Articulatory and Acoustic Inter-Speaker Variability in the Production of German Vowels

Melanie Weirich

Zentrum für Allgemeine Sprachwissenschaft, Berlin

This study examines articulatory and acoustic inter-speaker variability in the production of the German vowels /i/, /u/ and /a/. Our subjects are 3 monozygotic twin pairs (2 female and 1 male pair) and 2 dizygotic female twin pairs. All of them were born, raised and are still living in Berlin and see their twin brother or sister regularly. We assume that monozygotic twins that are genetically identical and share the same physiology should be more similar in their articulation than dizygotic twins but that the shared time and social environment of twins, regardless of their genetic similarity, also plays a crucial role in the acoustic similarity of twins. Articulatory measurements were made with EMA (Electromagnetic Articulography) and the target positions of the produced vowels were analyzed. Additionally, the formants F1-F4 of each vowel were measured and compared within the twin pairs. Our data seems to point out the importance of a shared environment and the strong influence of learning over the anatomical identity of the monozygotic twins regarding the production of vowels. But, additional results suggest (1) the impact of physiology on the production of a vowel following a velar consonant and (2) the interaction of physiology and stress in inter-speaker variability.

1   Introduction

Acoustic and articulatory variability are essential topics and crucial parameters in the field of speech production, and inter- and intra-speaker variability is one of the hallmarks of communication. To answer the question how intra- and inter-speaker variability is influenced by different factors is one aim of this paper. At this, the non-linear relation between articulation and acoustics (Stevens, 1972) has to be kept in mind, since differences in articulation do not necessarily result in differences in the acoustic output. Additionally, articulatory variability can occur in terms of motor equivalence, since some speech parameters can be

obtained by different alternatives in articulation and demonstrate "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 and Abbs, 1976, 199). 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).

In many research fields regarding the variability of certain human properties the impact of *nature* and *nurture* is discussed. However, what influence *nature* on the one hand and *nurture* on the other hand have on the acoustics and articulation in speech production, and how these determinants interact in terms of inter- and intra-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). Biology and hence physiological and biomechanical factors play an important role in terms of motor control and articulatory targets in speech production (Lindblom, 1983). The factor learning and social environmental influences come to the fore when we speak of auditory goals (Perkell et al. 1997).

The question we are addressing is what roles these two different influence sources play in terms of variability in speech production. Thus, when do physiological constraints prevail the impact of our social environment, and when does as a result biology have an impact on articulatory variability? And on the other hand, in which cases are we free in choosing different articulatory strategies to reach an auditory goal, and nurture seems to be the determining factor?

In psychology, learning theories like the social learning theory of Albert Bandura (1977) emphasize that people in general learn by observing. In terms of language acquisition it can be assumed that children learn the syntactical structures, phonological rules and lexical parameters of a language through imitation of the people surrounding and talking to them. Moreover, also dialectal pronunciation and socio-linguistic parameters of the parents are observed and absorbed by the child. Therefore, social environmental factors play an important role in speech production. Speech acquisition has to be at least partly independent of individual differences in the physiology of the vocal apparatus as it is in general possible to learn and speak any existing language for a child that is young enough and does not reveal any speech, language or hearing impairment. In exemplar-based models (Pierrehumbert, 2001, Johnson, 1997), social environmental influences are also seen to be crucial, since it is assumed that more recently encountered utterances are stored with higher activation levels than older utterances. Hence, socio-linguistic variation may be explained

partly by a change in nurture. Thus, some non-biological determinants of variability in speech could be shown, namely: social environment, observing and learning.

On the other hand, it cannot be denied that speech has a biological foundation. For example, the shape of the vocal tract influences its filter characteristics and therefore, speakers differ in their formant values (Fant, 1960). The speech signals we are able to produce and perceive are limited by our physiology (Fuchs et al., 2007). Regarding the economy of speech gestures, Lindblom (1983, 217) assumes that "languages tend to evolve sound patterns that can be seen as adaptations to biological constraints of speech production". A connection between biology and differences in articulatory token-to-token variability was found by Shiller et al. (2002). Physiological parameters seem to constrain the capability of our speech motor system. In their study an asymmetric relationship between jaw stiffness and kinematic variability could be shown: higher stiffness values were observed along the axis of jaw protrusion and retraction and go hand in hand with reduced kinematic variability. In addition, for high jaw positions, stiffness was greater and kinematic variability smaller. Perkell (1985) found in his 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 than in direction parallel to the midline and hence, supporting evidence of the physiological "saturation effects". Furthermore, studies by Mooshammer et al. (2004) and Brunner et al. (2005) have shown that variability appeared to be less when the amount of linguo-palatal contact was high, suggesting high biomechanical restrictions in the production of high vowels. In addition, velar stops are described to be strongly influenced by anatomical and physiological constrictions. Perrier (2003) investigated looping movements of the tongue during VCV sequences and showed in their study that the production of the consonant - which was a velar stop - is influenced by the vowels. They assume that biomechanical factors are at least partly responsible for the resulting looping movements in the production of a velar stop. That anatomical properties, in particular the shape of the palate, could also play a role in inter-speaker variability was shown 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 small variation of the tongue position has a larger impact on the area function and henceforth on the acoustics than in speakers with a dome-shaped palate.

It has to be kept in mind that when we talk about speech, first of all we talk about communication. And speaker-specific variability can also 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. The speech signal itself consists of different segments with different 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 time (de Jong, 1997). It may also be assumed that these stressed syllables correspond to learned auditory goals, and that the unstressed syllables are generally shorter in time, more influenced by coarticulation processes, less articulatory distinct and more variable. Thus, it can be hypothesized that unstressed syllables are more sensible to physiological factors, more influenced by the individual vocal apparatus and less oriented towards the learned auditory targets. Similarly, certain prosodic boundaries are lengthened and potentially strengthened in comparison to others. For example, Fougeron and Keating (1997) point to the process of articulatory strengthening at the edges of prosodic domains: the higher the prosodic domain, the more linguo-palatal contact was found. As a consequence of this, a stronger influence of physiology and anatomical parameters can be assumed at higher prosodic boundaries with more linguo-palatal contact. Hence, when we talk about articulatory and acoustic variability in speech, the factor stress and the parameter prosodic domain have to be kept in mind.

Different factors that can have an impact on variability in speech have been shown: the influence of social environment, learning, physiology and stress patterns. One possibility to investigate the interaction and impact of these factors and to describe the influence of nature and nurture are twin studies.

1.1 Twin studies

In order to study the influence of biological parameters (physiology, biomechanics) on the one hand and non-biological parameters (learning, environmental factors) on the other hand, we investigated inter-speaker variability in the speech of monozygotic twins (MZ, who are genetically identical) and dizygotic twins (DZ, who share approximately 50% of their genes). For example, if we find high inter-speaker variability in a certain speech parameter within a MZ twin pair that lives apart from each other, the influence of genetics and physiology on this parameter seems to be rather small. We will discuss our results in terms of articulatory and auditory targets. If underlying articulatory targets are assumed in speech production, MZ twin pairs should be identical in their articulation (and hence also acoustics) independent of the time they spend together. If auditory goals are assumed, they should differ in their acoustical output when living apart from each other. In addition to that, when a

DZ pair that spends most of their time together is very similar in their acoustical outputs of a certain speech parameter, auditory goals and not articulatory targets must be crucial in this parameter.

Twin studies are a common method in the field of psychology. The two influence factors investigated are (1) identical genes (and physiology) and (2) a shared 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 caused similarity, as studies of mislabelled twin pairs have shown (Spinath, 2005).

Comparing the similarity of DZ and MZ twins within the field of speech research is rather new and less common. Still, some studies regarding perceptual and acoustic differences within twin pairs have been conducted (for an overview see Loakes, 2006). Perception experiments revealed the striking similarity between twins' voices but also showed that twin pairs can be identified above chance by people familiar with their voices (Whiteside & Rixon, 2001). The most frequently investigated speech parameter in twins' speech is fundamental frequency, and results point to a great influence of physiology on this parameter as MZ twins reveal higher correlations than DZ twins (Przybyla et al., 1992, Debruyne et al. 2002). Also, voice quality characteristics analyzed by Van Lierde et al. (2005) such as perceptual voice characteristics, vocal performances, and the overall vocal quality by means of the Dysphonia Severity Index were very similar in MZ twins. For shimmer and jitter only, no significant correlation coefficient could be obtained. Smaller influence of identical genes and physiology and greater impact of environmental factors were found by Debruyne et al. (2002) for the variation of speaking fundamental frequency, as MZ and DZ twins revealed the same amount of similarity. Nolan & Oh (1996) showed similarities in coarticulation patterns of /l/ and /r/ but also differences in pronunciation alternatives of /r/. Regarding the production of vowels, the study by Loakes (2004) point to lax front vowels (F2, F3 of /I/) to be most speaker-specific regarding twins. There are still only few studies regarding inter-speaker variability in twins, but there is a greater lack of articulatory studies within this field. We are not aware of any study investigating articulation patterns in the speech of twins, although interesting findings regarding the impact of physiology on inter-speaker variability can be obtained by this.

1.2   Pilot Study

We have conducted a pilot study to reveal acoustic differences in the speech parameters of identical (MZ) and non-identical (DZ) twins (Weirich, 2009). This study should help to locate phonemes that show acoustic differences within twin

pairs and therefore are promising to show also differences in articulation, although the relation of 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 variance but low intra-speaker variance. 4 identical twin pairs and 1 non-identical twin pair were recorded with a large set of sentences repeated 5 times. The results revealed the vowels /u/ and /a/ to be most speaker-specific within all twin pairs. 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 [ɑ, ɑː, ʊ, uː, oː] revealed more significant differences than the front vowels [ɪ, iː, ɛ, eː, ʏ, yː]. The formant with most differences among F1-F4 was F3, followed by F1 and F4. F2 showed the least variation within all twin pairs. A clear discrepancy in the amount of differences in the formants could be found between non-related persons (twins from different pairs) and related persons (twin pairs). For the unrelated pairs the first three formants showed a relatively stable probability of over 50% of showing differences, F4 indicates a somewhat smaller value (45%), whereas the twins revealed an average probability of only 28%. Comparing the MZ with the DZ twins, the DZ twin pair reveals only in F3 a higher probability of showing differences. The differences between inter-speaker variability within DZ and MZ pairs were biggest 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. Furthermore, our pilot study revealed the impact of the factor stress on inter-speaker variability in plosives: 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. Since unstressed syllables are perceptually less important, they can be more variable and thus, they do not show significant differences. The stressed syllables are less variable and the learnt auditory goals are more crucial here. Another explanation could be that physiology has a greater influence on speech production in unstressed syllables than in stressed syllables. Of course, the results of the pilot study have to be treated very carefully, as the DZ twins are presented by only one pair.

1.3   Hypotheses

The aims of our twin study are to explore, (a) to what extent speaker specific variability is influenced by the speaker's respective articulatory constraints and (b) to what extent speakers are influenced by their social environment and yet free to aim for a given auditory goal. The purpose of the present articulatory and

acoustic study is to verify and cross-check the results of the pilot study regarding acoustical differences and similarities in vowels and additionally to examine the underlying articulatory patterns in the speech of MZ and DZ twins.

Our hypotheses are that (1) DZ twin pairs need not naturally show more differences in their acoustic outputs regarding vowels, as they adjust their speech production to each other and use auditory goals to produce vowels, but may reveal their anatomical and physiological differences in individual articulatory strategies, when this is possible. In addition to that, it is proposed that (2) the physiological influences of the tongue and the shape of the palate show a greater impact on the production of consonants (e.g. velar stops) and hence, also on vowels in their neighbourhood. Additionally, the factor stress is taken into account and it is hypothesized that (3) more inter-speaker variability can be found in stressed than in unstressed syllables within MZ twin pairs, mirroring the greater influence of physiology in the production of an unstressed syllable. Additionally, the DZ twins should reveal more differences in the unstressed syllables.

Therefore, an articulatory and acoustic analysis of inter-speaker variability within twin pairs was made of the vowel targets of /a/, /i/ and /u/ produced in a stressed syllable, and additionally of the vowel /i/ following a velar stop produced in the sequences /hagi/ (in a posttonic syllable) and /giba/ (in a tonic syllable).

2   Method

The presented study is part of a greater PhD project on acoustic and articulatory inter-speaker variability in the speech of twins. With respect to space limitation and for sake of simplicity, here, we will concentrate on our findings addressing vowels. Considerations regarding our speech material are based on earlier literature, studies that examined intra- and inter-speaker variability of speech parameters, and on the results of our pilot study. Because of the latter, the factor stress is taken into account. In the following section the experimental design, our speech material and subjects, and the acoustic and articulatory measurements that we conducted will be described.

2.1   Experimental set up

We investigated articulatory and acoustic inter-speaker variability within mono- and dizygotic twin pairs in the production of vowels. Articulatory recordings were carried out using a 2 D Electromagnetic Articulograph (Carstens AG 100). For the articulatory measurements two sensors, one at the upper incisors and one at the bridge of the nose, served as reference sensors to compensate for head

movements. Three coils were glued to the tongue: one at the tongue tip, one at the tongue back and one between these two on the tongue dorsum. Another sensor was placed below the lower incisors in order to track jaw movements, and two further sensors were glued to the upper and the lower lip to record lip movements. 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, we measured the distances between the glued coils, took photos of the set up, i.e. the tongue with the glued coils on top and created a template for the first twin to use it as a reference for the second twin.

In addition to the articulatory recordings, the audio tracks were recorded for our speakers for further acoustic analysis (48 kHz sampling rate). After the recordings, the sensor of 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 shape of the palates within the twin pairs. Second, a silicone dental and palate cast was taken, and the shape and steepness of the palate was examined more closely. This was important to us to verify the assumption of identical physiology in MZ pairs. 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.

2.2   Speech material

Our speech material consists of the stressed vowels /a/, /i/ and /u/ in a non-focused position. They were embedded in the verbs [ˈliːbə] (to love), [ˈvaʃə] (to wash) and [ˈzuːxə] (to search for) which were part of the carrier sentences. In that way we could increase the amount of renditions of the vowels: each speaker repeated the target vowels 40 times in different carrier sentences presented to them on a monitor. For each subject we examined articulatory targets (tongue positions) and acoustic targets (F1-F4) and explored the inter-speaker variability within the twin pairs.

As it is hypothesized that physiology has a stronger influence on consonants than on vowels and especially on velar stops, we conducted a second analysis. We investigated the influence of a velar stop on the production and inter-speaker variability within twins of the following vowel /i/. Therefore, the non-words (names) /ˈhagi/ and /ˈgiba/ that were presented in the carrier sentences were used to investigate articulatory and acoustic inter-speaker variability of /i/ in the syllable /gi/ in contrast to the analysis of the variability in the syllable /li/.

Additionally, the factor stress was taken into account to cross-check the results of our pilot study. We looked for an influence of stress on inter-speaker variability in vowels. Therefore, the production and variability of /i/ in a stressed syllable (/ˈgiba/) and in an unstressed syllable (/ˈhagi/) were compared. Table 1 gives an overview of our speech material. The number of repetitions (N) differs slightly between the speakers due to some articulatory data that had to be excluded from the analysis, and therefore, mean values and standard deviations of the analyzed repetitions are given in the table. Moreover, according to our experimental set-up and time restrictions during the EMA-recordings, the vowels /a/, /i/ and /u/ were iterated approximately 40 times, while the vowel /i/ in the syllable /gi/ was iterated just 9 times in each stress condition. Note also that in contrast to the verbs containing the vowel renditions, the target words /giba/ and /hagi/ were in a focused position.

Table 1: Speech material

vowel	stress condition	target word		$\varnothing$ N (std.dev.)
a	stressed	w**a**sche	/ˈvaʃə/	37.3 (3.1)
i:	stressed	l**ie**be	/ˈliːbə/	44.1 (5.2)
u:	stressed	s**u**che	/ˈzuːxə/	38.3 (2.6)
i:	stressed	G**i**ba	/ˈgiːba/	9.3 (2.0)
i:	unstressed	Hag**i**	/ˈhagiː/	8.8 (1.2)

2.3   Subjects

Our subjects are 3 monozygotic twin pairs (2 female, 1 male) and 2 dizygotic twin pairs (both female) between 20 and 34 years. The genetic similarity (zygosity) of these twins was determined by a genetic laboratory by their genotypic comparisons based upon 16 different genetic markers[1]. We suppose that the physiological and biomechanical properties of the vocal apparatus are rather similar in the former and different in the latter. To verify this assumption, the shape of the palate was examined and compared within the pairs. The palate contours that we recorded after the experiment were rotated and adjusted within the respective siblings, as they could differ in their sitting position and posture of the head. This was done by matching the highest point of the palate of one twin to his sibling and then rotating the palate until it fitted.

---

[1]   Monozygotic twin pairs are genetically identical. If a twin pair differs for any DNA marker, they must be dizygotic. When a reasonable number of markers reveal no differences, it can be concluded that the twin pair is monozygotic (Spinath, 2005).

All twin pairs were born, raised and are still living in Berlin, Germany, but differ with respect to the amount of time they recently spent together. The frequency of spending time together will additionally be considered as a factor influencing inter-speaker variability since the mutual influence of the twins and their shared social environment could play a role in shaping auditory goals. Another possible factor is the attitude towards being a twin. Therefore a separate interview with each subject was conducted and the twins made ratings on a 5 point Likert-scale from 1 (I don't like being a twin) to 5 (I like very much being a twin). The number 3 served as a neutral position with no positive or negative feelings about being a twin. All subjects showed a strong, positive attitude, and except for one pair concurred with their sibling. An overview of the characteristics of our subjects is given in Table 2.

Table 2: The twin pairs with the factors "shared time" and "attitude towards being a twin"

twin	zygosity	sex	age	Frequency of spending time together	Attitude towards being a twin (1-5)
SLCL	MZ	m	32	Twice a month	5 – 5
AFHF	MZ	f	34	Nearly every day	5 – 5
GSRS	MZ	f	26	Live together	5 – 5
LRSR	DZ	f	20	Live together	4 – 5
MGTG	DZ	f	20	Live together	5 – 5

2.4   Acoustic measurements

The vowels /a/, /i/ and /u/ in the target words mentioned in Table 2 were segmented and annotated in PRAAT 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 female) and 5000 Hz (for male), a window length of 0.025s and a real pre-emphasis from 50Hz. Figure 2 gives an example of a labelled vowel. Each measured formant value of every analyzed vowel was checked manually and corrected if necessary.
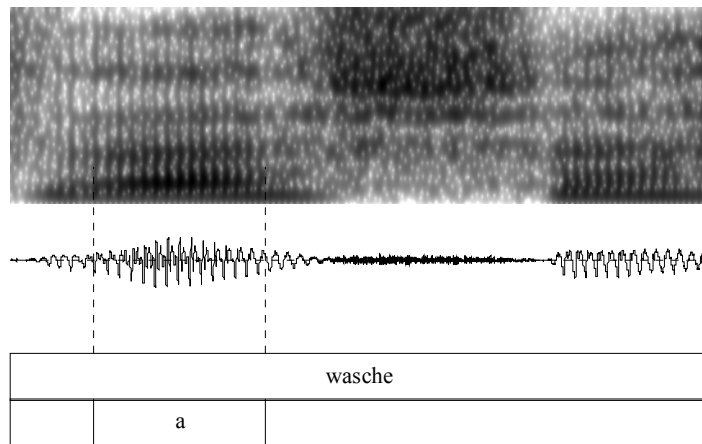
Figure 2: Example of a segmented and annotated vowel in Praat

For each twin pair, mean formant values were compared and the inter-speaker variability was examined. Scatterplots and dispersion ellipses (2 standard deviations) of F1-F2 variation were conducted for each subject and vowel using MATLAB 7.4.0. This was done by a principal component analysis with 2 main components: the highest amount of variability served to define the direction and length of the first axis, the second axis was perpendicular to the first. After that, statistical analyses in R (version 2.8.1) were done to find significant differences in mean formant values within the pairs. First we conducted an ANOVA over all speakers with the respective *formant* as dependent and *speaker* as the independent factor to look for a general effect of the speaker. After that we looked for differences *between the speakers of the same pair* with a Tukey Post Hoc test. 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.

2.5 Articulatory measurements

For each speaker and vowel, articulatory target positions were measured. In general, this is the point in time (and the position of an articulator) when the tongue or the jaw has reached a certain extreme position or maximum and the movement direction changes after reaching this target position. For each vowel a particular articulator was chosen to define the achievement of this articulatory target. Reaching the target position was assumed when the velocity of this particular articulator was minimal. The chosen parameters can be found in the following Table 3.

Table 3: Parameters defining articulatory targets

vowel	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 the upper and lower lip

Articulatory target positions of /a/ were easy to determine 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 already up because of the preceding /l/. Still, target positions for /i/ could be defined certainly in most cases. Defining an articulatory target position of /u/ in the target word /ˈzuːxə/ revealed to be the most difficult determination. Often, no minimum in the velocity of the tongue coils could be found, and therefore, the upper and lower lip and the orientation on the acoustical envelop were taken into account.

Again, scatterplots were made in MATLAB to visualize the articulatory targets of each subject, and vowel and statistical tests in R (ANOVA and Tukey Post Hoc Tests) were conducted to look for inter-speaker variability within the twin pairs in the target positions of each vowel.

3　Results

In the following section we will report our results addressing the analysis of the palates, acoustic and articulatory inter-speaker variability within the twin pairs, and the two factors *stress* and *velar stop* will be taken into account.

3.1　Similarities of the palates

Figure 1 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.
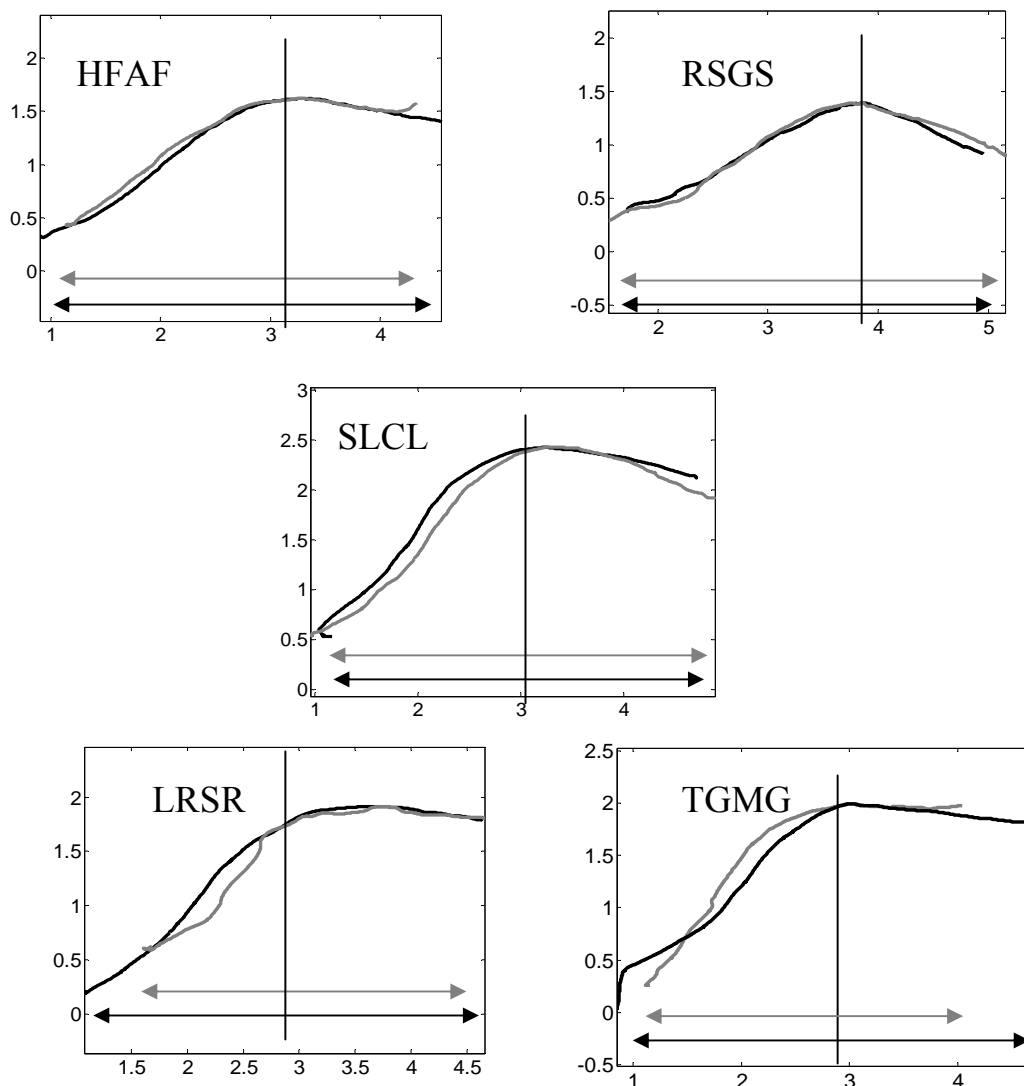
Figure 1: Palatal contours of each twin pair (generated in MATLAB), axis scales in cm, different colors show different subjects.

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 the 2 MZ pairs HFAF and RSGS. In contrast, the palate contours of the DZ pairs LRSR and TGMG are less similar and differ in shape and steepness. At a first glance the palate contours of the MZ pair SLCL also show some differences. But when you look at the size dimensions of the palate, both MZ twins reveal a remarkable high palate (2cm) and are identical in the distance from the highest point of the palate to the beginning of the incisors, whereas both DZ twins vary in these distances (cf. the different lengths of the lines under the respective figures). Especially the DZ pair LRSR varies in the distance from the highest point of the palate to the incisors and LR (grey) reveals a much smaller palate than here sister SR (black).

3.2 The vowel targets /a/, /i/ and /u/

Results regarding our first hypothesis, that is, differences in the formants F1-F4 and the articulatory targets of the vowels /a/, /i/ and /u/ within MZ and DZ twin pairs, are given in the following sections.

3.2.1 Acoustics

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 3 vowels were conducted. The following Figures 2a and 2b show the scatterplots for the 5 twin pairs (Figure 2a shows the 3 MZ pairs and Figure 2b the 2 DZ pairs). Each measured F1/F2 value is marked by a single dot. Ellipses were calculated and drawn to visualize the intra-speaker variability of each vowel. The two colors (black and grey) distinguish the two speakers of a twin pair.
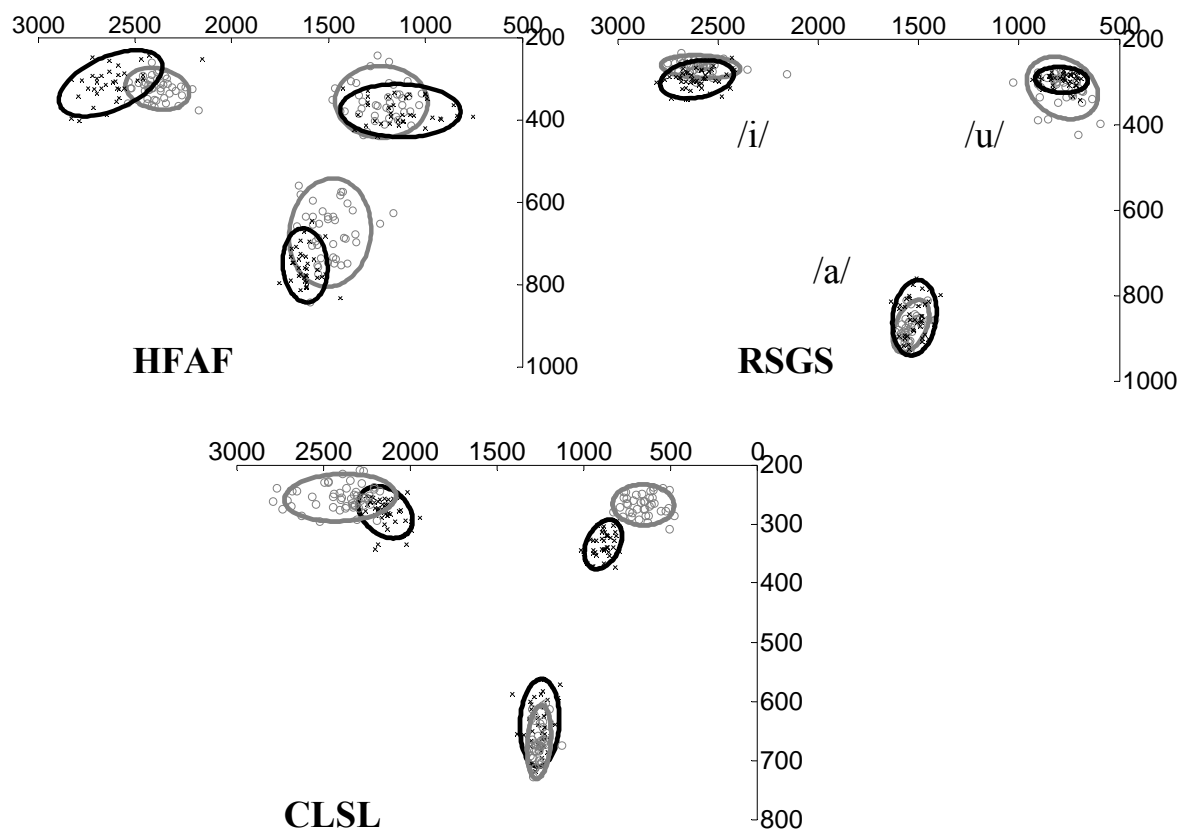

Figure 2a: Scatterplots of F1 (negative y-axis) and F2 (negative x-axis) for the female MZ pairs (HFAF, RSGS) and the male MZ pair (CLSL) and the vowels [a],[i:],[u:]

Figure 2a shows that the vowel spaces of the MZ pair RSGS are most similar; here, the ellipses overlap nearly 100%. For the MZ pairs SLCL and HFAF, differences within the pairs can be assumed in the mean formant values of /i/

and /a/ for HFAF, and /i/ and /u/ for SLCL, but similarities in the sizes of the ellipses and hence, the intra-speaker variability is strikingly apparent: Variability in F1 and F2 is small for SLCL and GSRS, but considerable for HFAF.
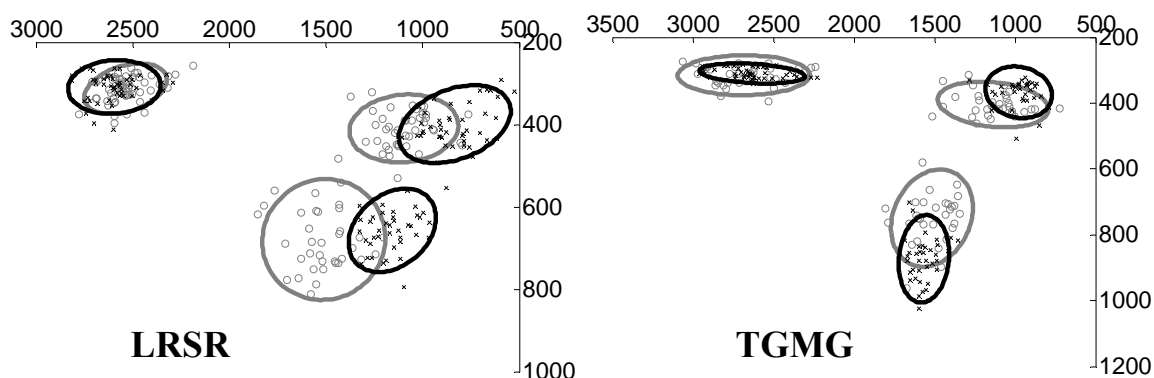

Figure 2b: Scatterplots of F1 and F2 for the female DZ pairs and the vowels /a/,/i/,/u/

In the scatterplots of the DZ twins (fig. 2b), the F1-F2-ellipses of /i/ are most similar within the pairs. For /a/ and /u/, differences can be seen, but especially for LRSR in F2. As was shown in figure 1 before (chapter 2.3), the DZ pair LRSR revealed differences in the sizes of the palates and it can be assumed that the different sizes of the vowel spaces are influenced by these anatomical differences: LR (grey), who showed a much smaller palate also reveals a smaller vowel space than her sister SR (black).

To look for statistically significant differences in formants within the twins, mean formant values of F1 – F4 of the three vowels were measured of each subject and compared with the corresponding sibling. Table 4 shows the significant differences found between speakers of the same twin pair. The MZ pairs showed on average 5.3 differences in F1-F4 of the three vowels, the DZ pairs 7 (of 12 (3 vowels x 4 formants) possible differences), but an ANOVA revealed no significant effect of the factor *zygosity*. Within the MZ pairs, the least inter-speaker variability in formants was found for the twin pair that shares genetics as well as environment (GSRS, as indicated in the scatterplots before), and the most differences were found for the MZ pair which sees each other only twice a month (CLSL). Concerning the amount of differences, the DZ pair LRSR (that lives together) even comes before this MZ pair. As in the pilot test, the results point to the a shared environment as the greatest impact factor on the acoustics of stressed vowels.

An influence of vowel height on acoustical variability as assumed in the introduction could not be found in terms of less inter-speaker variation in /i/ due to the strong influence of physiology on the production of this vowel. In fact, the MZ twins showed most differences in the formants of /i/ and least in /u/, whereby inter-speaker variability within the DZ twins was biggest in /u/ and

smallest in /i/. A possible explanation could be that our investigated speech material consisted of stressed vowels and physiology seems to play a more important role in unstressed vowels.

Table 4: Significant differences in F1-F4 within the twin pairs of /a/, /i/, /u/ (Tukey Post Hoc Test in R, Significance Level <.01)

Twin pair	zygosity	/u/	/a/	/i/	Amount of differences Total /12	
						% of F1/F2
GSRS	MZ	F3	F4	F1, F4	4/12	25 %
AFHF	MZ	-	F1, F2, F3	F2, F4	5/12	60 %
SLCL	MZ	F1, F2	F3	F1, F2, F3, F4	7/12	57 %
SRLR	DZ	F2,F3, F4	F2, F4	F4	6/12	33 %
TGMG	DZ	F1,F2, F3, F4	F1, F3, F4	F3	8/12	37 %

For two of the three MZ pairs the inter-speaker variability is quite equally distributed between F1/F2 on the one hand and F3/F4 on the other, it is even higher for the lower formants. In contrast to this, within the DZ twins, F1 and F2 account for only approximately 35 % of the differences. Since size and form of the vocal tract have a strong influence on the speaker specific higher formants of a speaker, the MZ twins are expected to show less inter-speaker variability in F3 and F4 than in F1 and F2. Our results confirm the assumption that the higher formants are more dependent on physiological factors and less affected by the speakers' alternative articulation strategies than the lower formants, since the DZ twins show here more variability than the MZ twins. As indicated in the scatterplot of the DZ pair LRSR (see Figure 2b), there were no differences at all for F1, pointing again to a small impact of physiology on this formant[2].

3.2.2 Articulation

Table 5 gives an overview of the articulatory inter-speaker variability within the pairs in the production of /a/, /i/ and /u/. As described before, different coils were used as a reference for determining the articulatory targets of the vowels (cf. tab. 3). To compare the articulatory targets between the speakers we analyzed the positions of the tongue at the point in time of the determined targets. We are concentrating now on the positions of the tongue dorsum coil for

---

[2] Note that differences in F1 can also appear because of differences in loudness and hence a bigger jaw opening. Since we did not yet check for similar amplitudes in the speech signals, this has to be kept in mind for further experiments and assumptions.

/i/ and the tongue back coil for /u/ and /a/. The positions in a horizontal (X) and vertical (Y) direction of these coils at the point in time of the determined target positions were compared within the twins, and results with significant differences are given in the following table.

Table 5: Differences in target tongue positions (in cm) of the three vowels within the twin pairs, significant differences ($p < .01$) in bold.

Twin pair	zygosity	coil position	/a/ (tback)	/i/ (tdorsum)	/u/ (tback)
GSRS	MZ	vertical (Y)	**0.223**	0.029	**0.235**
		horizontal (X)	0.049	**0.168**	0.006
AFHF	MZ	vertical (Y)	**0.445**	0.044	**0.114**
		horizontal (X)	**0.250**	**0.193**	**0.165**
SLCL	MZ	vertical (Y)	**0.743**	**0.171**	**0.176**
		horizontal (X)	**0.226**	**0.428**	**0.326**
SRLR	DZ	vertical (Y)	**0.491**	0.011	**0.591**
		horizontal (X)	**0.235**	**0.126**	0.035
TGMG	DZ	vertical (Y)	0.057	0.077	0.049
		horizontal (X)	**0.824**	**0.469**	**0.583**

In general, most similarities in the target positions within the pairs were found for the vowel /i/ and least for the vowel /a/. This could be explained in terms of differences in intra-speaker variability (see Brunner et al. 2005, Mooshammer et al. 2004), with high vowels showing less articulatory variability than low vowels. If more articulatory intra-speaker variation can be expected for /a/ than for /i/, less inter-speaker variation could be assumed in siblings with similar palate shapes. In contrast to the similarities in the articulation in /i/, in the acoustical analysis, many differences were found for /i/. This points to the fact that in higher vowels less articulatory variance is necessary to achieve differences in the acoustical output. The pairs with the least inter-speaker variability in all vowels were the MZ pair GSRS and the DZ pair TGMG (but note that the differences in the vertical position of the tongue dorsum of /i/ of TGMG reached significance with $p < 0.02$). Most differences were revealed by the MZ pair SLCL. This seems to point to a minor role of similar physiology in our investigated vowels and again to the influence of a shared environment (as SLCL is the pair which spends least time together). A possible explanation could also be the factor stress, as all vowels were produced within a stressed syllable, and the effect of physiology is suggested to be stronger in unstressed syllables, and stressed syllables are supposed to be less variable and therefore more sensitive to differences. Also, the high palate of SLCL could lead to

greater intra-speaker variability in general and hence, a greater variability between the two speakers.

3.3 Influence and interaction of the factors stress and velar stop

The next section contains our results regarding our second and third hypothesis , i.e. the factor stress and the influence of a velar stop on the articulatory and acoustic inter-speaker variability of the vowel /i/ are discussed. Additionally to the production of /i/ in the stressed syllable /li/, the vowel is analyzed and compared in the syllable /gi/ in a stressed and in an unstressed position.

3.3.1 Acoustics

Note that due to our speech material, the amount of repetitions differ between the three conditions. For the formants of /i/ in the stressed and unstressed syllable /gi/ only 9 repetitions per condition (on average) could be taken into account. The mean formants of /i/ in /liebe/ could be investigated in approximately 40 repetitions. Therefore, different sample sizes were used for the Tukey Post Hoc Test, 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 6: Significant differences in formants of /i/ within the twin pairs for the three conditions: /i/ produced in the unstressed syllable /gi/, in the stressed syllable /gi/, in the stressed syllable /li/ (Tukey Post Hoc Test in R, Significance Level <.01)

Twin pair	Zygosity	Stressed /i/ in /liebe/	Stressed /i/ in /giba/	Unstressed /i/ in /hagi/
GSRS	MZ	F1, F4	F1	-
AFHF	MZ	F2, F4	F1	-
SLCL	MZ	F1, F2, F3, F4	F3 (.011)	F2, F3
SRLR	DZ	F4	-	F1
TGMG	DZ	F3	F1	F1, F2 (.012)

Comparing the groups of the MZ twins with the DZ twins, it is noteworthy that there is a clear favour of significant differences in the formants of /i/ produced in the stressed syllable /li/ for the MZ twins, but a quite equally distributed amount of differences in formants for all three conditions for the DZ twins. As we noted before, more differences were expected for the stressed /li/ condition because of the bigger sample size. In spite of this fact, the DZ twins revealed more

differences in F1 and F2 in the /gi/ conditions, mirroring the stronger influence on physiology in the first two formants of a vowel following a velar consonant.

By comparing the two /gi/ conditions, results support our hypothesis of an interaction of physiology and the factor stress: The MZ twins GSRS and AFHF that revealed a strikingly similar palatal contour (see Figure 1) show no differences in the unstressed /gi/ condition, but in the stressed condition (F1). Both DZ twin pairs reveal more inter-speaker variability in the unstressed than in the stressed syllable, pointing to auditory goals to be crucial in stressed vowels.

As was said before, it has to be considered that due to the larger sample size, the overall probability of detecting significant differences is greater in the /li/ condition than in both /gi/ conditions. Moreover, the /i/ in /liebe/ is in a non-focused position, whereas the /i/ in /hagi/ and /giba/ is part of a word under focus. Nevertheless, here, we are concentrating on a comparison between inter-speaker variability of MZ and DZ pairs and not on one between the three conditions for all speakers, and thus, the requirements are equally balanced and comparable.

3.3.2 Articulation

Table 7 shows the results of the articulatory inter-speaker variability within the twin pairs in the production of /i/ in the three conditions: in the unstressed syllable /gi/, in the stressed syllable /gi/ and in the stressed syllable /li/. We will focus on the target tongue positions, and therefore, the horizontal and vertical position of the tongue dorsum coil, to compare the articulation between the speakers.

Table 7: Differences in target tongue positions (in cm) of the vowel /i/ within the twin pairs, significant differences (p < .01) in bold.

twin	zygosity	stress	Tongue dorsum Y	Tongue dorsum X
AFHF	MZ	Unstressed /gi/	0.0628	0.1618
		Stressed /gi/	0.0404	**0.4844**
		Stressed /li/	0.0446	**0.1931**
GSRS	MZ	Unstressed /gi/	0.1516	0.0763
		Stressed /gi/	0.0452	0.1189
		Stressed /li/	0.0290	**0.1677**
CLSL	MZ	Unstressed /gi/	0.1403	0.1887
		Stressed /gi/	**0.2351**	**0.4352**
		Stressed /li/	**0.1712**	**0.4280**
SRLR	DZ	Unstressed /gi/	0.1510	0.0092

		Stressed /gi/	**0.1312**	**0.2342**
		Stressed /li/	0.0383	**0.2590**
MGTG	DZ	Unstressed /gi/	0.0628	**0.3254**
		Stressed /gi/	**0.1699**	**0.3013**
		Stressed /li/	0.0728	**0.4700**

From the results given in Table 7, it can be said that the factor stress has an impact on the articulatory inter-speaker variability in the production of /i/ within twin pairs. None of the MZ pairs shows significant differences in their target positions of the unstressed /i/ (p <. 01). Of the two DZ pairs, one pair reveals significant differences in the horizontal position of the tongue dorsum. For the stressed condition, 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 sample size would favour significance in the /li/ condition). For the MZ pairs, there was either no difference in variability (AFHF, CLSL) or more differences in the syllable /li/ (GSRS) were found. This could be 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/).

4 Discussion

In the outline of the paper, 3 hypotheses were formulated. The first one addressed the great influence of a shared environment over the identical physiology of MZ twins in general, and assumed that:

(1) DZ pairs need not naturally show more differences in their acoustic outputs regarding vowels, as they adjust their speech to each other and use auditory goals to produce vowels, but may reveal their anatomical and physiological differences in individual articulatory strategies, when this is possible.

Our results support this hypothesis, as no differences in the amount of acoustic inter-speaker variability in the formants of the stressed vowels /i/, /a/, and /u/ between MZ and DZ twin pairs were found. The MZ pair that lives together revealed the least differences regarding formants, but one DZ pair that also lives together showed less acoustic inter-speaker variability than the MZ pair that sees each other only twice a month. Nothing can be said about individual articulatory strategies. Of course, the validity of the hypothesis is limited by our speech material (i.e. the vowels /i/, /a/, /u/ in a stressed position), and due to the time-consuming articulatory recording, the relatively small group of speakers, or pairs (3 MZ and 2 DZ twin pairs), respectively. Additionally, acoustic results may

indicate that in stressed vowels MZ and DZ twins differ in their amount of variability in the higher formants. MZ pairs were less likely to show differences in F3/F4 than the DZ pairs, mirroring a stronger influence of physiology on higher formants.

To look more closely at the possible influence factors on acoustic and articulatory inter-speaker variability within twins, two more hypotheses were formulated and investigated:

(2) The physiological influences of the tongue and the shape of the palate show a greater impact on the production of consonants (especially velar stops) and hence, also on vowels in their neighbourhood.

Results suggest that articulatory and acoustic inter-speaker variability of the vowel /i/ seems to be influenced by the production of a preceding velar stop: MZ twins, who are supposed to have identical anatomy and physiology of the vocal apparatus, are more similar in their articulatory targets (vertical and horizontal tongue positions) and acoustic outputs (formants, especially F1 and F2) of the vowel /i/ when a velar consonant precedes the vowel. DZ twins that show differences in shape and steepness of the palate, reveal more inter-speaker variability in their articulatory targets and acoustic outputs of /i/ produced in the syllable /gi/ than in the syllable /li/. This result points to a stronger influence of physiology on the production of /i/ following a velar stop.

As our pilot study revealed that stress could be a possible influence factor on inter-speaker variability, our third hypothesis assumed that:

(3) More inter-speaker variability can be found in stressed than in unstressed syllables within MZ twin pairs, mirroring the greater influence of physiology on the production of an unstressed syllable.

We found supporting evidence that there is an interaction between physiology and the factor stress: Physiology seems to have a stronger influence on the production of /i/ when produced in an unstressed syllable. Both DZ twins revealed more differences in formants in the unstressed condition, and the 2 female MZ twins with the remarkable similar palatal shape showed more differences in formants in the stressed condition. In their articulatory targets the MZ pairs revealed no inter-speaker variability in the unstressed condition, but one of the DZ twins did.

To sum up, it can be said that a shared environment plays a very important role in inter-speaker variability in vowels. But, there are several factors that contribute to this variability and intensify the impact of the identical physiology of the vocal apparatus of MZ twins, namely, the production of a velar consonant preceding the vowel and the factor stress.

Acknowledgments

References

Bandura, Albert (1977). *Social Learning Theory*. Prentice-Hall, Englewood Cliffs, N.J.

Brunner, Jana, Susanne Fuchs, & Pascal Perrier (2005). The influence of the palate shape on articulatory token-to-token variability. ZAS *Papers in Linguistics 42*, 43-67.

Brunner, Jana, Susanne Fuchs, & Pascal Perrier (2009). On the relationship between palate shape and articulatory behaviour. *Journal of the Acoustical Society of America 125* (6), 3936-3949.

De Jong, Kenneth (1997). An articulatory and acoustic analysis of English flaps: Evidence for attentional reduction in production. *Journal of the Acoustical Society of America 101,* 3178.

Debruyne, Frans, Wivine Decoster, Annemie Van Gijsel, & Julie Vercammen (2002). Speaking fundamental frequency in monozygotic and dizygotic twins. *Journal of Voice 16(4),* 466-471.

Fant, Gunnar (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Fougeron, Cécile & Patricia A. Keating (1997). Articulatory strengthening at the edges of prosodic domains. *Journal of the Acoustical Society of America 101,* 3728-3740.

Fuchs, Susanne, Bernd Pompino-Marschall & Pascal Perrier (2007): Is there a biological grounding of phonology? Determining factors, optimization and communicative usage. *Proceedings of the ICPhS,* Saarbrücken. 219-224.

Hughes, Olive M. & James H. Abbs, (1976). Labial-mandibular coordination in the production of speech: Implications for the operation of motor equivalence. *Phonetica 33 (3),* 199-221.

Johnson, Keith (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson and J. Mullenix (eds.): *Talker Variability in Speech Processing,* San Diego: Academic Press, 145-166.

Keating, Patricia, A. (1990). The window model of coarticulation: Articulatory evidence. In J. Kingston & M. Beckman (eds.): *Papers in Laboratory Phonology I,* Cambridge: University Press, 451-470.

Lindblom, Björn (1983). Economy of speech gestures. In P. MacNeilage (ed): *The production of speech,* New York: Springer Verlag, 217-245.

Loakes, Deborah (2004). Front vowels as speaker-specific: Some evidence from Australian Englisch. *Proceedings of the Australian International Conference on Speech Science.* Sydney, 289-294.

Loakes, Deborah (2006). A forensic phonetic investigation into the speech patterns of identical and non-identical twins. Ph.D. Dissertation, The University of Melbourne.

Mooshammer, Christine, Pascal Perrier, Susanne Fuchs, Christian Geng & Daniel Pape (2004). An EMMA and EPG study on token-to-token variability. *AIPUK 36*, 47-63.

Mooshammer, Christine, Pascal Perrier, & Susanne Fuchs (2008): Speaker-specific patterns of token-to-token variability. *Journal of the Acoustical Society of America 123 (5),* 3076-3076.

Nolan, Francis & Tomasina Oh (1996). Identical twins, different voices. *Forensic Linguistics 3*, 39-49.

Perkell, Joseph S., Melanie L. Matthies, Mario, A. Svirsky & Michael I. Jordan (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, Joseph S., Melanie L. Matthies, Harlan Lane, Frank H.Guenther, Reiner Wilhelms-Tricarico, Jane Wozniak, & Peter Guiod (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models, *Speech Communication* 22, 227-250.

Perrier, Pascal (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.

Pierrehumbert, Janet (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In Bybee, J. and P. Hopper (eds): *Frequency effects and the emergence of linguistic structure*. Amsterdam: John Benjamins, 137-157.

Przybyla, Beata D., Yoshiyuko Horii & Michael H. Crawford (1992). Vocal fundamental frequency in a twin sample: looking for a genetic effect. *Journal of Voice 6(3),* 261-266.

Savariaux, Christoph, Pascal Perrier & Jean-Pierre Orliaguet (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

Shiller, Douglas M., Rafael Laboissière & David J. Ostry (2002). Relationship between jaw stiffness and kinematic variability in speech. *Journal of Neurophysiology 88.* 2329-2340.

Spinath, Frank M. (2005). Twin designs. In B. S. Everitt & D. C. Howell (eds.), *Encyclopedia of Statistics in Behavioral Science,* Chichester: John Wiley & Sons. 2071-2074.

Stevens, Kenneth N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In P.B. Denes & E.E. Davis (eds.), *Human Communication: A Unified View*. New York: McGraw-Hill, 51-66.

Van Lierde, Kristiane, Bart M. Vinck, Sofia De Ley, Gregory Clement & Paul Van Cauwenberge (2005). Genetics of vocal quality characteristics in monozygotic twins: A multiparameter approach. *Journal of Voice 19(4),* 511-518.

Weirich, Melanie (2009). *Twins in the lab: Reasons for speaker-specific variability*. Paper/Talk presented at Phonetics & Phonology 5, Cologne. (http://www.uni-koeln.de/philfak/phonetik/aktuell/PuP5_Praesentationen/Weirich_Zwillinge_im_Sprachlabor.pdf)

Whiteside, Sandra P. & Emma Rixon (2001). Identification of twins from pure (single speaker) and hybrid (fused) syllables: An acoustic and perceptual case study. *Perceptual and Motor Skills (1)*. 319.