speech styles. Retrieved from http://www.speech-language-therapy.com/codemix.htm.

Bradlow, A. R.; Pisoni, D. B.; Akahane-Yamada, R. & Tohkura, Y. (1997) Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. Journal of the Acoustical Society of America 101, 2299-2310.

Browman, C. P. & Goldstein, L. (1986) Towards an articulatory phonology. Phonology Yearbook 3, 219-252.

Browman, C. P. & Goldstein, L. (1989) Articulatory gestures as phonological units. Phonology 6, 201-251.

Browman, C. P. & Goldstein, L. (1990) Representation and reality: Physical systems and phonological structure. Journal of Phonetics 18, 411-424.

Browman, C. P. & Goldstein, L. (1992) Articulatory phonology: An overview. Phonetica 49, 155-180.

Brunner, J. (2009) Perturbed Speech. How Compensation Mechanisms Can Inform Us about Phonemic Targets. Saarbrücken: Südwestdeutscher Verlag für Hochschulschriften.

Brunner, J.; Fuchs, S. & Perrier, P. (2005) The influence of the palate shape on articulatory token-to-token variability. ZAS Papers in Linguistics 42, 43-67.

Brunner, J.; Fuchs, S. & Perrier, P. (2009) On the relation of palate shape and articulatory behavior. Journal of the Acoustical Society of America 125(*6*), 3936–3949.

Brunner, J., Fuchs, S. & Perrier, P. (2011) Supralaryngeal control in Korean velar stops. Journal of Phonetics 39, 178-195.

Buchaillard, S., Perrier, P., & Payan, Y. (2008) Muscle saturation effect in /i/ production: Counterevidence from a 3D biomechanical model of the tongue. Journal of the Acoustical Society of America 123, 3321.

Bybee, J. (2001) Phonology and Language Use. Cambridge: Cambridge University Press.

Cai, S.; Ghosh, S.; Guenther, F. & Perkell, J. S. (2010) Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization. Journal of the Acoustical Society of America 128(*4*), 2033-2048.

Chambers, J. (2003) Sociolinguistic Theory. Oxford: Blackwell.

Chartrand, T. L., & Bargh, J. A. (1999) The chameleon effect: The perception-behavior link and social interaction. Journal of Personality and Social Psychology 76(*6*), 893-910.

Chomsky, N. (1975) Reflections on Language. New York: Pantheon Press

Coker, C. (1976) A model of articulatory dynamics and control. Proceedings of the IEEE 64, 452-460.

Cole, R. A. & Jakimik, J. (1980) How are syllables used to recognize words? Journal of the Acoustical Society of America 67, 965-970.

Cook, V. (ed.) (2003) Effects of the Second Language on the First. Clevedon: Multilingual Matters.

Cornut, G. (1971) Génèse de la voix de l'enfant. J Fr Otorhinolaryngol-Audiophonol-Chir-Maxillofac 20(*2*), 411-416.

Cutler, A. (2005) Lexical stress. *in* D. B. Pisoni, & R. E. Remez (eds), The Handbook of Speech Perception. Oxford: Blackwell Publishing, 264-289.

Daniloff, R.; Wilcox, K. & Stephens, M. (1980) An acoustic-articulatory description of children's defective /s/ productions. Journal of Communication Disorders 13, 347-363.

Dart, S. N. (1998) Comparing French and English coronal consonant articulation. Journal of Phonetics 26, 71-94.

Davidson, L. (2006) Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. Journal of the Acoustical Society of America 120(*1*), 407-415.

Davis, S. (1979) Acoustic characteristics of normal and pathological voices. *in* N. J. Lass (ed.), Speech and Language: Basic Advances in Research and Practice, New York: Academic Press, 271-335.

Deal, R. & Emanuel, F. (1978) Some waveform and spectral features of vowel roughness. Journal of Speech and Hearing Research 21(*2*), 250-264.

Debruyne, F.; Decoster, W.; Van Gijsel, A. & Vercammen, J. (2002) Speaking fundamental frequency in monozygotic and dizygotic twins. Journal of Voice 16(*4*), 466-471.

Decoster, W. & Debruyne, F. (2000) Longitudinal voice changes: Facts and interpretation. Journal of Voice 14, 184-193.

Decoster, W.; Van Gijsel, A.; Vercammen, J. & Debruyne, F. (2001) Voice similarity in identical twins. Acta Otorhinolaryngol Belg. 50(*1*), 49-55.

de Jong, K. J. (1995) The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. Journal of the Acoustical Society of America 97, 491-504.

de Jong, K. J. (1998) Stress-related variation in the articulation of coda alveolar stops: flapping revisited. Journal of Phonetics 26, 283-310.

de Jong, K. J., Beckman, M.E. & Edwards, L. (1993) The interplay between prosodic structure and coarticulation. Language and Speech 36(*2-3*), 197-212.

Diehl, R. L. & Kingston, J. (1991) Phonetic covariation as auditory enhancement: The case of the [+voice]/[-voice] distinction. i*n* O. Engstrand & C. Kylander (eds.), Current Phonetic Research Paradigms: Implications for Speech Motor Control, PERILU (volume 14), Stockholm: University of Stockholm, 139-143.

Diehl, R. L. & Kluender, K. R. (1989) On the objects of speech perception. Ecological Psychology 1, 121-144.

di Pellegrino, G.; Fadiga, L.; Fogassi, L.; Gallese, V. & Rizzolatti, G. (1992) Understanding motor events: A neurophysiological study. Experimental Brain Research 91, 176-180.

Dukiewicz, L. (1970) Frequency band dependence of speaker identification. *in* W. Jassem (ed.), Speech Analysis and Synthesis (volume II), Warsaw: Polish Academy of Sciences, 41-50.

Evers, V.; Reeth, H. & Lahiri. A. (1998) Crosslinguistic acoustic categorization of sibilants independent of phonological status. Journal of Phonetics 26, 345-370.

Fant, G. (1960) Acoustic Theory of Speech Production. The Hague: Mouton.

Farnetani, E. (1999) Coarticulation and connected speech processes. *in* W. J. Hardcastle; J. Laver & F. E. Gibbon (eds.), The Handbook of Phonetic Sciences, Oxford and Cambridge: Blackwell Publishers, 371-404.

Feiser, H. (2009) Evaluierung von gemeinsamen und unterschiedlichen Sprech- und Stimmmerkmalen von gleichgeschlechtlichen Geschwisterpaaren. Frankfurt/Main: Verlag für Polizeiwissenschaft.

Fitch, W. & Giedd, J. (1999) Morphology and development of the human vocal tract: A study using magnetic resonance imaging. Journal of the Acoustical Society of America 106(*3*), 1511-1522.

Flach, M.; Schwickardi, H. & Steinert, R. (1968) Zur Frage des Einflusses erblicher Faktoren auf den Stimmklang (Zwillingsuntersuchungen). Folia Phoniatrica et Logopaedica 20(*6*), 369-378.

Flege, J. E. (1995) Second language speech learning: Theory, findings and problems. *in* W. Strange (ed.), Speech Perception and Linguistic Experience: Issues in Cross-Language Research, Timonium, MD: York Press, 233-272.

Flipsen, P.; Shriberg, L.; Weismer, G.; Karlsson, H. & McSweeney, J. (1999) Acoustic characteristics of /s/ in adolescents. Journal of Speech, Language and Hearing Research 42(*3*), 663-677.

Forrest, K.; Weismer, G.; Milenkovic, P. & Dougall, R. N. (1988) Statistical analysis of word-initial voiceless obstruents: Preliminary data. Journal of the Acoustical Society of America 84(*1*), 115-123.

Fowler, C. A. (1980) Coarticulation and theories of extrinsic timing. Journal of Phonetics 8, 113-133.

Fowler. C. A. (1983) Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. Journal of Experimental Psychology: General 112(*3*), 386-412.

Fowler, C. A. (1986) An event approach of the study of speech perception from a direct-realist percpective. Journal of Phonetics 14, 3-28.

Fowler, C. A. (1991) Auditory perception is not special: We see the world, we feel the world, we hear the world. Journal of the Acoustical Society of America 89, 2910-2915.

Fuchs, M.; Oeken, J. H. T.; Täscher, R.; Hentschel, B. & Behrendt, W. (2000) Die Ähnlichkeit monozygotischer Zwillinge hinsichtlich Stimmleistungen und akustischer Merkmale und ihre mögliche klinische Bedeutung. HNO, 462-469.

Fuchs, S.; Perrier, P.; Geng, C. & Mooshammer, C. (2006) What role does the palate play in speech motor control? Insights from tongue kinematics for German alveolar obstruents. *in* J. Harrington & M. Tabain (eds.), Speech Production: Models, Phonetic Processes, and Techniques, New York: Psychology Press, 149-164.

Fuchs, S.; Perrier, P. & Mooshammer, C. (2001) The role of the palate in tongue kinematics: An experimental assessment in VC sequences from EPG and EMMA data. Proceedings of Eurospeech, Aalborg, Denmark. 1487-1490.

Fuchs, S.; Pompino-Marschall, B. & Perrier, P. (2007) Is there a biological grounding of phonology? Determining factors, optimization, and communicative usage. Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, 219-224.

Fuchs, S. & Toda, M. (2008) Inter-speaker variability and the articulatory-acoustic relations in German and English /ʃ/. Journal of the Acoustical Society of America 123, 3079.

Fuchs, S.; Winkler, R. & Perrier, P. (2008) Do speakers vocal tract geometries shape their articulatory behavior? Proceedings of the 8th International Seminar on Speech Production, Strasbourg, 333-336.

Galton, F. (1876) The history of twins as a criterion of the relative powers of nature and nurture. Royal Anthropological Institute of Great Britain and Ireland Journal 6, 391-406.

Gedda, L.; Fiori-Ratti, L. & Bruno, G. (1960) La voix chez les jumeaux monozygotiques. Folia Phoniatrica 12, 81-94.

Geng, C.; Fuchs, S.; Mooshammer, C. & Pompino-Marschall, B. (2003) How does vowel context influence loops? Proceedings of the 6th International Seminar on Speech Production, Sydney, 67-72.

Ghosh, S.; Matthies, M.; Maas, E.; Hanson, A.; Tiede, M.; Ménard, L.; Guenther, F.; Lane, H. & Perkell, J. S. (2010) An investigation of the relation between sibilant production and somatosensory and auditory acuity. Journal of the Acoustical Society of America 125(5), 3079-3087.

Gordon, M.; Barthmaier, P. & Sands, K. (2002) A cross-linguistic acoustic study of voiceless fricatives. Journal of the International Phonetic Association 32, 141-174.

Guenther, F. H. (1995) Speech sound acquisition, coarticulation and rate effects in a neural network model of speech production. Psychological Review 102(*3*), 594-621.

Guenther, F. H.; Ghosh, S. S. & Tourville, J. A. (2006) Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang 96, 280-301.

Guenther, F. H.; Hampson, M. & Johnson, D. (1998) A theoretical investigation of reference frames for the planning of speech movements. Psychological Review 105, 611-633.

Guzik, K. & Harrington, J. (2007) The quantification of place of articulation assimilation in electropalatographic data using the similarity index (SI). Advances in Speech Language Pathology 9(*1*), 109-119.

Gracco, V. L. (1988) Timing factors in the coordination of speech movements. Journal of Neuroscience 8(*12*), 4628-4639.

Grimme, B., Fuchs, S., Perrier, P. & Schöner, G. (2011) Limb versus Speech Motor Control: A Conceptual Review. Motor Control 15, 5-33.

Hall, T. A. (1992) Syllable Structure and Syllable Related Processes in German. Tübingen: Niemeyer.

Hardcastle, W. & Hewlett, N. (eds.) (1999) Coarticulation: Theory, Data, and Techniques. Cambridge: Cambridge University Press.

Harrington, J. (2000) Monophthongal vowel changes in received pronunciation: An acoustic analysis of the Queen's Christmas broadcasts. Journal of the International Phonetic Association 30, 63-78.

Harrington, J. (2005) An acoustic analysis of happy-tensing in the Queen's Christmas broadcasts. Journal of Phonetic*s* 34, 439-457.

Harrington, J. (2007) Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. *in* J. Cole & J. Hualde (eds.), Laboratory Phonology 9, Berlin: Mouton, 125-143.

Harrington, J.; Palethorpe, S. & Watson, C. I. (2000) Does the Queen speak the Queen's English? Nature 408, 927-928.

Harshman, R.; Ladefoged, P. & Goldstein, L. (1977) Factor analysis of tongue shapes. Journal of the Acoustical Society of America 62, 693-707.

Hawkins, S. (1999) Reevaluating assumptions about speech perception: Interactive and integrative theories. *in* J. M. Pickett (ed.), The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology, Boston: Allyn and Bacon, 232-288.

Helfrich, H. (1979) Age markers in speech. *in* K. Scherer & H. Giles (eds.), Social Markers in Speech, Cambridge: Cambridge University Press, 63-107.

Hermes, A.; Becker, J.; Mücke, D.; Baumann, S. & Grice, M. (2008): Articulatory gestures and focus marking in German. Proc. 4th International Conference of Speech Prosody, Campinas, Brazil, 457-460.

Honda, M.; Fujino, A. & Kaburagi, A. (2002) Compensatory responses of articulators to unexpected perturbation of the palate shape. Journal of Phonetics 30(*3*), 281-302.

Honda, M. & Murano, E. (2003) Effects of tactile and auditory feedback on compensatory articulatory response to an unexpected palatal perturbation. Proceedings of the 6th International Seminar on Speech Production, Sydney, 97-100.

Hoole, P. (1996a) Theoretische und methodische Grundlagen der Artikulationsanalyse in der experimentellen Phonetik. Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München 34, 3-173.

Hoole, P. (1996b) Issues in the acquisition, processing, reduction and parameterization of articulographic data. Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München 34, 158-173.

Hoole, P. (1999) On the lingual organization of the German vowel system. Journal of the Acoustical Society of America 106(*2*), 1020-1032.

Hoole, P. & Kühnert, B. (1996) Tongue-jaw coordination in German vowel production. Proceedings of the 1st ESCA Tutorial and Research Workshop on Speech Production Modelling, Autrans, 97-100.

Hoole, P.; Munhall, K. & Mooshammer, C. (1998) Do airstream mechanisms influence tongue movement paths? Phonetica 55, 131-146.

Houde, R. (1967) A study of tongue body motion during selected speech sounds. PhD thesis, University of Michigan.

Hughes, G. & Halle, M. (1956) Spectral properties of fricative consonants. Journal of the Acoustical Society of America 28, 303-310.

Hughes, O. M. & Abbs, J. H. (1976) Labial-mandibular coordination in the production of speech: Implications for the operation of motor equivalence. Phonetica 33(*3*), 199-221.

Huttenlocher, D. P. (1984) Acoustic-phonetic and lexical constraints in word recognition: lexical access using partial information. MS. Thesis, Massachussetts Institute of Technology.

Jannedy, S.; Weirich, M.; Brunner, J. & Mertins, M. (2010) Perceptual evidence for allophonic variation of the palatal fricative /ç/ in spontaneous Berlin German. Journal of the Acoustical Society of America 128, 2458.

Johnson, K. (1997) Speech perception without speaker normalization. An exemplar model. *in* K. Johnson & J. W. Mullennix (eds.), Talker Variability in Speech Processing, San Diego: Academic Press, 145-166.

Johnson, K. (2007) Decision and mechanisms in exemplar-based phonology. *in* M. J. Sole; P. Beddor & M. Ohala (eds.), Experimental Approaches to Phonology. Oxford: Oxford University Press, 25-40.

Johnson, K. & Azara, M. (2000) The perception of personal identity in speech: Evidence from the perception of twins speech. Unpublished manuscript, Retrieved from http://linguistics.berkeley.edu/~kjohnson/papers/twinPerc.pdf.

Johnson, K.; Ladefoged, P. & Lindau, M. (1993) Individual differences in vowel production. Journal of the Acoustical Society of America 94, 701-714.

Jones, J. A. & Munhall, K. (2000) Perceptual calibration of F0 production: Evidence from feedback perturbation. Journal of the Acoustical Society of America 108(*3*), 1246–1251.

Jones, J. A. & Munhall, K. (2002) The role of auditory feedback during phonation: Studies of Mandarin tone production. Journal of Phonetics 30(*3*), 303-320.

Jones, J. A. & Munhall, K. (2003) Learning to produce speech with an altered vocal tract: The role of auditory feedback. Journal of the Acoustical Society of America 113(*1*), 532-543.

Jongman, A.; Blumstein, S. & Lahir, A. (1985) Acoustic properties for dental and alveolar stop consonants: A cross-language study. Journal of Phonetics 13, 235-251.

Jongman, A.; Wayland, R. & Wong, S. (2000) Acoustic characteristics of English fricatives. Journal of the Acoustical Society of America 108, 1252-1263.

Katz, W.; Kripke, C. & Tallal, P. (1991) Anticipatory coarticulation in the speech of adults and young children: Acoustic, perceptual, and video data. Journal of Speech and Hearing Research 34, 1222-1232.

Kent, R. (1983) The segmental organization of speech. *in* P. MacNeilage (ed.), The Production of Speech, New York: Springer, 57-89.

Kent, R. & Moll, K. (1972) Cinefluorographic analyses of selected lingual consonants. Journal of Speech and Hearing Research 15, 453-473.

Klatt, D. & Klatt, L. (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers. Journal of the Acoustical Society of America 87, 820-857.

Koenig, L.; Lucero, J. C. & Perlman, E. (2008) Speech production variability in fricatives of children and adults: Results of functional data analysis. Journal of the Acoustical Society of America 124(*5*), 3158-3170.

Koeppen-Schomerus, G.; Spinath, F. M. & Plomin, R. (2003) Twins and non-twin siblings: Different estimates of shared environmental influence in early childhood. Twin Research 6, 97-105.

Köpke, B. & Schmid, M. S. (2007) Bilingualism and attrition. *in* B. Köpke; M. S. Schmid; M. Keijzer & S. Dostert, (eds.) Language Attrition. Theoretical Perspectives, Amsterdam, Philadelphia: John Benjamins, 1-7.

Kühnert, B. & Nolan, F. (1999) The origin of coarticulation. *in* W. J. Hardcastle & N. Hewlett (eds.), Coarticulation: Theory, Data and Techniques in Speech Production, Cambridge: Cambridge University Press, 7-30.

Künzel, H. (1987) Sprechererkennung: Grundzüge forensischer Sprachverarbeitung. Heidelberg: Kriminalistik Verlag.

Künzel, H. (2010) Automatic speaker recognition of identical twins. Journal of Speech, Language and the Law 17(*2*), 251-277.

Labov, W. (1980) The social origins of sound change. *in* W. Labov (ed.), Locating Language in Time and Space, New York: Academic Press, 251-266.

Labov, W. (1994) Principles of Linguistic Change. Volume 1: Internal Factors. Oxford: Blackwell.

Ladefoged, P. (1984) Out of chaos comes order: Physical, biological, and structural patterns in phonetics. Proceedings of the 10th International Congress of Phonetic Sciences, Utrecht, 83-95.

Ladefoged, P. & Broadbent, D. (1957) Information conveyed by vowels. Journal of the Acoustical Society of America 29(*1*), 98-104.

Ladefoged, P.; DeClerk, J.; Lindau, M. & Paptun, G. (1972) An auditory-motor theory of speech production. UCLA Working Papers in Phonetics 22, 48-75.

Ladefoged, P. & Maddieson, I. (1996) The Sounds of the Worlds Languages. Oxford: Blackwell.

Lane, H.; Denny, M.; Guenther, F.; Matthies, M.; Menard, L.; Perkell, J.; Stockman, E.; Tiede, M.; Vick, J. & Zandipour, M. (2005) Effects of bite blocks and hearing status on vowel production. Journal of the Acoustical Society of America 118, 1636–1646.

Lane, H.; Wozniak, J.; Matthies, M.; Svirsky, M. & Perkell, J. (1995) Phonemic resetting vs. postural adjustments in the speech of cochlear implant users: An exploration of voice-onset time. Journal of the Acoustical Society of America 98, 3096–3106.

Langer, P.; Tajtáková, M.; Bohov, P. & Klimes, I. (1999) Possible role of genetic factors in thyroid growth rate and in the assessment of upper limit of normal thyroid volume in iodine-replete adolescents. Thyroid 9(*6*), 557-562.

Laver, J. (1980) The Phonetic Description of Voice Quality. Cambridge: Cambridge University Press.

Laver, J. (1994) Principles of Phonetics. Cambridge: Cambridge University Press.

Lavoie, L. (2002) Subphonemic and suballophonic consonant variation: The role of the phoneme inventory. ZAS Papers in Linguistics 28, 39-54.

Lee, S. B. D. & Krivokapic, J. (2006) Functional data analysis of prosodic effects on articulatory timing. Journal of the Acoustical Society of America 119, 1661-1671.

Levon, E. (2006) Hearing "gay": Prosody, interpretation, and the affective judgments of men's speech. American Speech 81, 56-78.

Lewis, D. & Tuthill, C. (1940) Resonant frequencies and damping constants of resonators involved in the production of sustained vowels "o" and "ah". Journal of the Acoustical Society of America 11, 451-456.

Liberman, A. M.; Cooper, F.; Shankweiler, D. & Studdart-Kennedy, M. (1967) Perception of the speech code. Psychological Review 74, 431-461.

Liberman, A. M. & Mattingly, I. G. (1985) The motor theory of speech perception revised. Cognition 21, 1-36.

Lieberman, P. (1963) Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. Journal of the Acoustical Society of America 35, 344-353.

Lindblom, B. (1983) Economy of speech gestures. *in* P. MacNeilage (ed.), The Production of Speech, New York: Springer, 217-245.

Lindblom, B. (1984) Can the models of evolutionary biology be applied to phonetic problems? Proceedings of the 10th International Congress of Phonetic Sciences, Utrecht, 67-81.

Lindblom, B. (1988) Phonetic invariance and the adaptive nature of speech. *in* B.A. Elsendoom & H. Bouma (eds.), Working Models of Human Perception, London: Acadamic Press, 139-173.

Lindblom, B. (1990) Explaining phonetic variation: A sketch of the H&H theory. *in* W. J. Hardcastle & A. Marchal (eds.), Speech Production and Speech Modelling, Dordrecht: Kluwer, 403-439.

Lindblom, B. & Sundberg, J. (1971) Acoustical consequences of lip, tongue, jaw and larynx movement. Journal of the Acoustical Society of America 50, 1166-1179.

Linker, W. (1982) Articulatory and acoustic correlates of labial activity in vowels: A cross linguistic study. UCLA Working Papers in Phonetics 56, 1-134.

Linville, S. & Rens, J. (2001) Vocal tract resonance analysis of aging voice using long-term average spectra. Journal of Voice 15, 323-330.

Lipski, S.; Unger, S.; Grice, M. & Meister, I. (2011) Masked auditory feedback affects speech motor learning of a plosive duration contrast. Motor Control 15(*1*), 68-84.

Loakes, D. (2004) Front vowels as speaker-specific: Some evidence from Australian English. Proceedings of the 10th Australian International Conference on Speech Science & Technology, Sydney, 289-294.

Loakes, D. (2006) A forensic phonetic investigation into the speech patterns of identical and non-identical twins. PhD thesis. University of Melbourne, School of Languages.

Locke, J. L. & Mather, P. L. (1989) Genetic factors in the ontogeny of spoken language: Evidence from monozygotic and dizygotic twins. Journal of Child Language 16(*3*), 553-559.

Löfqvist, A. & Gracco, V. L. (1997) Lip and jaw kinematics in bilabial stop consonant production. Journal of Speech, Language and Hearing Research 40, 877-893.

Löfqvist, A. & Gracco, V. L. (2002) Control of oral closure in lingual stop consonant production. Journal of the Acoustical Society of America 111(*6*), 2811-2827.

Lucero, J. C. & Koenig, L. L. (2000) Time normalization of voice signals using functional data analysis. Journal of the Acoustical Society of America 108, 1408-1420.

Lucero, J. C. & Löfqvist, A. (2005) Measures of articulatory variability in VCV sequences. Acoustic Research Letters Online 6(*2*), 80-84.

Lucero, J. C.; Munhall, K. G.; Gracco, V. L. & Ramsay, J. O. (1997) On the registration of time and the patterning of speech movements. Journal of Speech, Language and Hearing Research 40, 1111-1117.

Luchsinger, R. & Arnold, G. (1965) Voice, Speech, Language. Clinical Communicology: Its Physiology and Pathology. Belmont, CA: Wadsworth Publishing Company.

Lundström, A. (1948) Tooth Size and Occlusion in Twins. Basel: Karger.

Mack, S. (2011) A sociophonetic analysis of /s/ variation in Puerto Rican Spanish. *in* L. A. Ortiz-López (ed.), Selected Proceedings of the 13th Hispanic Linguistics Symposium. Somerville, MA: Cascadilla Proceedings Project, 81-93.

Manuel, S. (1990) The role of contrast in limiting vowel-to-vowel coarticulation in different languages. Journal of the Acoustical Society of America 88, 1286-1298.

Mather, P. L. & Black, K. (1984) Heredity and environmental influences on preschool twin's language skills. Developmental Psychology 20, 303-308.

McAllister, J. (1991) The Processing of Lexically Stressed Syllables in Read and Spontaneous Speech. Language and Speech 34(*1*), 1-26.

McGurk, H. & MacDonald, J. (1976) Hearing lips and seeing voices. Nature 264, 746-748.

Mehta, G. & Cutler, A. (1988) Detection of target phonemes in spontaneous and read speech. Language and Speech 31 (*2*), 135-156.

Ménard, L.; Polak, M.; Denny, M.; Burton, E.; Lane, H.; Matthies, M. L..; Marrone, N.; Perkell, J. S.; Tiede, M. & Vick, J. (2007) Interactions of speaking condition and auditory feedback on vowel production in postlingually deaf adults with cochlear implants. Journal of the Acoustical Society of America 121(*6*), 3790-3801.

Meurer, E.; Wender, M.; Corleta, H. & Capp, E. (2004) Phono-articulatory variations of women in reproductive age and postmenopausal. Journal of Voice 18, 369-374.

Mitsuya, T.; MacDonald, E. N. & Munhall, K. G. (2009) Auditory feedback and articulatory timing. Journal of the Acoustical Society of America 126(*4*), 2223.

Mooshammer, C. (1998) Experimentalphonetische Untersuchungen zur artikulatorischen Modellierung der Gespanntheitsopposition im Deutschen. Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation (FIPKM), Universität München, 3-192.

Mooshammer, C. & Geng, C. (2008) Acoustic and articulatory manifestations of vowel reduction in German. Journal of the International Phonetic Association 38, 117-136.

Mooshammer, C.; Hoole, P. & Kühnert, B. (1995) On loops. Journal of Phonetics 23, 3-21.

Mooshammer, C.; Perrier, P.; Fuchs, S.; Geng, C. & Pape, D. (2004) An EMMA and EPG study on token-to-token variability. AIPUK 36, 47-63.

Munson, B. (2007) The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. Language and Speech 50, 125-142.

Nasir, S. M. & Ostry, D. J. (2008) Speech motor learning in profoundly deaf adults. Nature Neuroscience 11, 1217-1222.

Nasir, S. M. & Ostry, D. J. (2009) Auditory plasticity and speech motor learning. PNAS 106, 20470-20475.

Newman, R. S. (2003) Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. Journal of the Acoustical Society of America 113(*5*), 2850-2860.

Newman, R. S.; Clouse, S. A. & Burnham, J. L. (2001) The perceptual consequences of within-talker variability in fricative production. Journal of the Acoustical Society of America 109(*3*), 1181-1196.

Niemi, M.; Laaksonen, J.-P.; Ojala, S.; Aaltonen, O. & Happonen, R.-P. (2006) Effects of transitory lingual nerve impairment on speech: An acoustic study of sibilant sound /s/. International Journal of Oral Maxillofac Surgery 35, 920–923.

Nittrouer, S; Studdart-Kennedy, M. & McGowan, R. S. (1989) The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. Journal of Speech and Hearing Research 32, 120-132.

Nolan, F. (1983) The Phonetic Bases of Speaker Recognition. Cambridge: Cambridge University Press.

Nolan, F. & Oh, T. (1996) Identical twins, different voices. Forensic Linguistics 3, 39-49.

Nolan, F.; Oh, T.; McDougal, K.; de Jong, G. & Hudson, T. (2006) A forensic phonetic study of "dynamic" sources of variability in speech: The DyViS project. Proceedings of the 11th Australian International Conference on Speech Science & Technology, Auckland, 13-18.

Nordström, P. & Lindblom, B. (1975) A normalization procedure for vowel formant data. Proceedings of the 8th Intenational Congress of Phonetic Sciences, Leeds, England, 212.

Öhman, S. (1966) Coarticulation in VCV utterances: spectrographicmeasurements. Journal of the Acoustical Society of America 39, 151-168.

Ohala, J. (1983) The origin of sound patterns in vocal tract constraints. *in* P. McNeillage (ed.), The Production of Speech, New York: Springer, 189-216.

Ooki, S. (2005) Genetic and environmental influences on stuttering and tics in Japanese twin children. Twin Research and Human Genetics 8(*1*), 69-75.

Parush, A.; Ostry, D. & Munhall, K. (1983) A kinematic study of lingual coarticulation in VCV sequences. Journal of the Acoustical Society of America 74(*4*), 1115-1125.

Payan, Y. & Perrier, P. (1997) Synthesis of V–V sequences with a 2D biomechanical tongue model controlled by the equilibrium point hypothesis. Speech Communication 22(*2/3*), 185-205.

Perkell, J. S. (1997) Articulatory processes. *In* W. J. Hardcastle; J. Laver & F. E. Gibbon (eds.), The Handbook of Phonetic Sciences, Oxford and Cambridge: Blackwell Publishers, 333-370.

Perkell, J. S.; Denny, M.; Lane, H.; Guenther, F. H.; Matthies, M. L.; Tiede, M.; Vick, J.; Zandipour, M. & Burton, E. (2007) Effects of masking noise on vowel and sibilant contrasts in normal-hearing speakers and postlingually deafened cochlear implant users. Journal of the Acoustical Society of America 121, 505-518.

Perkell, J. S.; Guenther, F. H.; Lane, H.; Matthies, M. L..; Stockmann, E.; Tiede, M. & Zandipour, M. (2004) The distinctiveness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. Journal of the Acoustical Society of America 116(*4*), 2338-2344.

Perkell, J. S.; Guenther, F. H; Lane, H.; Marrone, N.; Matthies, M. L.; Stockmann, E.; Tiede, M. & Zandipour, M. (2006) Production and perception of phoneme contrasts covary across speakers. *in* J. Harrington & M. Tabain (eds.), Speech Production: Models, Phonetic Processes, and Techniques, New York: Psychology Press, 69-84.

Perkell, J. S.; Lane, H.; Ghosh, S.; Matthies, M. L.; Tiede, M.; Guenther, F. & Ménard, L. (2008) Mechanisms of vowel production: Auditory goals and speaker acuity. Proceedings of the 8th International Seminar on Speech Production, Strasbourg, 28-32.

Perkell, J. S.; Matthies, M. L.; Svirsky, M. A. & Jordon, M. I. (1993) Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot "motor equivalence" study. Journal of the Acoustical Society of America 93(*5*), 2948-2961.

Perkell, J. S.; Matthies, M. L.; Tiede, M.; Lane, H.; Zandipour, M.; Marrone, N.; Stockmann, E. & Guenther, F. (2004) The distinctness of speakers /s/-/ʃ/ contrast is related to their auditory discrimination & use of an articulatory saturation effect. Journal of Speech, Language & Hearing Research 47(*6*), 1259-1269.

Perkell, J. S. & Nelson, W. L. (1985) Variability in the production of the vowels /i/ and /a/. Journal of the Acoustical Society of America 77, 1889-1895.

Perkell, J. S.; Numa, W.; Vick, J.; Lane, H.; Balkany, T. & Gould, J. (2001) Language-specific, hearing-related changes in vowel spaces: A preliminary study of English- and Spanish-speaking cochlear implant users. Ear and Hearing 22, 461-470.

Perrier, P. (2005) Control and representations in speech production. ZAS Papers in Linguistics 40, 109-133.

Perrier, P.; Payan, Y.; Zandipour, M. & Perkell, J. S. (2003) Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. Journal of the Acoustical Society of America 114(*3*), 1582-1599.

Peterson, G. E. & Barney, H. L. (1952) Control methods used in a study of the vowels. Journal of the Acoustical Society of America 24(*2*), 175-184.

Pierrehumbert, J. (2001) Exemplar dynamics: Word frequency, lenition, and contrast. *in* J. Bybee & P. Hopper (eds.), Frequency Effects and the Emergence of Linguistic Structure, Amsterdam: John Benjamins, 137-157.

Pierrehumbert, J. (2002) Word-specific phonetics. *in* C. Gussenhoven & N. Warner (eds.), Papers in Laboratory Phonology 7, Berlin: Mouton de Gruyter, 101-139.

Pinheiro, J. & Bates, D. (2000) Mixed-Effects Models in S and S-Plus. Statistics and Computing Series. New York: Springer.

Pompino-Marschall, B. (2003) Einführung in die Phonetik (2nd ed.). Berlin: de Gruyter.

Pouplier, M. & Goldstein, L. (2005) Asymmetries in the perception of speech production errors. Journal of Phonetics 33, 47-75.

Przybyla, B. D.; Horii, J. & Crawford, M. H. (1992) Vocal fundamental frequency in a twin sample: Looking for a genetic effect. Journal of Voice 6(*3*), 261-266.

R Development Core Team (2008) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Ramishvili, G. (1966) Automatic voice recognition. Engineering Cybernetics 5, 84-90.

Ramsay, J. O. & Silverman, B. W. (1997) Functional Data Analysis. New York: Springer.

Rizzolatti, G. & Arbib, M. (1998) Language within our grasp. Trends in Neuroscience 21, 188-194.

Rizzolatti, G. & Craighero, L. (2004) The mirror neuron system. Annual Review of Neuroscience 27, 169-192.

Rizzolatti, G.; Fogassi, L. & Gallese, V. (2001) Neurophysiological mechanisms underlying the understanding and imitation of action. Nature Review Neuroscience 2, 661-670.

Rose, P. (2002) Forensic Speaker Identification. London: Taylor & Francis.

Rosenberg, A. (1973) Listener performance in speaker verification tasks. IEEE Transactions on Acoustics, Speech, and Signal Processing 21(*3*), 221-225.

Rvachew, S. (1994) Speech perception training can facilitate sound production learning. Journal of Speech, Language, and Hearing Research 37, 347-357.

Ryalls, J.; Shaw, H. & Simon, M. (2004) Voice onset time production in older and younger female monozygotic twins. Folia Phoniatrica et Logopaedica 56, 165-169.

Sambur, M. R. (1975) Selection of acoustic features for speaker identification. IEEE Transactions on Acoustics, Speech, and Signal Processing 23(*2*), 176-182.

Sataloff, R. (1995) Genetics of the voice. Journal of Voice 9, 16-19.

Savariaux, C.; Perrier, P. & Orliaguet, J.-P. (1995) Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. Journal of the Acoustical Society of America 98, 2428-2442.

Scarr, S. & Carter-Saltzman, L. (1979) Twin method: Defense of a critical assumption. Behavior Genetics 9(*6*), 527-542.

Schiller, N. & Köster, O. (1996) Evaluation of a foreign speaker in forensic phonetics: A report. Forensic Linguistics: International Journal of Speech, Language and the Law 3, 176-185.

Shadle, C. (1985) The acoustics of fricative consonants. Technical Report 506, Cambridge: MIT Research Laboratory of Electronics.

Shiller, D. M.; Laboissière, R. & Ostry, D. J. (2002) Relationship between jaw stiffness and kinematic variability in speech. Journal of Neurophysiology 88, 2329-2340.

Shiller, D. M.; Sato, M.; Gracco, V. L. & Baum, S. R. (2009) Perceptual recalibration of speech sounds following speech motor learning, Journal of the Acoustical Society of America 125(*2*), 1103-1113.

Simberg, S.; Santtila, P.; Soveri, A.; Varjonen, M.; Sala, E. & Sandnabba, N. K. (2009) Exploring genetic and environmental effects in dysphonia: A twin study. Journal of Speech, Language and Hearing Research 52, 153-163.

Sørensen, M. (acc.) Voice line-ups: Speakers' F0 values influence the reliability of voice recognitions. International Journal of Speech, Language and the Law.

Spinath, F. M. (2005) Twin designs. *in* B. S. Everitt & D. C. Howell (eds.), Encyclopedia of Statistics in Behavioral Science, Chichester: John Wiley & Sons, 2071-2074.

Spinath, F. M.; Riemann, R.; Hempel, S.; Schlangen, B.; Weiß, R.; Borkenau, P. & Angleitner, A. (1999) A day in the life: Description of the German Observational Study on Adult Twins (GOSAT) assessing twin similarity in controlled laboratory settings. *in* I. Mervielde; I. Deary; F. DeFruyt & F. Ostendorf (eds.), Personality Psychology in Europe 7, Tilburg, NL: Tilburg University Press, 311-328.

Stevens, K. N. (1972) The quantal nature of speech: Evidence from articulatory-acoustic data. *in* P. B. Denes & E. E. David Jr. (eds.), Human Communication: A Unified View. NewYork: McGraw Hill, 51-66.

Stevens, K. N. (1989) On the quantal nature of speech. Journal of Phonetics 17, 3-46.

Stevens, K. & Blumstein, S. (1978) Invariant cues for place of articulation in stop consonants. Journal of the Acoustical Society of America 64, 1358-1368.

Stevens, K.; Williams, C.; Carbonell, J. & Woods, B. (1968) Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material. Journal of the Acoustical Society of America 44(*6*), 1596-1607.

Stone, M. (1995) How the tongue takes advantage of the palate during speech. *in* F. Bell-Berti & L. Raphael (eds.), Producing Speech: Contemporary Issues, New York: American Institute of Physics, 143-153.

Stone, M. & Lundberg, A. (1994) Tongue-palate interactions in consonants vs. vowels. Proceedings of the International Conference on Spoken Language Processing, Yokohama, Japan, 49-52.

Stone, M. & Lundberg, A. (1996) Three-dimensional tongue surface shapes of English consonants and vowels. Journal of the Acoustical Society of America 99, 3728-3737.

Strand, E. (1999) Uncovering the role of gender stereotypes in speech perception. Journal of Language and Social Psychology 18(*1*), 86-99.

Strange, W., Jenkins, J. J., & Johnson, T. (1983) Dynamic specification of coarticulated vowels. Journal of the Acoustical Society of America 74, 694-705.

Sussman, H. M.; Hoemeke, K. & McCaffrey, H. A. (1992) Locus equation as an index of coarticulation for place of articulation distinctions in children. Journal of Speech and Hearing Research 35, 769-781.

Sussman, H. M.; McCaffrey, H. A. & Matthews, S. A. (1991) An investigation of locus equations as a source of relational invariance for stop place articulation. Journal of the Acoustical Society of America 90(*3*), 1309-1325.

Svirsky, M. A. & Tobey, E. A. (1991) Effect of different types of auditory stimulation on vowel formant frequencies in multichannel cochlear implant users. Journal of the Acoustical Society of America 89, 2895–2904.

Takahashi, H. & Koike, Y. (1975) Some perceptual dimensions and acoustical correlates of pathologic voices. Acta Oto-Laryngologica (suppl) 338, 1-24.

Templin, M. C. & Darley, F. L. (1969) The Templin-Darley Tests of Articulation: A Manual and Discussion of Articulation Testing, Iowa City: Bureau of Educational Research and Service, Division of Extension and University Services, University of Iowa.

Tjaden, K. & Turner, G. (1997) Spectral properties of fricatives in amyotrophic lateral sclerosis. Journal of Speech, Language and Hearing Reserach 40, 1358-1372.

Toda, M. (2006) Deux stratégies articulatoires pour la réalisation du contraste acoustique des sibilantes /s/ et /ʃ/ en français. Actes des XXVI ès Journées d'Étude de la Parole, Dinard, 65-68.

Toda, M.; Maeda, S. & Honda, K. (2010) Formant-cavity affiliation in sibilant fricatives. *in* S. Fuchs; M. Toda & M. Zygis (eds.), Turbulent Sounds – An Interdisciplinary Guide, Berlin: Mouton de Gruyter, 343-374.

van Bergem, D. R. (1993) Acoustic vowel reduction as a function of sentence accent, word stress, and word class. Speech Communication 12(*1*), 1-23.

van Bergem, D. R. (1994) A model of the coarticulatory effects on the schwa. Speech Communication 14, 143-162.

van Lierde, K.; Vinck, B.; Ley, S.; Clement, G. & van Cauwenberge, P. (2005) Genetics of vocal quality characteristics in monozygotic twins: A multiparameter approach. Journal of Voice 19(*4*), 511-518.

Venables, W. & Ripley, B. (2003) Modern Applied Statistics with S. New York: Springer.

Vick, J.; Lane, H.; Perkell, J.; Gould, J. & Zandipour, M. (2001) Covariation of cochlear implant users perception and production of vowel contrasts and their identification by listeners with normal hearing. Journal of Speech, Language and Hearing Research 44, 1257-1267.

Villacorta, V. M.; Perkell, J. S. & Guenther, F. H. (2007) Sensorimotor adaptation to feedback perturbation of vowel acoustics and its relation to perception. Journal of the Acoustical Society of America 122(*4*), 2306-2319.

Warren, P. (2005) Patterns of late rising in New Zealand English: Intonational variation or intonation change? Language Variation and Change 17, 209–230.

Watson, C. I. & Harrington, J. (1999) Acoustic evidence for dynamic formant trajectories in Australian English vowels. Journal of the Acoustical Society of America 106, 458-468.

Wendahl, R. (1966a) Laryngeal analog synthesis of jitter and shimmer auditory parameters of harshness. Folia Phoniatrica 18(*2*), 98-108.

Wendahl, R. (1966b) Some parameters of auditory roughness. Folia Phoniatrica 18(*1*), 26-32.

Whiteside, S. P. & Rixon, E. (2000) Identification of twins from pure (single) speaker and hybrid (fused) syllables: An acoustic and perceptual case study. Perceptual and Motor Skills 91, 933-947.

Whiteside, S. P. & Rixon, E. (2001) Speech patterns of monozygotic twins: An acoustic case study of monosyllabic words. Phonetician 84, 9-22.

Whiteside, S. P. & Rixon, E. (2003) Speech characteristics of monozygotic twins and a same-sex sibling: An acoustic case study of coarticulation patterns in read speech. Phonetica 60, 273-297.

Wiese, R. (1996) The Phonology of German. Oxford: Clarendon Press.

Winkler, R.; Fuchs, S. & Perrier, P. (2006) The relation between differences in vocal tract geometry and articulatory control strategies in the production of French vowels: Evidence from MRI and modelling. Proceedings of the 7th International Seminar on Speech Production, Ubatuba, 509-516.

Wolf, H.; Spinath, F. M. & Angleitner, A. (2003) Similar and yet still different? German Observational Study of Adult Twins. German Research 28, 8-10.

Xue, S. & Hao, G. (2003) Changes in the human vocal tract due to aging and the acoustic correlates of speech production: A pilot study. Journal of Speech, Language and Hearing Research 46, 689-701.

Yumoto, E.; Gould, W. & Baer, T. (1982) Harmonics-to-noise ratio as an index of the degree of hoarseness. Journal of the Acoustical Society of America 71(*6*), 1544-1550.

Yumoto, E.; Sasaki, Y. & Okamura, H. (1984) Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness. Journal of Speech and Hearing Research 27(*1*), 2-6.

List of Tables

List of Figures

APPENDIX A

Methods

Table A.1: *Questionnaire of the pilot study.*

Name/ Wohnort:
E-mail/Telefonnr.:
Geburtsdatum/-ort:
Aufgewachsen in:
Schulabschluss/Beruf:
Auslandsaufenthalte:
Fremdsprachen:
Familienstand/Kinder:
Physische Statur: Größe/Gewicht
Seid ihr eineiige Zwillinge?
Wie weit weg voneinander wohnt ihr?
jetzt (seit wann):
früher (wie lange):
Wie oft siehst du deine(n) Zwillingsbruder/schwester?
Früher/ heute:
Wie gut verstehst du dich mit ihm/ihr?
Hast du noch weitere Geschwister? Wie ist das Verhältnis zu ihnen?
Wart ihr in der Kindheit immer zusammen/gleiche Hobbies?
Tragt ihr gleiche Kleidung?
Früher/ heute:
Habt ihr gleiche Freunde?

Früher/ heute:
Legten deine Eltern Wert auf Individualität?
Bist du gerne ein „Zwilling“?
Siehst du mehr Vorteile oder mehr Nachteile darin, ein Zwilling zu sein?
Wurdet/werdet ihr oft verwechselt?
Aussehen:
Stimme:
Krankheiten/Unfälle/Operationen, die Sprach-/Hörfähigkeit beeinträchtigen/verändert haben?

Table A.2: *Questionnaire of the EMA study.*

Name: _____Kürzel: _____

Schon bei PILOT-AKUSTIK teilgenommen?

Wenn NEIN:

- anderen Fragebogen ausfüllen (cf. Table A.1)!

Wenn JA:

- hat sich etwas verändert hinsichtlich Beziehung zu Zwilling? (Zusammensein?)

Fragen zu ANATOMIE Sprachapparat:

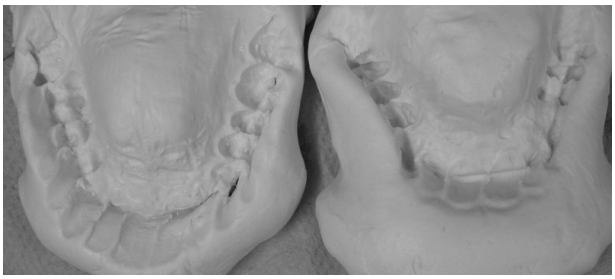
- Operationen? (welche?)
- Zahnspange? (wie lange?)
- Zahnprothesen? Zahnlücken?
- Schnuller? Daumen gelutscht?
- Raucher? (wie lange?)

Sprech-, Sprach- Stimmtherapie? Hörprobleme? (schwere Mittelohrentzündungen?)

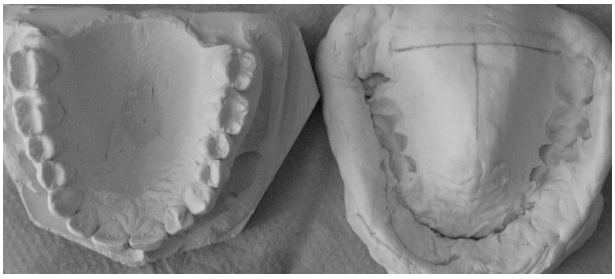
Gesangsausbildung? Chor?

Figure A.1: *Silicone dental-palatal casts for the male twins.*

DZm1 (HMFM)



MZm1 (SLCL)



MZm2(MIMA)

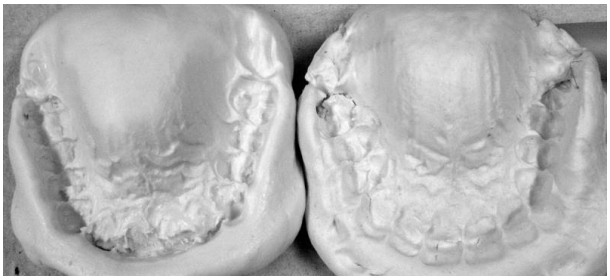


Figure A.2: *Adjustment of palatal contour and articulatory data of MZf1 (HF = blue, AF (raw) = blue, AF (rotated) = red).*

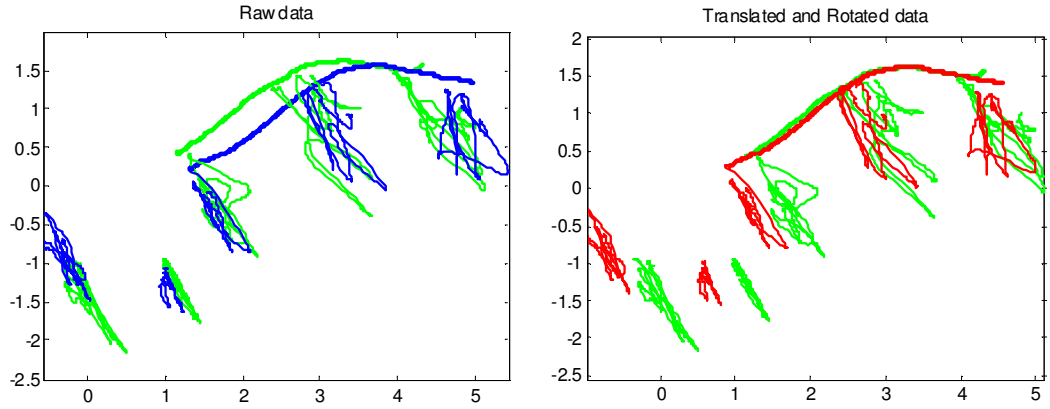


Figure A.3: *Adjustment of palatal contour and articulatory data of MZm1 (CL = green, SL(raw) = blue, SL(rotated) = red)*

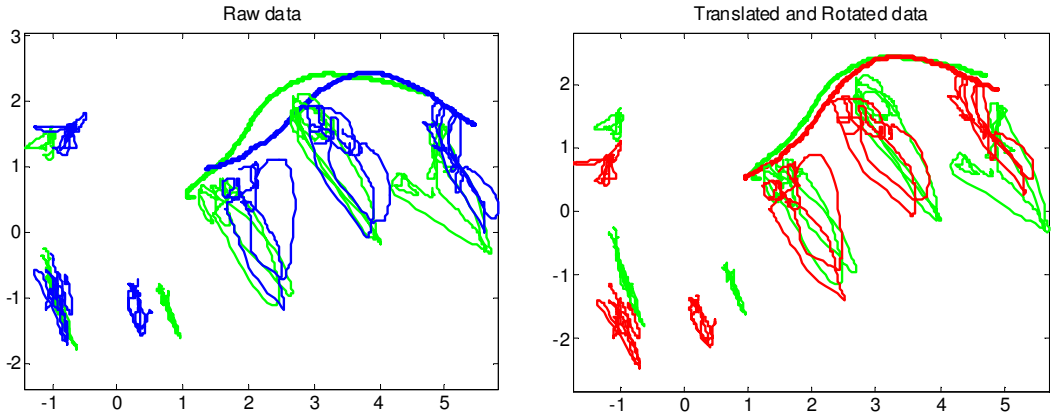


Figure A.4: *Adjustment of palatal contour and articulatory data of MZm2 (MI = green, MA(raw) = blue, MA(rotated) = red).*

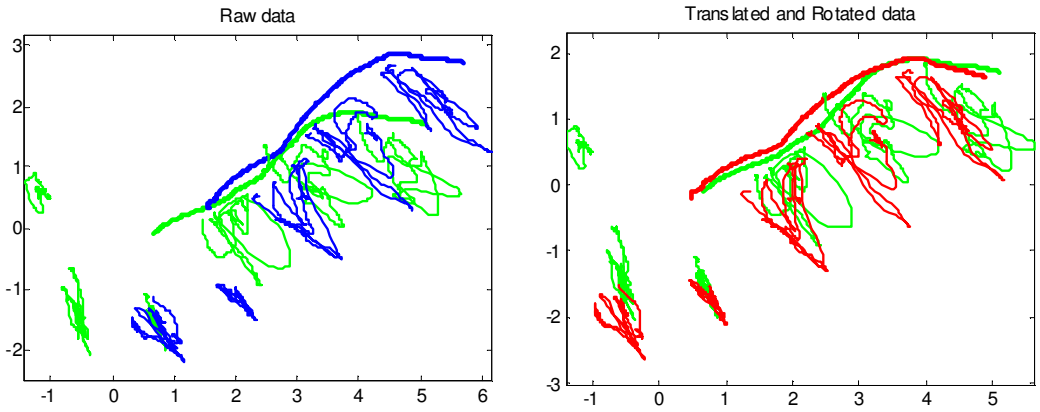


Figure A.5: *Adjustment of palatal contour and articulatory data of DZf1 (SR = green, LR(raw) = blue, LR(rotated) = red)*

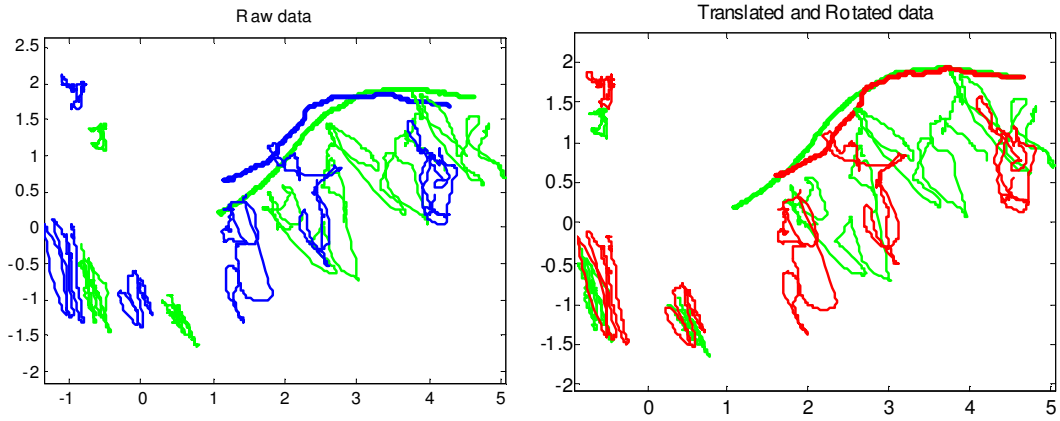


Figure A.6: *Adjustment of palatal contour and articulatory data of DZf2 (MG = green, TG(raw) = blue, TG(rotated) = red).*

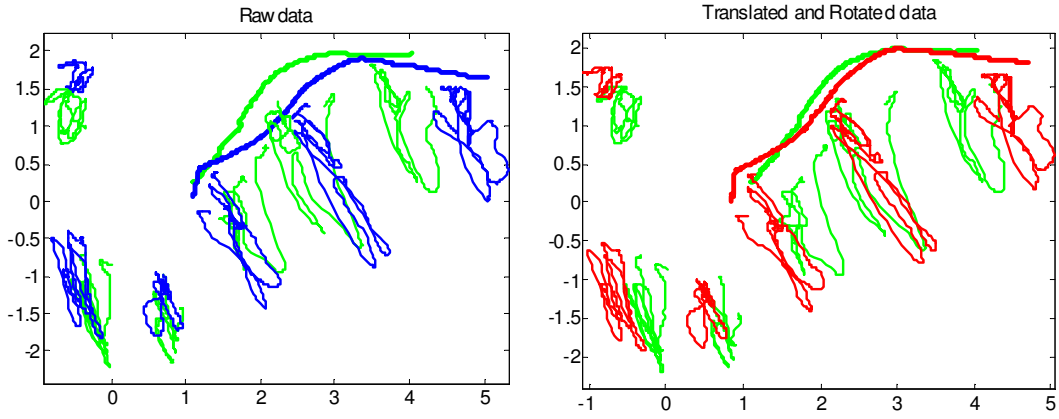


Table A.3: *Recorded Speech Material.*

Recording (A)

Ich bitte Datei zu sagen.
Ich bitte Datum zu sagen.
Ich bitte Balu zu sagen.
Ich bitte Bali zu sagen.
Ich küsse Gaba im Garten.
Ich küsse Guba im Garten.
Ich küsse Giba im Garten.
Ich küsse Kiba im Garten.
Ich suche Taler im Garten.
Ich suche Talent im Garten.
Ich suche Pape im Garten.
Ich suche Papier im Garten.
Ich summe Gala im Garten.
Ich summe Galopp im Garten.
Ich summe Kakadu im Garten.
Ich summe KakAdu im Garten.
Ich wasche Haga im Garten.
Ich wasche Hagi im Garten.
Ich wasche Hagu im Garten.
Ich wasche Haku im Garten.
Ich grüße Haka im Garten.
Ich grüße Haki im Garten.
Ich grüße Kuba im Garten.
Ich grüße Kaba im Garten.
Ich liebe Bali am Montag.
Ich liebe Papier am Montag.
Ich liebe Pape am Montag.
Ich liebe Galopp am Montag.
Ich liebe Kakadu am Montag.
Ich liebe Taler am Montag.
Ich liebe Talent am Montag.
Ich liebe Balu am Montag.
Ich liebe KakAdu am Montag.
Ich liebe Gala am Montag.
Ich liebe Datum am Montag.
Ich liebe Datei am Montag.

Recording (B)

Gestern war ich bei Peter. Kakadus mag er am liebsten.
Gestern sah ich bei Peter Kakadus und andere Vögel.
Ihm flog ein blauer und kein roter Kakadu vor die Kamera.
Ich suche nicht DEN Rad, aber DAS Rad.
Ich suche nicht DAS Rat, aber DEN Rat.

APPENDIX B

Statistics for Vowels

Table B.1: *Results of ANOVAs for each vowel and articulatory target position with factor SPEAKER and dependent variable HORIZONTAL or VERTICAL TONGUE POSITION.*

Vowel	articulatory position	F values, df and significance level
/a/	Horizontal position of tongue back	$F(9, 363) = 313.58$, $p < 0.001$
/a/	Vertical position of tongue back	$F(9, 363) = 534.78$, $p < 0.001$
/i:/	Horizontal position of tongue dorsum	$F(9, 431) = 185.20$, $p < 0.001$
/i:/	Vertical position of tongue dorsum	$F(9, 431) = 421.19$, $p < 0.001$
/u:/	Horizontal position of tongue back	$F(9, 373) = 187.39$, $p < 0.001$
/u:/	Vertical position of tongue back	$F(9, 373) = 540.36$, $p < 0.001$

Table B.2: *Results of post hoc Tukey tests for effect of SPEAKER on dependent variable HORIZONTAL/VERTICAL TONGUE POSITION for the vowels /a/, /i:/ and /u:/, Significance levels: $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*'.*

Comparison	vowel	Mean diff in horizontal tongue position	Mean diff in vertical tongue position
MZf1a-MZf1b	/a/	0.25***	-0.45***
MZf2a-MZf2b		-0.05	-0.22***
MZm1a-MZm1b		-0.22***	0.74***
DZf1a-DZf1b		0.24***	0.49***
DZf2a-DZf2b		0.82***	-0.06
MZf1a-MZf1b	/i:/	-0.19***	-0.05
MZf2a-MZf2b		-0.17***	-0.03
MZm1a-MZm1b		-0.43***	-0.17***
DZf1a-DZf1b		0.13**	-0.01

DZf2a-DZf2b		0.47***	0.08*
MZf1a-MZf1b	/u:/	-0.17***	0.11***
MZf2a-MZf2b		-0.01	-0.24***
MZm1a-MZm1b		-0.33***	0.18***
DZf1a-DZf1b		0.04	0.59***
DZf2a-DZf2b		0.58***	0.05

Table B.3: *Results of ANOVAs for each stress condition and articulatory target position with factor SPEAKER and dependent variable HORIZONTAL or VERTICAL TONGUE POSITION.*

Vowel	Articulatory position	F values, df and significance level
/i:/ (stressed)	Horizontal position of tongue dorsum	$F(9, 89) = 64.07$, $p < 0.001$
/i:/ (stressed)	Vertical position of tongue dorsum	$F(9, 89) = 230.20$, $p < 0.001$
/i/ (unstressed)	Horizontal position of tongue dorsum	$F(9, 84) = 33.97$, $p < 0.001$
/i/ (unstressed)	Vertical position of tongue dorsum	$F(9, 84) = 95.32$, $p < 0.001$

Table B.4: *Results of post hoc Tukey tests for effect of SPEAKER on dependent variable HORIZONTAL or VERTICAL TONGUE POSITION for /i:/ in /agi/ in a stressed and unstressed position, Significance levels: p < 0.001 '***', p < 0.01 '**', p < 0.05 '*'.*

Comparison	Stress condition	Mean diff in horizontal tongue position	Mean diff in vertical tongue position
MZf1a-MZf1b	stressed	-0.48***	0.04
MZf2a-MZf2b		-0.12	0.05
MZm1a-MZm1b		-0.44***	-0.24***
DZf1a-DZf1b		0.23**	0.13**
DZf2a-DZf2b		0.30**	0.17***
MZf1a-MZf1b	unstressed	0.16	0.06
MZf2a-MZf2b		0.19	0.15*
MZm1a-MZm1b		-0.19	-0.14
DZf1a-DZf1b		-0.01	0.15*
DZf2a-DZf2b		0.32**	0.06

Table B.5: *Results of ANOVAs for each vowel and F1-F4 with factor SPEAKER and dependent variable FORMANT.*

Vowel	Formant	F values, df and significance level
/a/	F1	$F(13, 492) = 140.74$, $p < 0.001$
/a/	F2	$F(13, 492) = 182.47$, $p < 0.001$
/a/	F3	$F(13, 492) = 59.71$, $p < 0.001$
/a/	F4	$F(13, 492) = 61.93$, $p < 0.001$
/i:/	F1	$F(13, 566) = 35.07$, $p < 0.001$
/i:/	F2	$F(13, 566) = 274.21$, $p < 0.001$
/i:/	F3	$F(13, 566) = 118.51$, $p < 0.001$
/i:/	F4	$F(13, 566) = 156.79$, $p < 0.001$
/u:/	F1	$F(13, 505) = 68.59$, $p < 0.001$
/u:/	F2	$F(13, 505) = 94.00$, $p < 0.001$
/u:/	F3	$F(13, 505) = 189.08$, $p < 0.001$
/u:/	F4	$F(13, 505) = 116.60$, $p < 0.001$

Table B.6: *Results of post hoc Tukey tests for effect of SPEAKER on dependent variable FORMANT for the vowels /a/, /i:/ and /u:/, Significance levels: $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*'.*

Comparison	vowel	Mean diff in F1	Mean diff in F2	Mean diff in F3	Mean diff in F4
MZf1a-MZf1b	/a/	79.48***	128.79***	196.09***	-49.85
MZf2a-MZf2b		19.74	17.52	-8.50	349.12***
MZm1a-MZm1b		-34.08	-4.24	-324.36***	-162.36
MZm2a-MZm2b		47.16**	-87.00**	-115.53	161.70
DZf1a-DZf1b		-14.51	-310.89***	-3.41	-241.98**
DZf2a-DZf2b		123.93***	50.07	-462.26***	284.65***
DZm1a-DZm1b		2.80	79.04**	10.06	303.55***

MZf1a-MZf1b	/i:/	-13.51	241.97***	-8.08	246.08***
MZf2a-MZf2b		-29.13***	-19.21	99.54	-185.74**
MZm1a-MZm1b		23.81***	-262.87***	-638.95***	-423.52***
MZm2a-MZm2b		3.41	38.14	-206.73**	-313.42***
DZf1a-DZf1b		-3.24	56.18	28.46	-233.23***
DZf2a-DZf2b		-5.36	51.93	-186.49***	38.11
DZm1a-DZm1b		-12.97	104.32**	51.99	10.42
MZf1a-MZf1b	/u:/	21.52	-98.18	68.95	-18.94
MZf2a-MZf2b		18.86	-4.88	-735.44***	31.21
MZm1a-MZm1b		66.94***	232.63***	-101.94	-131.14
MZm2a-MZm2b		-3.29	-0.62	-212.78***	473.58***
DZf1a-DZf1b		-12.31	-268.90***	-272.63***	-261.98***
DZf2a-DZf2b		-37.23***	-159.01***	-275.65***	-272.13***
DZm1a-DZm1b		-25.14*	123.38***	-111.35	-38.74

Table B.7: *Results of ANOVAs for each stress condition (stressed = /'gi:ba/, unstressed = /'hagi/) and F1-F4 with factor SPEAKER and dependent variable FORMANT.*

condition	Formant	F values, df and significance level
stressed	F1	$F(13, 116) = 34.85$, $p < 0.001$
stressed	F2	$F(13, 116) = 146.87$, $p < 0.001$
stressed	F3	$F(13, 116) = 32.04$, $p < 0.001$
stressed	F4	$F(13, 116) = 36.63$, $p < 0.001$
unstressed	F1	$F(13, 109) = 18.02$, $p < 0.001$
unstressed	F2	$F(3, 109) = 98.69$, $p < 0.001$
unstressed	F3	$F(3, 109) = 23.72$, $p < 0.001$
unstressed	F4	$F(3, 109) = 87.42$, $p < 0.001$

Table B.8: *Results of post hoc Tukey tests for effect of SPEAKER on dependent variable FORMANT for /i/ in /agi/ in a stressed and unstressed position, Significance levels: $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*'.*

Comparison	condition	Mean diff in F1	Mean diff in F2	Mean diff in F3	Mean diff in F4
MZf1a-MZf1b	stressed	-54.86***	79.45	-161.27	31.64
MZf2a-MZf2b		88.80 ***	-90.89	151.31	-123.87
MZm1a-MZm1b		11.76	-90.89	-352.94**	-178.23
MZm2a-MZm2b		7.90	40.13	-251.63	15.61
DZf1a-DZf1b		-38.95	-65.20	12.71	101.26
DZf2a-DZf2b		-55.78***	-80.69	-138.70	-157.46
DZm1a-DZm1b		27.61	67.97	-144.22	-27.29
MZf1a-MZf1b	unstressed	-15.43	67.55	-213.48	-130.65
MZf2a-MZf2b		-24.24	21.58	140.25	-86.97
MZm1a-MZm1b		25.69	-269.73***	-377.96***	-189.44
MZm2a-MZm2b		10.81	-49.73	-265.89*	-627.08***
DZf1a-DZf1b		40.95*	-49.83	44.66	-59.59
DZf2a-DZf2b		-93.07***	191.74*	136.47	81.34
DZm1a-DZm1b		-13.83	-119.62	-211.07	-86.36

APPENDIX C

Statistics for Sibilants

Table C.1: *Results of ANOVAs for each sibilant and articulatory target position with factor SPEAKER and dependent variable HORIZONTAL or VERTICAL TONGUE POSITION.*

sibilant	Articulatory position	F values, df and significance level
/s/	Horizontal position of tongue tip	$F(11, 434) = 253.89$, $p < 0.001$
/s/	Vertical position of tongue tip	$F(11, 434) = 290.68$, $p < 0.001$
/ʃ/	Horizontal position of tongue tip	$F(11, 436) = 344.82$, $p < 0.001$
/ʃ/	Vertical position of tongue tip	$F(11, 436) = 1042.9$, $p < 0.001$

Table C.2: *Results of post hoc Tukey tests for effect of SPEAKER on dependent variable HORIZONTAL or VERTICAL TONGUE POSITION for /s/ and /ʃ/, Significance levels: p < 0.001 '***', p < 0.01 '**', p < 0.05 '*'.*

Comparison	sibilant	Mean diff in horizontal tongue position	Mean diff in vertical tongue position
MZf1a-MZf1b	/s/	0.17***	-0.01
MZf2a-MZf2b		-0.55***	-0.36***
MZm1a-MZm1b		-0.47***	0.06
MZm2a-MZm2b		0.52***	0.23***
DZf1a-DZf1b		-0.46***	0.16***
DZf2a-DZf2b		0.41***	0.37***
MZf1a-MZf1b	/ʃ/	0.11**	0.04
MZf2a-MZf2b		-0.70***	-0.41***
MZm1a-MZm1b		-0.56***	-0.14***
MZm2a-MZm2b		-0.03	0.05
DZf1a-DZf1b		-0.56***	0.29***
DZf2a-DZf2b		0.34***	1.29***

Table C.3: *Results of ANOVAs for each sibilant with factor SPEAKER and dependent variable COG or PEAK.*

sibilant	Acoustic parameter	F values, df and significance level
/s/	COG	$F(13, 447) = 33.59$, $p < 0.001$
/s/	PEAK	$F(13, 433) = 11.56$, $p < 0.001$
/ʃ/	COG	$F(13, 503) = 55.47$, $p < 0.001$
/ʃ/	PEAK	$F(13, 481) = 15.87$, $p < 0.001$

Table C.4: *Results of post hoc Tukey tests for effect of SPEAKER on dependent variable COG or PEAK for /s/ and /ʃ/, Significance levels: p < 0.001 '***', p < 0.01 '**', p < 0.05 '*'.*

Comparison	sibilant	Mean diff in COG	Mean diff in PEAK
MZf1a-MZf1b	/s/	281.59	680.09
MZf2a-MZf2b		481.43***	292.98
MZm1a-MZm1b		508.08***	850.59*
MZm2a-MZm2b		-487.66***	462.58
DZf1a-DZf1b		175.72	-249.98
DZf2a-DZf2b		403.35	131.67
DZm1a-DZm1b		-641.14***	583.69
MZf1a-MZf1b	/ʃ/	69.56	-214.24
MZf2a-MZf2b		250.14**	401.28
MZm1a-MZm1b		95.11	-440.57
MZm2a-MZm2b		-178.74	-560.81
DZf1a-DZf1b		826.97***	1333.57***
DZf2a-DZf2b		81.94	581.82
DZm1a-DZm1b		-987.22***	-1448.81***

Table C.5: *Results of ANOVAs for each sibilant with factor SPEAKER and dependent variable DCT1, DCT2 and DCT3.*

Sibilant	Parameter	F values, df and significance level
/ʃ/	DCT1	$F(13, 504) = 33.90$, $p < 0.001$
	DCT2	$F(13, 504) = 92.31$, $p < 0.001$
	DCT3	$F(13, 504) = 58.98$, $p < 0.001$
/s/	DCT1	$F(13, 448) = 38.18$, $p < 0.001$
	DCT2	$F(13, 448) = 46.12$, $p < 0.001$
	DCT3	$F(13, 448) = 28.16$, $p < 0.001$

Table C.6: *Mean values (and standard deviations) for DCT coefficients 1-3 for /s/, significant differences within twin pairs marked with * ($p < 0.01$) (calculated by a post hoc Tukey test).*

Twin	Mean DCT1 (SD) Twin A – Twin B	Mean DCT2 (SD) Twin A – Twin B	Mean DCT3 (SD) Twin A – Twin B
MZf1 (afhf)	-38.8 (25.7) / -47.9 (36.6)	-122.3 (17.4)/ -117.9 (16.9)	11.2 (14.2) / 12.6 (15.7)
MZf2 (gsrs)	-10.3 (28.6) / -91.6 (24.2)*	-128.7 (23.0) -108.7 (13.3)*	-3.0 (19.2) / -35.2 (14.2)*
MZm1 (slcl)	-5.7 (33.0) / 0.5 (27.7)	-149.4 (21.2) / -118.9 (22.3)*	8.1 (27.0) / -22.9 (11.5)*
MZm2 (mima)	8.5 (57.4) / -12.0 (52.5)	-71.8 (25.6) / -62.2 (30.3)	-32.9 (26.1) / 9.1 (24.5)*
DZf1 (srlr)	-65.9 (32.6) / -28.4 (39.2)*	-130.8 (30.6) / -98.3 (21.0)*	-2.9 (18.9) / 1.8 (16.1)
DZf2 (tgmg)	15.9 (48.9) / -82.8 (44.7)*	-65.4 (26.8) / -121.8 (25.4)*	-27.9 (20.6) / -21.1 (24.9)
DZm1 (fmhm)	-52.2 (28.6) / -128.6 (27.7)*	-81.4 (21.5) / -108.6 (17.9)	-16.3 (18.3) / -29.8 (14.9)

Table C.7: *Mean values (and standard deviations) for DCT coefficients 1-3 for /ʃ/, significant differences within twin pairs marked with * ($p < 0.01$) (calculated by a post hoc Tukey test).*

Twin	Mean DCT1 (SD) Twin A – Twin B	Mean DCT2 (SD) Twin A – Twin B	Mean DCT3 (SD) Twin A – Twin B
MZf1 (afhf)	-5.4 (20.0) / 6.3 (28.8)	-118.0 (21.2) / -125.5 (23.2)	-45.1 (13.9) / -30.8 (17.4)
MZf2 (gsrs)	28.4 (23.4) / -13.8 (18.7)*	-186.3 (15.3) / -170.4 (17.7)	-79.3 (18.3) / -75.9 (13.1)
MZm1 (slcl)	39.4 (20.8) / 46.8 (24.9)	-150.9 (17.2) / -131.7 (13.9)	-40.6 (18.8) / -61.3 (12.2)*
MZm2 (mima)	32.6 (43.0) / 40.2 (27.1)	-64.1 (20.3) / -82.5 (29.6)	-31.4 (20.8) / -46.1 (20.2)
DZf1 (srlr)	66.7 (39.3) / 66.1 (48.8)	-100.8 (28.1) / -111.9 (36.2)	-37.6 (17.8) / -2.4 (15.1)*
DZf2 (tgmg)	54.1 (21.9) / 16.8 (50.5)*	-97.3 (28.4) /-85.1 (26.0)	-72.4 (19.7) /-24.4 (15.2)*
DZm1 (fmhm)	27.6 (17.2) / 87.4 (19.5)*	-105.5 (27.6) / -59.1 (20.7)*	-42.3 (19.9) / -55.1 (15.7)

Significance levels: $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*'

Table C.8: *Results of ANOVAs for each sibilant and F2 and F3 with factor SPEAKER and dependent variable formant TARGET (Ftarget) or formant TRANSITION (Ftrans– Ftarget).*

Sibilant	Parameter	F values, df and significance level
/ʃ/	F2_target	F(13, 454) = 97.32, p < 0.001
	F2_transition	F(13, 454) = 7.53, p < 0.001
	F3_target	F(13, 454) = 75.34, p < 0.001
	F3_transition	F(13, 454) = 16.01, p < 0.001
/s/	F2_target	F(13, 182) = 33.23, p < 0.001
	F2_transition	F(13, 182) = 4.48, p < 0.001
	F3_target	F(13, 182) = 43.56, p < 0.001
	F3_transition	F1(13, 182) = 13.52, p < 0.001

Table C.9: *Results of post hoc Tukey tests for effect of SPEAKER on dependent variable formant TARGET or TRANSITION for /s/ and /ʃ/, Sign. levels: p < 0.001 '***', p < 0.01 '**', p < 0.05 '*'.*

Comparison	sibilant	Diff in F2_Target	Diff in F3_Target	Diff in F2_Transition	Diff in F3_Transition
MZf1a-MZf1b	/s/	65.45	113.78	75.25	73.59
MZf2a-MZf2b		-139.38	51.33	-104.06	10.52
MZm1a-MZm1b		75.56	303.38***	-86.13	-101.69
MZm2a-MZm2b		-6.14	3.99	10.61	115.93
DZf1a-DZf1b		-224.14***	-141.56	187.87*	183.56
DZf2a-DZf2b		-130.45	-215.79*	34.70	-57.811
DZm1a-DZm1b		-138.59	196.29	-87.09	-430.46***
MZf1a-MZf1b	/ʃ/	118.05***	202.61**	-58.91	-108.58
MZf2a-MZf2b		57.25	224.75**	-4.80	121.11
MZm1a-MZm1b		42.15	39.53	-70.88	-142.08
MZm2a-MZm2b		56.76	-433.03**	-152.15	165.23
DZf1a-DZf1b		-63.76	413.97**	161.90	-294.09*
DZf2a-DZf2b		59.615	-204.69*	234.85	452.89***
DZm1a-DZm1b		201.26**	-373.23**	-81.60	127.89

APPENDIX D

Statistics for Loops

Table D.1: *Number of compared records (N), mean Euclidean distances (Mean ED) and Standard Deviations (SD) for each speaker comparison.*

Group	Code	Speaker pair	Mean ED	SD	N
SSp	DZf1a	LRLR	0.026	0.018	28
SSp	DZf1b	SRSR	0.024	0.013	36
SSp	DZf2a	MGMG	0.086	0.079	45
SSp	DZf2b	TGTG	0.046	0.041	28
SSp	MZf1a	AFAF	0.032	0.022	45
SSp	MZf1b	HFHF	0.053	0.047	45
SSp	MZf2a	GSGS	0.014	0.012	45
SSp	MZf2b	RSRS	0.026	0.015	45
SSp	MZm1a	CLCL	0.137	0.136	45
SSp	MZm1b	SLSL	0.018	0.010	45
MZ	MZf1	AFHF	0.153	0.076	100
MZ	MZf2	GSRS	0.065	0.026	100
MZ	MZm1	CLSL	0.253	0.127	100
DZ	DZf1	LRSR	0.266	0.062	72
DZ	DZf2	MGTG	0.746	0.185	80
UN	UNf01	AFGS	0.378	0.089	100
UN	UNf02	AFSR	0.419	0.118	90
UN	UNf03	AFLR	0.181	0.074	80
UN	UNf04	AFRS	0.326	0.094	100
UN	UNf05	AFMG	1.835	0.351	100
UN	UNf06	AFTG	0.333	0.118	80

UN	UNf07	HFRS	0.392	0.149	100
UN	UNf08	HFSR	0.744	0.217	90
UN	UNf09	HFGS	0.349	0.097	100
UN	UNf10	HFRS	0.392	0.149	100
UN	UNf11	HFMG	2.568	0.492	100
UN	UNf12	HFTG	0.669	0.214	80
UN	UNf13	GSSR	0.628	0.102	90
UN	UNf14	GSLR	0.131	0.045	80
UN	UNf15	GSMG	1.799	0.358	100
UN	UNf16	GSTG	0.606	0.166	80
UN	UNf17	RSSR	0.489	0.093	90
UN	UNf18	RSLR	0.102	0.050	80
UN	UNf19	RSMG	1.481	0.347	100
UN	UNf20	RSTG	0.443	0.144	80
UN	UNf21	MGSR	0.670	0.169	90
UN	UNf22	MGLR	1.408	0.307	80
UN	UNf23	TGSR	0.051	0.028	72
UN	UNf24	TGLR	0.347	0.293	64

Table D.2: *Overview of the fit of the model and the random effects.*

Linear mixed model fit by REML

MODEL: lmer(log(distance) ~ group + (1|speak1) + (1|speak2))

AIC	BIC	logLik	deviance	REMLdev
5307	5349	-2646	5280	5293

Random effects:

Groups	Name	Variance	SD
speak1	(Intercept)	0.48955	0.69968
speak2	(Intercept)	0.36890	0.60737
Residual	(Intercept)	0.33228	0.57644

Number of obs: 2985, groups: speak1, 10; speak2, 10

APPENDIX E

Perception Experiment

Table E.1: *Stepwise regression to find the best fitted model.*

Models:

test0: correctness ~ group + (1 | listener)

test1: correctness ~ group + (1 | listener) + (1 | pair)

test2: correctness ~ group + (1 | listener) + (1 | pair) + (1 | stimulus)

	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)
test0	4	15503.5	15535.2	-7747.8			
test1	5	14766.3	14805.8	-7378.1	739.25	1	< 2.2e-16 ***
test2	6	14297.5	14344.9	-7142.7	470.82	1	< 2.2e-16 ***

Table E.2: *Overview of the fit of the model and the random effects.*

Generalized linear mixed model fit by Laplace approximation

correctness ~ group + (1 | listener) + (1 | pair) + (1 | stimulus)

AIC	BIC	logLik	deviance
14297	14345	-7143	14285

Random effects:

Groups	Name	Variance	SD
stimulus	(Intercept)	0.66057	0.81276
pair	(Intercept)	0.53921	0.73431
listener	(Intercept)	0.76764	0.87615

Number of obs: 20160, groups: stimulus, 1216; pair, 28; listener, 28

Selbstständigkeitserklärung zur Dissertation

Ich erkläre ausdrücklich, dass es sich bei der von mir eingereichten schriftlichen Arbeit mit dem Titel:

> *‚The Influence of NATURE and NURTURE on Speaker-Specific Parameters in Twins' Speech: Acoustics, Articulation and Perception'*

um eine von mir selbstständig und ohne fremde Hilfe verfasste Arbeit handelt.

Ich erkläre ausdrücklich, dass ich *sämtliche* in der oben genannten Arbeit verwendeten fremden Quellen, auch aus dem Internet (einschließlich Tabellen, Grafiken u. Ä.) als solche kenntlich gemacht habe. Insbesondere bestätige ich, dass ich ausnahmslos sowohl bei wörtlich übernommenen Aussagen bzw. unverändert übernommenen Tabellen, Grafiken u. Ä. (Zitaten) als auch bei in eigenen Worten wiedergegebenen Aussagen bzw. von mir abgewandelten Tabellen, Grafiken u. Ä. anderer Autorinnen und Autoren (Paraphrasen) die Quelle angegeben habe.

Mir ist bewusst, dass Verstöße gegen die Grundsätze der Selbstständigkeit als Täuschung betrachtet und entsprechend der Prüfungsordnung und/oder der Allgemeinen Satzung für Studien- und Prüfungsangelegenheiten der HU (ASSP) geahndet werden.

Datum …….. Unterschrift