

Melanie Weirich

Friedrich-Schiller-Universität Jena

Organic Sources of Inter-Speaker Variability in Articulation: Insights from Twin Studies and Male and Female Speech

Abstract: This chapter presents three studies dealing with articulatory inter-speaker variability in German. In particular, organic sources (such as biomechanics of the tongue muscles, palatal shape and vocal tract dimensions) of idiosyncratic variation are discussed. Two studies deal with the within-pair similarity of identical (monozygotic) and non-identical (dizygotic) twin pairs; the third study describes differences between male and female speakers. The speech material comprises looping movements of the tongue in /aCV/-sequences, the production of the sibilant contrast /s/-/ʃ/ and the tense vowels /i: e: a: o: u:/ in different accent conditions. Results show that individual differences in articulatory strategies can at least in part be explained by idiosyncratic physiological restrictions and that the investigation of phonemic contrasts instead of targets, and the emphasis on speech dynamics are particularly relevant.

1. What we can learn from variability in speech

Research within the framework of speech perception has long dealt with the question of *invariance* in the speech signal. Possible invariant correlates of the speech production task in the physical space have been claimed to exist in various dimensions including articulation, acoustics and neural patterns (*Acoustic Invariance Theory*, Stevens and Blumstein, 1978; *Adaptive Variability Theory*, Lindblom 1988, 1990; *Motor Theory* Liberman et al., 1967; Liberman and Mattingly, 1985). However, we know that speech perception is multimodal and the articulatory movements and the acoustic signal are taken into account when both modalities are available (e.g. McGurk and MacDonald, 1976). In addition, numerous studies investigating intra- and inter-speaker variability – both in acoustic and articulatory terms – show that no true invariance exists and various combinations of physical correlates in both the acoustic and articulatory domain are present. Moreover,

the variability found is not random, and should not be considered as distracting noise. Rather, we should consider it to be highly informative, telling us something about the speaker (or the respective speaker group), comprising both physiologically based restrictions as well as learned speech behavior (Foulkes and Docherty, 2006). In other words, a main question we are dealing with is, which variability is not just noise but systematic and is explainable due to which classifiable factors? Most generally, these factors can be separated into two potential sources, i. e. organic and learned (Ladefoged and Broadbent, 1957), and thus can be discussed within the *nature-nurture* framework. Of course, in most cases just one of these factors is never sufficient to fully explain the variability found, but it might be that sometimes one of the factors outweighs the other. The question is: When? And why? If we understand the reasons (when is which factor more important) we can learn something essential about the functioning of the speech production process.

The aim of studying variability is thus not to describe speaker-specific behavior per se but to determine particular groups of speakers that show the same “speaker-specific” behavior or strategies and to relate this variability with particular factors that classify the respective groups, arising from different biological, social or cultural sources. The studies described in this chapter focus on inter-speaker variability that is due to biological/organic variability. Biological similarity is present in related speakers, and in the most extreme way, in *twins*, the speaker group investigated in the first two studies of section 3. Biological similarity is also a strong factor when sex-specific differences are concerned and section 4 deals with articulatory differences between male and female speakers.

2. Learned vs. organic sources of inter-speaker variability

From psychological theories of learning, e.g., *Social Learning Theory* (Bandura, 1977) we know that people in general learn by observing and mimicking. Regarding 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 around them (i. e. especially in the beginning, mothers). Also dialectal pronunciation and sociolinguistic parameters of the parents are observed and absorbed by the

child (Chambers, 2003). Moreover, this learning is a life-long process as has been shown very effectively in the analysis of pronunciation changes in the Christmas broadcasts of Queen Elisabeth II over a span of 40 years (Harrington, 2006). Sociolinguistic studies in general have shown that inter-speaker variation has numerous behavioral sources and can be used to create, express and attribute a certain social identity (for an overview see Foulkes and Docherty, 2006).

Nevertheless, organic sources of speaker-specific articulation exist which constrain the degrees of freedom a speaker has. Lindblom (1983, p. 217) assumes in his theory on the economy of speech gestures that “languages tend to evolve sound patterns that can be seen as adaptations to biological constraints of speech production.” These biological constraints are manifold and comprise the length and constitution of the vocal folds, the size and dimensions of the vocal tract, the functioning of the tongue muscles, the shape of the palate and also the teeth. All of these organic factors can differ to some degree between speakers (or speaker groups such as adults and children or male and female speakers) and thus influence variation in articulatory strategies. Speaker groups (e.g. males vs. females, adults vs. children) differ in formant patterns due to biologically determined differences in the individual cross-sectional area of the vocal tract, with children showing the highest formant frequencies and males the lowest (Fant, 1960). Other studies have found a relationship between vocal tract geometry and articulatory space (Winkler et al., 2006; Fuchs et al., 2008). In particular, the individual articulatory distances between corner vowels (investigated using MRI in 9 French speakers) depended on the length of the speakers’ pharynx: speakers with longer pharynxes showed larger degrees of freedom in the vertical direction and had larger vertical displacements than speakers with shorter pharynxes. Vocal tract size and dimension is the biological factor discussed in section 4 on sex-specific differences in articulatory spaces.

Several studies have emphasized the significant role of palate shape in articulatory variability (Lammert et al., 2013; Rudy and Yunusova, 2013; Brunner et al., 2009; Fuchs et al., 2006). For example, Brunner et al. (2009), found a relationship between a speaker’s variability in tongue height and the steepness of the palate. Speakers with flat palates were more constrained in their variability than speakers with domed palates. The authors suggest that this is due to the large consequences on the area function/the acoustic

output that small variation in tongue position can have in speakers with flat palates. Rudy and Yunusova (2013) showed that palate curvature and length can at least in part explain tongue position variability in the production of front consonants. They investigated VCV-sequences with C including stops, fricatives and affricates in 21 speakers of Canadian English. Lammert et al. (2013) investigated the interplay of hard palate morphology, articulation and acoustics in real vowel production data (MRI, five speakers) and in simulations. While simulations showed that palatal morphology affects formant frequencies, no significant correlation was found between real formant data and lingual articulation, leading the authors to conclude that speakers adapt their articulation strategies to accommodate palate shape differences. Palate shape as a potential organic source of inter-speaker articulatory variability is the factor investigated in the second twin study of section 3.

All three studies presented in this chapter concentrate on lingual inter-speaker variability that might be explained by organic sources, in particular, the palate shape (section 3, second twin study) and vocal tract dimensions (section 4, sex-specific differences in articulatory spaces). In addition, the first twin study of section 3 examines looping movements of the tongue during VCV-sequences in identical and non-identical twin pairs. By looking at the whole movement or gesture of the sequence, the influence of the tongue muscles, vocal tract dimensions and palate morphology is taken into account. It should be noted that in this chapter the term *gesture* is not used in the sense of an abstract idea (following Browman and Goldstein, 1992) but as a concrete movement of specific articulators.

3. Speaker-specific articulation in twins' speech

To investigate individual differences and to explain the variation in terms of the two possible influencing factors nature (i. e., genes and physiology) and nurture (i. e., environmental factors), a standard procedure in the field of behavioral genetic research is conducting twin studies (Spinath, 2005). Twin studies comprise a systematic comparison of the within-pair similarity of monozygotic (MZ) twins (who are 100% genetically identical) and dizygotic (DZ) twins (who share only about 50% of their genes, same as normal siblings). That anatomical and physiological characteristics are genetically

determined and more similar in MZ twin pairs than in DZ twin pairs has been shown in several medical and dental studies, e.g. regarding the size and position of the jaw, the tooth size and the occlusal morphology (Lundström, 1948; Kabban et al., 2001) but also regarding the thyroid volumes (Langer et al., 1999). Eguchi et al. (2004) found in their comprehensive study of 78 male and female MZ and DZ twin pairs a high genetic contribution to speaker-specific variation in dental arch width, length, and also palatal height. While the twin types differ in their genetic/physiological similarity, they do not differ in terms of social environmental factors that contribute to the resemblance between individuals who grow up in the same family. If no particular emphasis by the parents is laid on treating the twins differently, they go to the same school, share most of their friends and also hobbies. In addition, for both twin types, the siblings have the same age at the same time, thereby being influenced by historical events in a similar way (*Equal Environments Assumption*, Scarr and Carter-Saltzman, 1979). Regarding language, both twin types share a) their environment during the speech acquisition process and b) social factors (such as school, hobbies and peer groups) which influence the speech of an individual. Thus, by comparing the within-pair similarity between MZ and DZ twins (who have grown up together, shared their speech acquisition process, and have a history that is not significant for differences in external factors such as surgeries, accidents, drug abuse or even the use of a pacifier that affects palatal shape during maturation), the role of physiological determinants and inherited morphological parameters can be analyzed. In other words, if MZ twins are more similar than DZ twins in a particular parameter, this parameter is affected by organic (genetic) factors.

While twin studies have a long tradition in the field of behavioral genetics research and go back to the late 19th century (Sir Frances Galton, 1876), analyses of twins' speech is rather new. Several studies have investigated speech acquisition and speech pathology in twins (Locke and Mather, 1989; Ooki, 2005; Simberg et al., 2009). However, only few have examined inter-twin variability in normal speech, and here, perceived similarity and acoustic features have been the predominant topics (see Loakes, 2006 and Weirich, 2012 for a more comprehensive overview of these studies). In summary, MZ twins have been found to be more similar than same-sex DZ twins or age-matched siblings in their average fundamental frequency

(Przybyla et al., 1992; Debruyne et al., 2002), voice quality parameters (van Lierde et al., 2005) and coarticulatory/dynamic patterns (Nolan and Oh, 1996; Whiteside and Rixon, 2003; Weirich, 2012). While perception studies have shown that familiar listeners can distinguish MZ twins (Whiteside and Rixon, 2000), unfamiliar listeners succeed in distinguishing unrelated speakers by using only one short bi-syllabic word but fail to do so in both MZ and DZ twin pairs (Weirich and Lancia, 2011).

Articulatory studies in twins have rather been neglected (but see Weirich, 2012). A reason for the uncommonness of articulatory studies in twins might be a methodological one: articulatory analyses in general involve only small subject groups due to their time-consuming character, and together with the fact that in twin studies we usually compare the similarity of speaker pairs (i. e., MZ twin pairs vs. DZ twin pairs), the problem of a small subject group becomes even more crucial. Moreover, the participating twins have to fulfill several requirements concerning environmental factors such as the time they have spent (and are still spending) together, surgical interventions and also the attitude they have towards being a twin (a negative attitude could lead to an enhancement of individuality also expressed in an idiosyncratic speech style).

However, despite these difficulties, twin design studies have a high potential for helping us to distinguish physiological determinants and environmental factors responsible for individual differences in speech. The impact of physiological factors is especially relevant with regard to speaker-specific articulation strategies. Thus, if we are interested in understanding the reasons for inter-speaker variability, analyzing articulatory variability in MZ and DZ twin pairs is a promising source of information. Therefore, one of the two main aspects of this chapter is the discussion of two recent studies that we have conducted on the speech of MZ and DZ twins, concentrating on within-pair articulatory variability in VCV sequences (Weirich et al., 2013) and in the realization of the sibilant contrast /s-/ʃ/ (Weirich and Fuchs, 2013).

3.1. Individual articulatory strategies in looping movements

The first study presented here on articulatory variability in twins' speech is on a particularly interesting articulatory gesture: the looping movement

of the tongue (for a more detailed description of the study see Weirich et al., 2013). Loops are curved trajectories of the tongue back found in VCV-sequences where C is a velar consonant. The trajectories of the sequence do not simply consist of straight lines between vowel and consonant targets but an elliptical movement of the tongue back – a loop – is found (Kent and Moll, 1972, Mooshammer et al., 1995; Hoole et al., 1998; Löfqvist and Gracco, 2002; Geng et al., 2003; Perrier et al., 2003; Brunner et al., 2011). Curved paths in movements in general have been shown to be potentially explained by anatomical factors and muscle mechanics (see Flanagan et al., 1993; Gribble and Ostry, 1996; Gribble et al., 1998 for arm movement, Perrier et al., 2003; Perrier and Fuchs, 2008 for orofacial movements). Thus, it is argued that the loops of the tongue are also a result of the biomechanical characteristics of the muscles and the surrounding vowel targets (Perrier et al., 2003). From that we can hypothesize that loops should be more similar in MZ twins than in DZ twins or unrelated speakers. If, on the other hand, loops are actively controlled (Löfqvist and Gracco, 2002) and reflect learned behavior independent of individual physiology, the degree of variability within a twin pair should depend less on the twin type (MZ vs. DZ).

3.1.1. *Articulatory analysis*

3.1.1.1. *Participants*

The participants were ten German speakers (20–34 years old): two female DZ twin pairs and two female and one male MZ twin pair. All speakers were born, raised and still living in Berlin, Germany. The twins grew up together and were still seeing each other at least twice a month. With a comprehensive questionnaire we controlled for differences in potential influencing factors such as relevant surgeries, habits or attitudes towards being a twin, but also the behavior of the parents in raising their children was checked (e.g. treating them particularly different or making them like the same things). All of the participants liked being a twin and no pair differed with respect to surgeries, accidents or habits (e.g. use of pacifier as a child, singing, smoking) that might have affected physiological characteristics of the speech apparatus. Also, all twins reported being treated similarly by the parents and having shared friends and hobbies especially during childhood and adolescence.

In addition to the comparison of speakers within the same twin pair (groups MZ and DZ), speakers of different twin pairs were paired to form the group of unrelated (sex-matched) speakers (group UN).

3.1.1.2 Recordings, speech material and measurements

Acoustic and articulatory recordings were conducted in the speech lab of ZAS (Zentrum für Allgemeine Sprachwissenschaft, Berlin) by means of 2D electromagnetic articulography (EMA, Carstens AG100). Two coils, one above the upper incisors and one at the bridge of the nose served as reference coils and were used for head movement correction. Three coils were attached to the tongue (one approximately 0.5 cm behind the tongue tip, one approximately 5 cm behind the tip on the tongue back, and a third one halfway in-between the two, on the tongue dorsum). Comparable positioning of the coils within the twin pairs was attained using a true-to-scale template of the tongue with the coils of one of the twins (created with the help of a printed photograph) being used as a reference for the second twin. For the analysis of the looping movement, we concentrated on the coil positioned furthest back on the tongue (henceforth, tongue back coil).

The speech material was obtained during a larger recording session with different target phonemes and carrier sentences (see Weirich, 2012). For the analysis of the looping pattern of the tongue back the sequence /aCV/ within the names “Haga”, “Hagu”, “Haka”, and “Haku” was chosen. The target words were part of the sentence “Ich grüße/wasche Haka/Haga/Haku/Hagu im Garten” (I greet/wash Haka/Haga/Haku/Hagu in the garden). On average 9.45 repetitions for each speaker and each /aCV/ sequence could be used for the analysis.

For a comparison of the shape of the looping patterns between speakers, including all repetitions, the data had to be processed in several ways. Briefly, first, the shape of the looping movement had to be parameterized. Therefore, instead of taking absolute positions of the tongue back coil (in vertical and horizontal dimension), curvature was calculated for each measurement point throughout the /aCV/-sequence (cf. Tasko and Westbury, 2004; O’Neill, 2006). This was also done to prevent a confound of the potentially more similar coil positions in MZ twins than in DZ twins or unrelated speaker pairs. Second, multiple pairwise comparisons were done (separately for each /aCV/ sequence) consisting of either trajectories from

two speakers of the same twin pair (MZ or DZ) or trajectories from two unrelated (sex matched) speakers (e. g. twin 1 from twin pair A and twin 1 from twin pair B). Third, the articulatory trajectories had to be temporally normalized. For this reason we adopted the functional data analysis tool proposed by Lucero et al. (1997) and used a registration method described in detail in Lancia and Tiede (2012) to get time-aligned trajectories for each /aCV/-sequence and each speaker pair separately. Fourth, distances were measured between all points of each pair of aligned curvature data and mean distances were calculated for each comparison to be used for the statistical analysis. For further details on the recording session, the labeling procedure and the different processing steps see Weirich et al. (2013).

3.1.2. *The impact of physiology on looping movements*

For purposes of exemplification Figure 1 shows looping trajectories during /aka/ for two speakers (in grey and black) of a female MZ twin pair (left) and two speakers (grey and black) of a female DZ twin pair (right). The plots show the positional data (in horizontal and vertical dimensions) of the average tongue movement (without time-alignment), the arrow marks the direction of movement. The MZ twins in the left plot exhibit similar shapes of their loops, starting with a rather straight/slightly curved upwards movement from the vowel to the velar stop, a horizontal – possibly sliding – movement along the palate and a straight or slightly backwards oriented downwards movement to the second /a/. The loops of the DZ twins in the right plot show more obvious differences in shape and looping characteristics. While speaker DZa (grey) resembles the MZ speakers in terms of a slightly forward directed upwards lifting, horizontal movement (even though to a lesser extent) and a downward and backward movement to the second /a/, DZb (black) shows a more s-like shape of the tongue lifting, no sliding along the palate with a very steep angle at the turning point and a backwards directed movement to the second vowel. This pair also differs with respect to the relative positions of the two vowels: while DZa (grey) produces V2 at a more fronted position than V1 (as the two speakers of the MZ pair), speaker DZb (black) produces V2 at a more retracted position than V1. Note that for the statistical analysis, curvature and not positional data was used. A high value in curvature corresponds to

a change in movement direction, while low curvature values reflect rather straight movements. Thus, DZb is the speaker revealing the clearest peak in curvature over the whole movement.

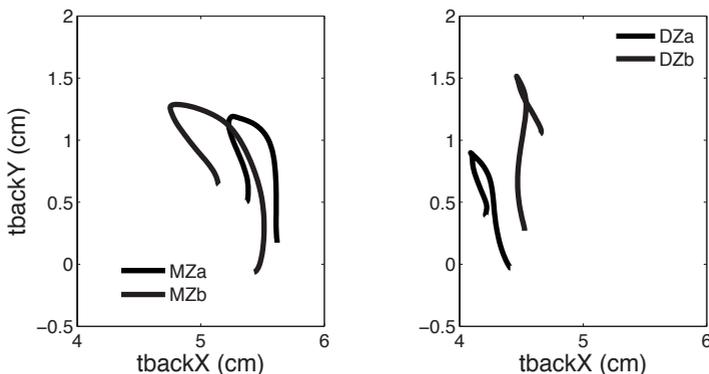


Figure 1: Mean looping trajectories of the tongue back coil during /aka/ for two twin pairs (left MZ, right DZ). Different speakers of each pair are indicated by grey and black. Vertical movement of the tongue back is displayed on the y-axis, horizontal movement on the x-axis (in cm). Reprinted from *Journal of Acoustical Society of America* 134, 5, 3766–3780. Weirich, M., Lancia, L., Brunner, J. *Interspeaker articulatory variability during vowel-consonant-vowel sequences in twins and unrelated speakers*. Reproduced with permission from AIP Publishing LLC. Copyright 2013.

For the statistical analysis a linear mixed model (Pinheiro and Bates, 2000) as implemented in the lme4 package of the R software (version 2.14.1, R Development Core Team, 2008) was run. The dependent variable was the measured pairwise mean absolute distance between the aligned curvature data. The logarithmic values of these distances were used to normalize the residuals, a mandatory assumption in linear mixed models (Pinheiro and Bates, 2000). We included speaker group (with the levels MZ, DZ and UN = unrelated speakers), vowel (/a/ vs. /u/) and voice (voiced vs. voiceless) as fixed factors and a pair specific random intercept for vowel and voice.

Figure 2 shows the distribution of the log transformed distance measures for all /aCV/-sequences together but separated by speaker group. As is apparent in the figure, the statistical analysis revealed a significant difference between MZ twins and DZ twins ($p_{\text{MCMC}} < 0.001$), but not between DZ twins and unrelated speaker pairs. No interaction between vowel and

speaker group or voice and speaker group was found but a three way interaction between all factors suggested that for the vowel /a/ a stronger effect of the comparison between MZ and DZ twins exists in the voiced than in the voiceless condition.

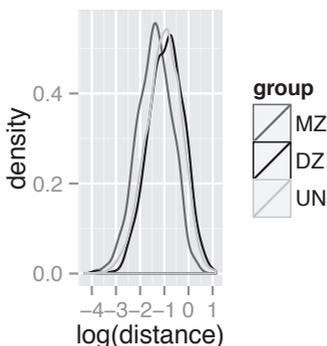


Figure 2: Distribution of logarithmic distances in curvature for all /aCV/-sequences separated by the three groups: monozygotic twins (MZ), dizygotic twins (DZ) and unrelated speakers (UN).

3.1.3. Discussion

The results of the study reveal a significant influence of shared physiology on articulatory inter-speaker variability. In detail, more similar looping patterns in VCV sequences were found for MZ than for DZ twins or unrelated speakers. By investigating articulatory movements (such as loops) and not articulatory target positions the focus has moved from static to dynamic aspects of the speech signal. Particularly regarding inter-speaker variability this might be an essential factor. Nolan et al. (2006) suggested that the speech signal can be described by two different aspects: 1) linguistically determined targets and 2) organically determined transitions. It is proposed that while the targets are constrained by the shared language system and carry the linguistic information, the transitions link the adjacent targets and are more prone to reflect speaker-specific characteristics that are due to individual physiology.

Recently, studies on inter-speaker variability have focused not only on phonemic targets (or transitions) but also on the realization of phonemic *contrasts*. In this way, the phonetic inventory of a language is better reflected and taken into account. The next twin study deals with the realization of the sibilant contrast /s-/ʃ/ in German.

3.2. Individual articulatory strategies in realizing the sibilant contrast

Toda (2006) found two different speaker-dependent strategies in the realization of the sibilant contrast /s/ - /ʃ/ in French: 1) tongue placement strategy (where speakers only retract their tongue horizontally) and 2) tongue adjustment strategy (where speakers additionally elevate their tongue). With respect to sibilants, some of the most important work in recent years has been conducted by Perkell and colleagues (Perkell, 2010; Perkell et al., 2004). Their work particularly emphasizes the link between speech production and perception. In other words, they find that speakers with poorer auditory acuity of a phonemic contrast also tend to produce this contrast less distinctively. Ghosh et al. (2010) went one step further by including a speaker's somatosensory acuity (which implies the sensation of touch, i. e. tactile feedback) into the analysis of the acoustic realization of /s/ and /ʃ/. They found a positive correlation between a speaker's acoustic distance between the sibilants and their auditory and somatosensory acuities. When tactile feedback plays a role in the realization of a phonemic contrast, as Ghosh and colleagues found, then individual differences in the respective morphological structures relevant for the sound production (i. e. the palatal shape) might also affect the realization of this contrast. Perkell et al. (2004) included some morphological parameters in their analysis of the sibilant contrast. They examined palatal height, length and width but could not find any significant correlations. They did, however, not include a parameter that is essential for the production of sibilants: the palatal and in particular the alveolo-palatal steepness. Thus, in our study (Weirich and Fuchs, 2013) we investigated the potential relationship between speaker-specific realizations of the /s/-/ʃ/ contrast in German and the speaker's palatal shape, parameterized by two angles describing the overall steepness of the palate and the steepness of the alveolo-palatal ridge, where the contrast is realized.

3.2.1. *Articulatory analysis*

3.2.1.1. *Participants*

The study consisted of two different experiments (EMA and EPG) with different speaker samples. The EMA study comprised the same DZ and MZ

pairs from the VCV study (4 female and 1 male pair). In addition, another male MZ pair, part of the twin corpus recorded at the ZAS (Weirich, 2012) could be included. The EPG experiment comprised 12 unrelated German speakers (7 females and 5 males) with no hearing or speech impairments, aged between 24 and 56.

3.2.1.2. Recordings, speech material and measurements

The speech material of the EMA experiment was acquired during the larger recording session of the Weirich (2012) study. The target sounds were the sibilants /s/ and /ʃ/ that were part of the German verbs /kʏsə/ (1. p. sg. of ‘to kiss’) and /vaʃə/ (1. p. sg. of ‘to wash’) embedded in carrier sentences. On average, 32 repetitions for each speaker and phoneme were included. The target positions of /s/ and /ʃ/ were labeled oriented on the minimal tangential velocity of the tongue tip sensor. We then investigated inter-speaker variability in realizing the contrast in terms of the horizontal and vertical position of the tongue tip following Toda’s (2006) idea of two different speaker specific strategies which vary in the amount of vertical tongue elevation. While we cannot compare the whole overall shape of the tongue, as Toda did, due to the use of EMA-data which gives us information only about the position of three flesh points on the tongue, we can compare the position of the tongue tip between the two sibilants for each speaker and thereby investigate the vertical/horizontal distance between the sound productions.

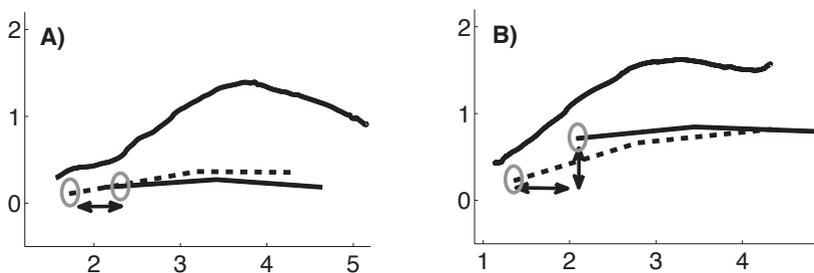


Figure 3: Distance measurement (in horizontal and vertical dimensions) between mean tongue tip positions (dashed line = /s/, solid line = /ʃ/) for two speakers of different twin pairs. Reprinted from *Journal of Speech, Language, and Hearing Research* 56, 8, 1894–1190, Weirich, M. and Fuchs, S. *Palatal morphology can influence speaker-specific realizations of phonemic contrasts*. Reproduced with permission from the American Speech-Language-Hearing Association (<http://jslhr.pubs.asha.org>). Copyright 2013.

Figure 3 shows two strategies in two of our participants and visualizes their mean interpolated tongue contours during their articulatory target positions for /s/ (black line) and /ʃ/ (dashed line): while speaker A only retracts the tongue for /ʃ/ in contrast to /s/, speaker B retracts and additionally elevates the tongue, following the palate contour. To quantify this, for each speaker the horizontal and vertical distances between the tongue tip positions of the two sounds were summed up to 100%. The horizontal and vertical distance was then expressed in percentages, too, in relation to the total amount.

For the EPG experiment, the sibilants /s/ and /ʃ/ were recorded on average 30 times per speaker, they occurred in the nonsense words /zasa/ and /ʃaʃa/ and were embedded in a carrier sentence. We defined the place of articulation for the two sounds using the articulatory center of gravity (COG, Hardcastle et al., 1991), which is a weighted index that attaches more importance to rows at the front of the palate. A higher COG thus reflects a more anterior place of articulation (typical for /s/). The differences in COG between the /s/ and /ʃ/ productions of each speaker were then calculated to get a distance measure comparable to the one used in the EMA study.

Physiological measures were taken regarding body size, body weight and tongue (for the twin study) and the palate (for both studies). The different measurements were taken to look for their potential impact on sibilant production but also to confirm the assumption that physiological parameters were more similar in the MZ than in the DZ twins (for further information, see Weirich and Fuchs, 2013). Here, we will concentrate on the most crucial parameter in sibilant production: the palatal shape. The palatal shape was parameterized by two different angles: the angle of the overall palatal steepness (δ), and the angle of the alveolo-palatal ridge (γ) shown in Figure 4.

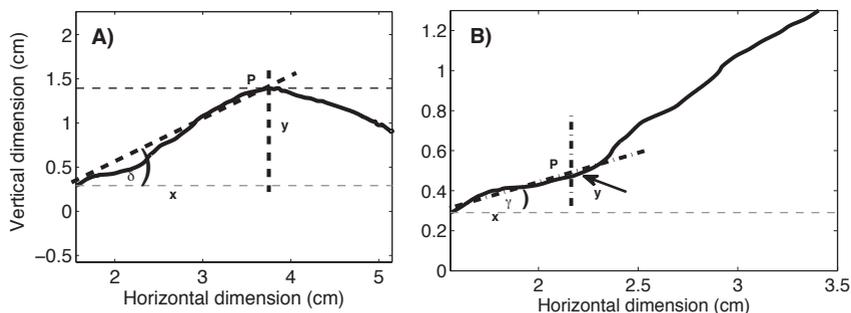


Figure 4: Visualization of angle measurements: angle of general palatal steepness δ (A) and angle of alveolo-palatal ridge γ (B, close up view). The thick black line shows the palate contour, the thinner dashed horizontal lines the minimal and maximal vertical positions of the palate. P defines the highest point of the palate (in A) or the alveolar step (in B, see arrow) with corresponding vertical (y) and horizontal (x) interval. Reprinted from *Journal of Speech, Language, and Hearing Research* 56, 8, 1894–1190, Weirich, M. and Fuchs, S. Palatal morphology can influence speaker-specific realizations of phonemic contrasts. Reproduced with permission from the American Speech-Language-Hearing Association (<http://jslhr.pubs.asha.org>). Copyright 2013.

The calculation of these angles was done in the same way for both experiments and is expressed in equation (1):

$$\tan(\delta, \gamma) = y(P) / x(P), \quad (1)$$

where P is the point on the palate that determines the height, y , and the length, x , necessary to calculate the particular angle. P differs for the two angles and reflects either the maximal vertical point on the contour for the angle δ (see plot A in Figure 4) or the visually defined position of the alveolo-palatal ridge for the angle γ (in most cases easily identifiable by a small dip as seen in plot B of Figure 4).

To look for a potential relationship between morphology and articulation, correlations were run between the two palate angles and the horizontal distance between the sibilants (in % for the EMA study or expressed as COG difference for the EPG study).

3.2.2. *The impact of palatal shape on articulatory realization of sibilant contrast*

The first main result of the EMA twin study was that we found more similar articulatory strategies in MZ twins than in DZ twins. This reflects the findings of the looping study shown above. While no difference between any of the four MZ twins was found (Welch two sample t-tests), both DZ pairs revealed significant differences ($p < 0.01$) in terms of their horizontal tongue tip variation (in %) between the two sounds.

The second main result was that we found a clear effect of individual palatal shapes on the articulatory realization of the sibilant contrast. For the twin study both angles revealed a significant negative correlation (Spearman) with the horizontal distance measure (in %): while the overall palatal steepness angle δ showed a correlation of -0.53 ($p < 0.05$), the correlation was even higher for the alveolo-palatal angle γ (-0.78 , $p < 0.01$). The relationship of the latter angle to the articulatory realization is shown in Figure 5 (left plot). The smaller the angle (the flatter the palate) and the more horizontal distance (in %) is found. The figure also reveals the more similar articulation for the MZ twins (marked by the filled symbols) than the DZ twins (unfilled symbols).

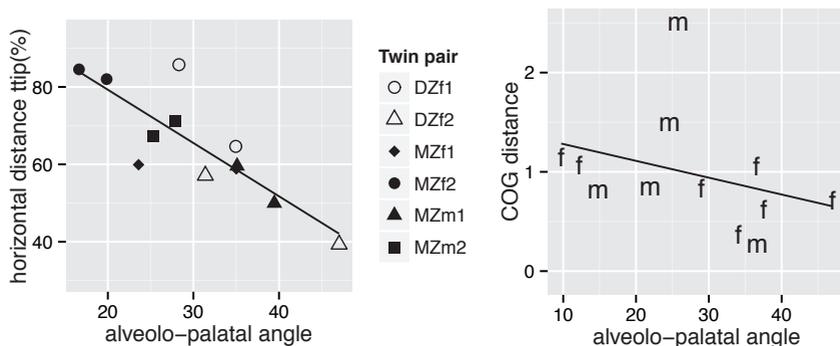


Figure 5: Relationship between alveolo-palatal angle (x-axis) and articulatory realization of the sibilant contrast (y-axis). The plot on left side shows the EMA-twin-data, the plot on the right side the EPG-data (m: male speaker, f: female speaker). The black line shows the regression line and the gray shadowed area defines the 95% confidence interval.

The remaining question then was whether we can find this influence of palatal shape also in a more heterogeneous group of unrelated speakers. Figure 5 (right plot) shows the COG distance measure of the EPG study in dependence of the alveolo-palatal angle. Despite one outlier (a male speaker with an extremely high COG distance measure) a significant negative correlation of -0.62 ($p < 0.05$) was found, mirroring the results of the twin data.

3.2.3. Discussion

The study revealed that not only does individual physiology play a role in dynamic aspects of articulation (such as loops), but also in the realization of a phoneme contrast. The articulatory organization of a speaker's targets is affected by his/her speaker-specific organically determined idiosyncrasies. Especially in sibilants, where the tongue-palate contact is crucial, these individual physiological characteristics come to the fore and show their impact. In detail, the shape of the alveolar ridge – which is the articulatory place where the sibilants are produced – can account for at least some of the inter-speaker variability found in the articulation of sibilants.

The question arises whether other phonemes that are less affected by physiological restrictions such as vowels might also be affected by individual differences in vocal tract anatomy. The final study focuses on inter-speaker variability in articulatory vowel spaces. Here, the speaker groups under investigation are male and female speakers, which have been found to differ in the physiological characteristics essential for the production of vowels (such as the overall vocal tract size, and the relationship between oral and pharyngeal cavity dimensions).

4. Speaker-specific articulation in male and female speech

Most studies on differences between male and female speech have concentrated on acoustic differences, fewer have investigated potential articulatory variability. A very salient and highly investigated aspect is the larger acoustic vowel space in females. It has been found for several languages, such as American English (Diehl et al., 1996), British English (Whiteside, 2001), German (Weirich and Simpson, 2014a) and Swedish (Simpson and Ericsson, 2007). The differences between vowel spaces are not uniform, with differences between different vowel categories increasing as F1 and

F2 increase. Thus, male and female speakers differ most in front and low vowels (such as /i:/ and /a/) and less in high back vowels (such as /u:/) (Fant 1966). Various hypotheses have been proposed to account for this variability. While some focus on purely behavioral reasons, such as the sociophonetic explanation of females aiming at speaking more clearly than males (Bladon et al., 1983; Henton, 1995), others emphasize physiological (sex-related) differences. One of the latter is the non-uniform difference between males and females in the relationship of pharyngeal and oral cavity (Chiba and Kajiyama, 1941; Fant, 1966, 1975; Nordström, 1977; Winkler et al., 2006; Fuchs et al., 2008).

A third strand of possible explanations is based on acoustic-perceptual compensation (Goldstein, 1980; Ryalls and Lieberman, 1982; Diehl et al., 1996). The reasoning is as follows: The higher the fundamental frequency, the sparser the harmonics. The greater inter-harmonic spacing in higher pitched voices causes a poorer definition of the spectral envelope (and in particular of the formants). From that it is hypothesized that the larger acoustic distance between female vowel targets compensates for the poorer spectral definition more typically found in high-pitched female voices. However, in a recent study of 56 female speakers with varying fundamental frequency (from 154 Hz to 234 Hz), we did not find a correlation between f_0 and acoustic vowel space size (Weirich and Simpson, 2013) suggesting other factors (organic and/or learned) must be responsible for the larger female acoustic vowel space.

Another explanation involves the underlying articulatory dynamics in producing the vowel space. Despite females having, on average, larger acoustic vowel spaces than males, Simpson (2001, 2002) found smaller articulatory vowel spaces in females than in males. In addition, Simpson (1998) found sex-specific differences in the relationship between formant values and duration (some expected correlations were only found for males but not females). Due to females exhibiting on average smaller vocal tracts than males they reach their articulatory targets earlier (in terms of time and space), and thus, might undershoot their targets less than males. Vowel undershoot can result from different degrees of coarticulation possibly induced by varying accent and stress conditions. Lindblom (1983, 1990) suggested in his H&H theory that speech varies along a continuum between output-oriented, hyperarticulated stressed syllables

at one end and system-oriented, reduced/hypoarticulated unstressed syllables at the other end. Since then the relationship between stressed and unstressed syllables and hyper- and hypoarticulation has been investigated intensively (e.g. de Jong et al., 1993; de Jong, 1995, 1998; Harrington et al., 2000; Cho, 2004). Mooshammer and Geng (2008) investigated articulatory manifestations of vowel reductions in German and found a greater degree of coarticulation with the consonant context in unstressed vowels than in stressed vowels. If females reach their articulatory targets earlier/more often than males (e.g. even in unstressed vowels), then they should be less influenced by accent-induced undershoot. If no differences in undershoot were found between the sexes, we would expect to find higher velocities or longer durations in males, but this is not the focus of the present investigation.

To test this assumption we conducted an articulatory analysis of 4 female and 5 male German speakers including speech material suitable to investigate a speaker's "extreme" articulatory vowel space only minimally affected by coarticulation and accent-induced undershoot. The aim was then to use this as a speaker-specific articulatory reference frame that all further analyses could be compared to (Weirich and Simpson, 2014b).

4.1. Articulatory analysis

4.1.1. *Participants and recordings*

Five male and four female German speakers took part in the study. The speakers were between 23 and 43 years old and came from the Eastern Central German dialect area but showing very little dialectal influence. Articulatory recordings were made at Potsdam University with the NDI-Wave system. Parallel to the twin studies, three coils were attached to the tongue and, for the present analysis, the movement of the coil positioned furthest back on the tongue (tongue back coil) was investigated. The articulatory labeling was done with the help of the MATLAB based software *mview* developed by Mark Tiede (Haskins Laboratories).

4.1.2. *Speech material*

The speech material was part of a larger corpus comprising 20 different target words (approximately 10 repetitions each) in different accent conditions

and varying carrier sentences. The data presented here is twofold: The first set of data included the three corner vowels /a: u: i:/ contained in the double vowel sequences in the abbreviations *IAA*, *AUU* and *BII*. The abbreviations were used because here the articulatory positions of the vowels were expected to be extreme and only minimally effected by coarticulatory influences. The second set of data included the sequence /gV/ with V being /i: e: a: o: u:/ in the German name *GVbi* embedded in the carrier sentence *Ich sah GVbi an* ('I looked at *GVbi*'). Here, three different accent conditions were recorded. First, the participants were asked to read the sentences presented to them from a screen (control condition, c). Second, participants produced the sentences in response to questions from the experimenter eliciting an answer with either the name under focus (accented condition, a) or the preceding verb (unaccented condition, u).

4.1.3. Sex-specific 'extreme' vowel spaces in IAU-polygon

Figure 6 shows the articulatory positions of the tongue back coil at the vowel targets /i: a: u:/ measured at the midpoint of the double vowel sequences of the abbreviations. The data was translated for each speaker with the midpoint of the vowel space set to the origin (0/0). This facilitated a better visual comparison between the sexes. The displayed data includes all repetitions of all male (black) and female (grey) speakers. As we can see, there is a tendency for male speakers to exhibit larger articulatory spaces than females (on average 93 mm² vs. 66 mm²), and this variability is due to a vowel-specific difference: while the articulatory positions completely overlap for /u:/, males exhibit a lower and more retracted tongue position for /a/ and a higher tongue position for /i:/. Thus, statistical tests revealed a significant difference between males and females only for the mean Euclidean distance (ED) between /i:/ and /a:/ (Welch two-sample t-tests, $t = -2.7$, $df = 5.9$, $p < 0.05$). This is also expressed in the sex-specific dimensions of the space: while males on average exhibit a 1.3 times larger vertical than horizontal expansion, this relationship is around 1 for the females.

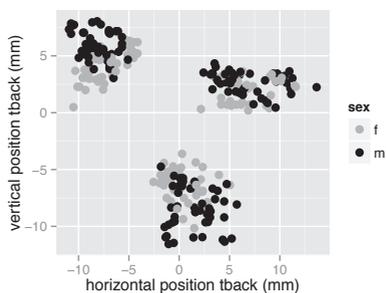


Figure 6: IAU-polygon: articulatory positions of the tongue back coil measured in the double vowel sequences of the abbreviation. Female speakers are in grey, male speakers in black.

4.1.4. Sex-specific differences in undershoot

Analyses of /gV/-sequences served two aims. The first one was to compare the vowel space resulting from the tense vowels produced in this sequence with the speaker-specific “extreme” reference vowel space resulting from the IAU-polygon. This made it possible to analyze the degree of coarticulation-induced undershoot individually for each speaker and then compare it between speakers, and ultimately sexes. The second aim was to compare the degree of accent-induced undershoot between males and females by analyzing the vowels of the /gV/-sequence in three different accent conditions (control, accented, unaccented).

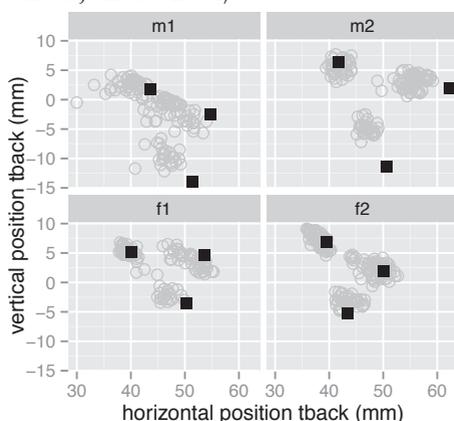


Figure 7: Articulatory positions of the tongue back coil during the IAU (extreme) vowel space (black squares) and the /gV/-sequence (grey circles) of two male (m1, m2) and two female (f1, f2) speakers.

Figure 7 gives a first hint of speaker- (or sex-) specific differences in the relationship between the “extreme” IAU vowel space and the coarticulatorily more affected /gV/-vowel space. The figure shows four subplots, visualizing the data of two male speakers (above, m1, m2), and two female speakers (below, f1, f2). For each speaker, the black squares show the mean IAU-polygon, the grey circles show the vowel space resulting from the /gV/-sequence for all three accent conditions. While for both female speakers the vowel spaces of the IAU and the /gV/-sequence overlap considerably, differences are apparent for both male speakers, especially in terms of a lower and more retracted position for /a/ in the IAU space compared to the /gV/-space.

The bars of Figure 8 show the average female (grey) and male (black) vowel space of the /gV/-sequence in absolute terms (in mm²) separately for the three accent conditions. The male speakers reveal higher values for all accent conditions; however, the difference is only considerable for the accented condition. In addition, the figure shows the average female and male vowel space of the /gV/-sequence in percent of the IAU-space (black and grey circles connected by the lines). The relationship between the vowel space produced within the /gV/-sequence and the “extreme” IAU vowel space was calculated for each speaker and accent condition separately. It is apparent that here, females reveal substantially higher values than males in the control and unaccented condition, while in the accented condition the considerable difference between males and females found for the absolute values is absent.

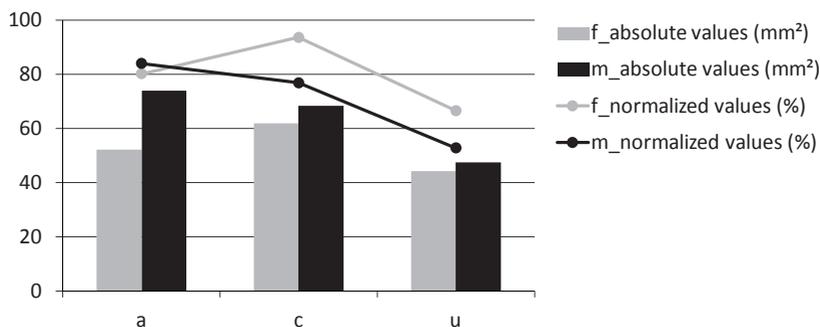


Figure 8: Average polygon sizes of the /gV/-data for male (black) and female (grey) speakers separated by accent condition (a, c, u). The bars represent the vowel spaces in absolute terms (mm²), the connected dots represent the vowel spaces in normalized terms (%).

For statistical analysis (linear mixed models), not the overall vowel space size per speaker, but rather EDs from the midpoint of the vowel space to each vowel were used as dependent variable. In this way, the number of data points could be increased and a vowel specific analysis could be undertaken. Two analyses were run, with either the absolute EDs as dependent variable or the ED expressed as a percentage of the EDs between vowels and the midpoint measured in the IAU data. Model comparisons (likelihood ratio tests) were conducted to find the model with the best fit to the data. For the absolute ED as dependent variable, we found a significant interaction of sex*vowel and sex*accent condition (random factors included were speaker and repetition). Regarding the first interaction, a significant difference between males and females was only found for the vowel /a:/ analogous to the results of the IAU polygon (estimate: 2.3, pMCMC-value < 0.01). Regarding the second interaction, males show a significant difference between accented and unaccented (estimate: -1.5, pMCMC < 0.01), while females do not.

For the normalized EDs as dependent variable, we found a significant interaction of sex*accent. In contrast to the absolute values, no sex-specific differences were found for the ED for /a:/ (or any other vowel). However, analogous to the absolute values, the factor accent condition showed its significance in terms of sex-specific differences: males differed between accented and unaccented condition (estimate: -17.6, pMCMC < 0.01), while females did not.

4.2. Discussion

Our results are in line with the hypothesized higher probability of accent-induced undershoot in males than in females: while males show the expected significantly smaller articulatory vowel spaces in unaccented conditions (in absolute and normalized values), females do not differ between the accent conditions. Additionally, the expected larger articulatory spaces in males were only found in the IAU-data, where articulatory positions are assumed to be “extreme” in terms of being minimally affected by coarticulation-induced undershoot. While we cannot rule out that females “do more” than males on purpose (in terms of reaching their articulatory targets irrespective of accent condition and coarticulatory influences to achieve a large acoustic

vowel space resulting in clear speech), we have indeed seen that sex-specific differences vary depending on whether we look at the absolute or relative values: while males exhibit larger articulatory distances in absolute values, the difference is leveled out or even reversed in normalized values. This leads us to suggest that the same articulatory distance results in different acoustic outputs in speakers who differ in their maximal articulatory spaces (such as an average male and an average female speaker). To examine this more closely we are currently investigating potential sex-specific differences in the articulatory-acoustic relationship in the production of diphthongs.

Thus, sex-specific differences in acoustic vowel spaces might be due to differences in anatomically restricted articulatory spaces between males and females. We suggest that the underlying dynamics of the articulatory gestures play a crucial role in sex-specific differences.

5. General discussion

Various sources of inter-speaker variability, including behavioral and organic factors exist and all of them are worthy of systematic investigation and categorization. In this chapter the impact of organic factors on inter-speaker variability was highlighted by investigating two speaker groups in which biological variation is a central issue: related speakers (twins) and male and female speakers. We have seen that individual differences in lingual strategies can at least in part be explained by idiosyncratic physiological restrictions. In particular, the shape of the palate, the physiological properties of the tongue muscles and the size and shape of the vocal tract seem to be crucial factors regarding articulatory inter-speaker variability. The variability found is systematic and explainable, and can help us understand some of the underlying principles of the speech production process. Following Lindblom's assumption that languages "tend to evolve sound patterns that can be seen as adaptations to biological constraints of speech production" (1983, p. 217), it is suggested, that also at least some of the inter-speaker variability we have discussed mirrors speaker-specific adaptations to individual biological restrictions (see also Lammert et al., 2013). While in many phonological theories speaker-specific behavior is considered a source of random noise with no impact on phonemic categories, we found a significant influence of individual differences in the alveolo-palatal

steepness on inter-speaker variability in realizing the phonemic contrast of /s-/ʃ/ in German and thus, on the phonetic realization of two phonemic categories.

Furthermore, in addition to investigating phonemic contrasts instead of targets, another crucial step in the analysis of inter-speaker variability is to focus more on the dynamic aspect of speech. In line with Nolan et al. (2006), who suggested the speech signal contains linguistically determined targets and organically determined transitions, we found a significant influence of organically determined differences on the looping movements of the tongue (i. e. transitions) during /aCV/-sequences. While it is recognized that the movement is of course also affected by the targets (and here especially the stop closure at the palate), we suggest that dynamic patterns in speech are especially appropriate for showing the influence of organic sources on inter-speaker articulatory variability. Whether the properties of the tongue muscles (biomechanics) or the palate shape/vocal tract dimensions (physical constraints) are the chief influencing factor remains to be examined.

Another way of highlighting the underlying dynamic nature of articulatory gestures is to set the lingual movement in relation to the size and shape of a speaker's individual and organically determined articulatory space, instead of comparing the absolute sizes of the movement. The same articulatory movement (in shape and size) might be extreme for a small female speaker but only half of the potential movement size of a large male speaker. If there is enough time for the gesture (as e. g. in an accented position) both speakers will reach their extreme target positions. These maximal positions will differ according to the speakers' respective physiological space, as was shown for the IAU vowel spaces. If time is short (for example due to a target's occurrence in an unaccented position) the female speaker might reach her extreme position, while the male speaker reaches only 50% of the movement amplitude he could reach when there were enough time. This might be one reason for the sex-specific differences we found in accent-induced undershoot: while males revealed significantly smaller amplitudes in unaccented conditions than in accented conditions, females did not differ. Going a step further, it could be suggested that the same articulatory gesture (in shape and size) results in different acoustic outputs. This is especially interesting in the light of the mismatch between articulation and acoustics regarding sex-specific differences in vowel space

sizes: despite having smaller articulatory vowel spaces, females exhibit larger acoustic vowel spaces than males (Simpson 2002). In our current work we are investigating the acoustic vowel spaces of the IAU-data, and while males showed larger articulatory distances, the acoustic distances between the vowels did not differ between the sexes. In addition, we are examining acoustic and articulatory diphthong realizations in males and females and while no significant sex-specific difference in absolute articulation is found, males and females differ in their respective acoustic output. In both cases females achieve more (in acoustic terms) by doing less (in articulatory terms) compared to males.

Acknowledgements

This work was supported by the German Federal Ministry of Education and Research (BMBF) (01UG0711) and two German Research Council Grants (SI 743/6-1,2) awarded to Adrian Simpson. I am grateful to the *Zentrum für Allgemeine Sprachwissenschaft* (ZAS) in Berlin particularly for its financial support regarding the publication of the loop paper, to Jörg Dreyer for technical support with the twin recordings, to the Dept. of Linguistics, University of Potsdam and Adamantios Gafos, Christian Geng and Jana Brunner for help with the gender recordings. Many thanks to the two reviewers of this chapter, the editors of this book, and of course the participating subjects. Thanks also to my marvelous co-authors involved in these studies Leonardo Lancia, Jana Brunner, Susanne Fuchs and Adrian Simpson. All mistakes are my own.

References

- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, N.J.: Prentice-Hall.
- Bladon, R., Henton, G., and Pickering, J. (1983). Towards an auditory theory of speaker normalization. *Language and Communication*, 4, 59–69.
- Browman, C.P., and Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica*, 49 (3-4), 155–180.
- Brunner, J., Fuchs, S., and Perrier, P. (2009). On the relationship between palate shape and articulatory behavior. *The Journal of the Acoustical Society of America*, 125, 3936–3949.

- Brunner, J., Fuchs, S., and Perrier, P. (2011). Supralaryngeal control in Korean velar stops. *Journal of Phonetics* 39(2), 178–195.
- Chambers, J. (2003). *Sociolinguistic theory*. Oxford: Blackwell.
- Chiba, T., and M. Kajiyama. (1941). *The vowel – its nature and structure*. Tokyo, Japan: Tokyo-Kaiseikan.
- Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics* 32(2), 141–176.
- Debruyne, F., Decoster, W., Van Gijssel, A., and Vercammen, J. (2002). Speaking fundamental frequency in monozygotic and dizygotic twins. *Journal of Voice* 16(4), 466–471.
- de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The 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., and Edwards, L. (1993). The interplay between prosodic structure and coarticulation. *Language and Speech* 36 (2-3), 197–212.
- Diehl, R. L., Lindblom, B., Hoemeke, K. A., and Fahey, R. P. (1996) On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics*, 24, 187–208.
- Enguchi, S., Townsend, G. C., Richards, L. C., Hughes, T., and Kasai, K. (2004). Genetic contribution to dental arch size variation in Australian twins. *Archives of Oral Biology*, 49, 1015–1024.
- Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.
- Fant, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scaling, *STL QPSR* 4, 22–30.
- Fant, G. (1975). Non-uniform vowel normalization. *STL-QPSR* 2-3, 1–19.
- Flanagan, J. R., Ostry, D. J., and Feldman, A. G. (1993). Control of trajectory modifications in target-directed reaching. *Journal of Motor Behaviour*, 25(3), 140–152.
- Foulkes, P. and Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, 34, 409–438.
- Fuchs, S., Perrier, P., Geng, C., and Mooshammer, C. (2006). What role does the palate play in speech motor control? Insights from tongue kinematics

- for German alveolar obstruents. In J. Harrington and M. Tabain, (eds.) *Speech production: Models, phonetic processes, and techniques*, New York: Psychology Press, 149–164.
- Fuchs, S., Winkler, R., and Perrier, P. (2008). Do speakers' vocal tract geometries shape their articulatory behavior? In *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.
- Geng, C., Fuchs, S., Mooshammer, C. and Pompino-Marschall, B. (2003). How does vowel context influence loops? In *Proceedings of the 6th International Seminar on Speech Production*, Sydney, Australia, 67–72.
- Ghosh, S., Matthies, M., Maas, E., Hanson, A., Tiede, M., Ménard, L., Guenther, F., Lane, H., and Perkell, J. S. (2010). An investigation of the relation between sibilant production and somatosensory and auditory acuity. *The Journal of the Acoustical Society of America*, 125, 3079–3087.
- Goldstein, U. (1980). *An articulatory model for the vocal tracts of growing children*. PhD Thesis, MIT.
- Gribble, P. L. and Ostry, D. J. (1996). Origins of the power law relation between movement velocity and curvature: Modeling the effects of muscle mechanics and limb dynamics. *Journal of Neurophysiology*, 76(5), 2853–2860.
- Gribble, P. L., Ostry, D. J., Sanguineti, V., and Laboissière, R. (1998). Are complex control signals required for human arm movement? *Journal of Neurophysiology*, 79(3), 1409–1424.
- Hardcastle, W. J., Gibbon, F., and Nicolaidis, K. (1991). EPG data reduction methods and their implications for studies of lingual coarticulation. *Journal of Phonetics*, 19, 251–266.
- Harrington, J. (2006). An acoustic analysis of happy-tensing in the Queen's Christmas broadcasts. *Journal of Phonetics*, 34, 439–457.
- Harrington, J., Fletcher, J., and Beckman, M. (2000). Manner and place conflicts in the articulation of accent in Australian English. In M. Broe and J. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the lexicon*. (pp. 40–51) Cambridge: Cambridge University Press.

- Henton, C. G. (1995). Cross-language variation in the vowels of female and male speakers. In *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm, Vol. 4, 420–423.
- Hoole, P., Munhall, K. and Mooshammer, C. (1998). Do airstream mechanisms influence tongue movement paths? *Phonetica*, 55, 131–146.
- Kabban, M., Fearne, J., Jovanovski, V., and Zou, L. (2001). Tooth size and morphology in twins. *International Journal of Paediatric Dentistry*, 11, 333–339.
- Kent, R. and Moll, K. (1972). Cinefluorographic analyses of selected lingual consonants. *Journal of Speech, Language and Hearing Research*, 15, 453–473.
- Ladefoged, P. and Broadbent, D. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, 29(1), 98–104.
- Lammert, A.; Proctor, M., and Narayanan, S. (2013). Interspeaker variability in hard palate morphology and vowel production. *Journal of Speech, Language, and Hearing Research* 56, S1924–S1933.
- Lancia, L. and Tiede, M. (2012). A survey of methods for the analysis of the temporal evolution of speech articulator trajectories. In S. Fuchs, M. Weirich, D. Pape and P. Perrier (eds.) *Speech production and perception: Speech planning and dynamics*, Frankfurt/Main: Peter Lang, Vol. 1, 239–277.
- Langer, P., Tajtáková, M., Bohov, P., and 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.
- Lieberman, A. M., Cooper, F., Shankweiler, D., and Studdart-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461.
- Lieberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36.
- Lindblom, B. (1983). Economy of speech gestures. In P.F. MacNeilage, (ed.) *The production of speech*. New York: Springer, 217–245.
- Lindblom, B. (1988). Phonetic invariance and the adaptive nature of speech. in B.A. Elsendoom and H. Bouma, (eds.) *Working models of human perception*. London: Academic Press, 139–173.

- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. in W. J. Hardcastle and A. Marchal, (eds.) *Speech Production and Speech Modelling*, Dordrecht: Kluwer, 403–439.
- 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.
- Lundström, A. (1948). *Tooth Size and Occlusion in Twins*. Basel: Karger.
- Locke, J. L. and 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. and Gracco, V. L. (2002). Control of oral closure in lingual stop consonant production. *The Journal of the Acoustical Society of America*, 111(6), 2811–2827.
- Lucero, J. C.; Munhall, K. G.; Gracco, V. L., and 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.
- McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748.
- Mooshammer, C., Hoole, P., and Kühnert, B. (1995). On loops. *Journal of Phonetics*, 23, 3–21.
- Mooshammer, C. and Geng, C. (2008). Acoustic and articulatory manifestations of vowel reduction in German. *Journal of the International Phonetic Association*, 38, 117–136.
- Nolan, F., Oh, T., McDougal, K., de Jong, G., and Hudson, T. (2006). A forensic phonetic study of ‘dynamic’ sources of variability in speech: The DyViS project, in *Proceedings of the 11th Australian International Conference on Speech Science and Technology*, Auckland, Australia, 13–18.
- Nolan, F. and Oh, T. (1996). Identical twins, different voices. *Forensic Linguistics: International Journal of Speech, Language and the Law*, 3, 39–49.
- Nordström, P.-E. (1977). Female and infant vocal tracts simulated from male area functions. *Journal of Phonetics*, 5, 81–92.
- O’Neill, B. (2006). *Elementary differential geometry*, 2nd ed., New York: Academic, Chap. 5, 202–263.

- Ooki, S. (2005). Genetic and environmental influences on stuttering and tics in Japanese twin children. *Twin Research and Human Genetics*, 8(1), 69–75.
- Perkell, J. S. (2010). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, 25, 382–407.
- Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E. and Guenther, F. H. (2004). The distinctness of speakers' /s/-/S/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of Speech, Language, and Hearing Research*, 47, 1259–1269.
- Perrier, P. and Fuchs, S. (2008). Speed-curvature relations in speech production challenge the 1/3 power law. *Journal of Neurophysiology*, 100(3), 1171–1183.
- Perrier, P., Payan, Y., Zandipour, M. and Perkell, J. S. (2003). Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *The Journal of the Acoustical Society of America*, 114, 1582–1599.
- Pinheiro, J. and Bates, D. (2000). *Mixed-Effects Models in S and S-Plus*. Statistics and Computing Series, New York: Springer.
- Przybyla, B. D., Horii, J., and 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*. R Foundation for Statistical Computing, Vienna, Austria.
- Rudy, K. and Yunusova, Y. (2013). The effect of anatomic factors on tongue position variability during consonants. *Journal of Speech, Language, and Hearing Research*, 56, 137–149.
- Ryalls, J. H. and Lieberman, P. (1982). Fundamental frequency and vowel perception. *The Journal of the Acoustical Society of America*, 72, 1631–1634.
- Scarr, S. and Carter-Saltzman, L. (1979). Twin method: Defense of a critical assumption. *Behavior Genetics*, 9(6), 527–542.
- Simberg, S.; Santtila, P.; Soveri, A.; Varjonen, M.; Sala, E., and Sandnabba, N. K. (2009). Exploring genetic and environmental effects in dysphonia:

- A twin study. *Journal of Speech, Language and Hearing Research*, 52, 153–163.
- Simpson, A. P. (1998). Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel* (AIPUK) 33).
- Simpson, A. P. (2001). Dynamic consequences of differences in male and female vocal tract dimensions. *The Journal of the Acoustical Society of America*, 109, 2153–2164.
- Simpson, A. P. (2002). Gender-specific articulatory-acoustic relations in vowel sequences, *Journal of Phonetics*, 30, 417–435.
- Simpson, A. P. and Ericsson, C. (2007). Sex-specific differences in f_0 and vowel space. In *Proceedings of the XVIth International Congress of Phonetic Sciences*, Saarbrücken, 933–936.
- Spinath, F. M. (2005). Twin designs. In B. S. Everitt and D. C. Howell, (eds.) *Encyclopedia of statistics in behavioral science*, Chichester: John Wiley & Sons, 2071–2074.
- Stevens, K. and Blumstein, S. (1978). Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, 64, 1358–1368.
- Tasko, S. M. and Westbury, J. R. (2004). Speech-curvature relations for speech-related articulatory movement. *Journal of Phonetics*, 32, 65–80.
- Toda, M. (2006). Deux stratégies articulatoires pour la réalisation du contraste acoustique des sibilantes /s/ et /ʃ/ en français. *Actes des XXVI^e Journées d'Étude de la Parole*, Dinard, 65–68.
- van Lierde, K., Vinck, B., Ley, S., Clement, G., and van Cauwenberge, P. (2005). Genetics of vocal quality characteristics in monozygotic twins: A multiparameter approach. *Journal of Voice*, 19(4), 511–518.
- Weirich, M. (2012). *The influence of nature and nurture on speaker-specific parameters in twins' speech: Articulation, acoustics and perception*, Ph.D. dissertation, HU Berlin.
- Weirich, M. and Lancia, L. (2011). Perceived auditory similarity and its acoustic correlates in twins and unrelated speakers. In *Proceedings of the XVII International Congress of Phonetic Sciences*, Hong Kong, 2118–2121.

- Weirich, M. and Fuchs, S. (2013). Palatal morphology can influence speaker-specific realizations of phonemic contrasts. *Journal of Speech, Language and Hearing Research*, 56, S1894–S1908. [<http://jslhr.pubs.asha.org/article.aspx?articleid=1802795>]
- Weirich, M., Lancia, L., and Brunner, J. (2013). Inter-speaker articulatory variability during vowel-consonant-vowel sequences in twins and unrelated speakers. *The Journal of the Acoustical Society of America*, 134 (5), 3766–3780. [<http://scitation.aip.org/content/asa/journal/jasa/134/5/10.1121/1.4822480>]
- Weirich, M. and Simpson, A.P. (2013). Investigating the relationship between average speaker fundamental frequency and acoustic vowel space size. *The Journal of the Acoustical Society of America*, 134 (4), 2965–2974.
- Weirich, M. and Simpson, A.P. (2014a). Differences in acoustic vowel space and the perception of speech tempo. *Journal of Phonetics*, 43, 1–10.
- Weirich, M. and Simpson, A.P. (2014b). Articulatory vowel spaces of male and female speakers. In *Proceedings of the 10th Intern. Seminar on Speech Production (ISSP)*, Cologne, 453–456.
- Whiteside, S. P. (2001). Sex-specific fundamental and formant frequency patterns in a cross-sectional study. *The Journal of the Acoustical Society of America*, 110, 464–478.
- Whiteside, S. P. and 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. and 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.
- Winkler, R., Fuchs, S., and 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 modeling. In *Proceedings of the 7th International Seminar on Speech Production*, Ubatuba, 509–516.

