ARTICULATORY AND ACOUSTIC ANALYSIS OF COARTICULATED /u/ AND /y/ PRODUCED BY FEMALE AND MALE SPEAKERS OF GERMAN

Lia Saki Bučar Shigemori, Rosa Franzke, Philip Hoole

Institute for Phonetics and Speech Processing (IPS), LMU Munich
lia | rosa.franzke | hoole@phonetik.uni-muenchen.de

ABSTRACT

Acoustic vowel spaces of female and male speakers differ in a non-uniform way, but less is known about whether this difference can be extended to coarticulatory effects. Eleven female and eight male German speakers produced single words containing tense /u/ or /y/ in bilabial, alveolar and velar contexts. Principal component analysis was used to analyse entire ultrasound image frames. For the acoustic analysis F2-values were extracted. At the acoustically determined vowel midpoint a significant fronting effect in alveolar context could be observed for both female and male speakers as a raised F2 in /u/; however in the articulatory data the effect was significant only for females. In /y/, F2 values differed significantly between bilabial and velar contexts for female speakers, while the articulatory data showed no significant effects for consonantal context. The data will be discussed from a sound change perspective and regarding articulatory-acoustic relationships.

Keywords: coarticulation, sex differences, acoustic-articulatory relations, ultrasound, vowels

1. INTRODUCTION

The aim of this study is to compare consonant to vowel coarticulatory effects between female and male speakers for phonological back and front vowels. Vowel spaces of female and male speakers differ in a non-uniform way. In phonological back vowels like /u/, F2 of females and males is similar, but in phonological front vowels like /y/, F2 of females is higher than for males (e.g. Figure 1 in [1]). While this non-uniformity can partially be attributed to physiological differences, different behavioural or sociophonetic explanations (e.g. different acoustic targets [2], or compensation for poorer harmonic sampling of spectral envelopes with higher F0 [3]) have been proposed in addition.

Several studies also reported systematic differences between females and males in speaking style, linking slower speech rate [4], more distinct vowel spaces [5] and greater spectro-temporal variation and consequently a lower degree of coarticulation with female speakers compared to male speakers [6]. It has also been suggested that female and male speakers apply different articulatory strategies due to physiological differences of their oral structures [7]. Although it has been shown that the degree of coarticulation differs for different vowels [8], to our knowledge, no study has compared the degree of coarticulation in different vowels separately for female and male speakers.

In this study, we focus on coarticulated /u/ and /y/ in German, for which it has been shown that greater coarticulatory effects in both articulation and resulting acoustics can be found in /u/ than in /y/ based on data by one female and six male speakers [9]. The results are presented as evidence for the universal preference for /u/ to diachronically change to /y/, rather than /y/ to become /u/. This so called /u/-fronting sound change has also been claimed to be led by females, although the suggested reasons are of sociolinguistic nature [10]. The literature does not link the leadership of sound change and one of its main sources, namely the misinterpretation of speaker-dependent differences in coarticulation. By looking at German, we want to explore whether differences in coarticulatory effect along the back-front dimension can be found between female and male speakers, and whether differences in the acoustic-articulatory relationship can be observed.

Given that a lower degree of coarticulation is reported for female compared to male speakers, it might be counterintuitive to suggest that we would find greater coarticulatory effects for females in our data. However, previous studies mainly looked at F2 locus equation slopes and linked a steeper slope with greater distinction between the consonant and vowel and consequently a lower degree of coarticulation. In the current study, we will instead look at data extracted at vowel midpoints. It is possible that the greater acoustic vowel space of female speakers allows them to convey consonantal context information in their vowel production while still keeping the vowel categories distinct, in contrast

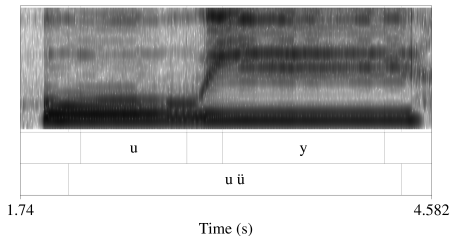

Figure 1: Example of an /u/ to /y/ continuum production by a female speaker. The spectrogram shows a Frequency range from 0 to 4000 Hz.

to male speakers who might be more likely to hypoarticulate.

2. METHODS

Eleven female and eight male German speakers were recorded in a soundproof booth. Simultaneous recordings of acoustic and articulatory data by means of ultrasound were carried out using the AAA-software [11]. The ultrasound probe was attached below the chin using the stabilizer developed by [12], which allows flexible jaw movement. A 5-8MHz, 10mm radius or a 2-4MHz 20mm radius microconvex probe were used depending on the speaker anatomy with adjusted recording settings to ensure good quality imaging but painless sessions. In most cases the frame rate was 81.6Hz. Simultaneous video recordings have been made which can be used to track the movement of the probe in relation to the head, but they have not been evaluated yet. Participants were instructed to read single words presented on a screen one by one. The target words analyzed in this study are shown in Table 1. They consist of nouns with tense vowels /u/ or /y/ occupying the stressed, initial syllable, preceded and followed by either alveolar, bilabial or velar consonants. Four repetitions were recorded per participant. In addition, participants were instructed to produce continua by moving their tongue from /u/ to /y/ or /y/ to /u/ (Example in Figure 1).

	alveolar	bilabial	velar
/u/	T**u**te	B**u**be	K**u**gel K**u**chen
/y/	T**ü**te	B**ü**bchen	K**ü**ken K**ü**gelchen

Table 1: Targetwords used in this study, column-wise for the three consonantal contexts. The target vowels /u/ or /y/ are indicated in bold.

2.1. Data preparation

Acoustic recordings were segmented using WebMAUS [13]. Formant data were extracted using LPC analysis in PRAAT by applying the Burg method with 25 ms Window length. The acoustic data were transformed into an emuDB [14] and segment boundaries and formant data were manually corrected if necessary.

To analyze the ultrasound data principal component analysis (PCA) on entire image frames was carried out using the `prcomp`-function in R [15], similarly to the method described in [16]. Our main interest was the movement of the tongue in the /u/-/y/ dimension. Therefore, the vowel continua served as the reference data, based on which the PC dimensions were calculated separately for each speaker. Ultrasound frames at the acoustically determined midpoint were then extracted and the PC scores were calculated. Upon visual inspection, for each speaker the PC dimension with the lowest number was selected which best separated /u/ and /y/ in word production and reflected a movement in opposite directions for /u/-/y/ continua compared to /y/-/u/ continua.[1] For each speaker the PC values were then scaled from -1 to 1 with reference to the lowest and highest values of the continua productions and of the target vowels, and if required flipped, so that the lower values corresponded to more /u/-like and the higher values to more /y/-like tongue configurations.

2.2. Statistics

To test the effect of vowel, consonantal context and speaker sex on articulation (scaled PC value) and acoustics (F2), linear mixed effects analyses were performed using lme4 [17] in R. CONSONANTAL CONTEXT (three levels: alveolar, bilabial, velar), VOWEL (two levels: /u/, /y/) and SPEAKER SEX (two levels: female, male) were fixed factors, and SPEAKER was a random factor (with by-SPEAKER random intercept for CONSONANTAL CONTEXT and VOWEL). To simplify the analysis the data was further split up by vowel, especially because the consonantal context is expected to have a different effect on front and back vowels. This allowed us to still compare the effect of sex and consonantal context and their interaction, but separately for /u/ and /y/. The significance of each effect was retrieved from the ANOVA type III table with p-values calculated using Satterthwaite's method. Effects were considered to be significant when $p < 0.05$. Post-hoc Tukey-tests were carried out using the emmeans-package [18].

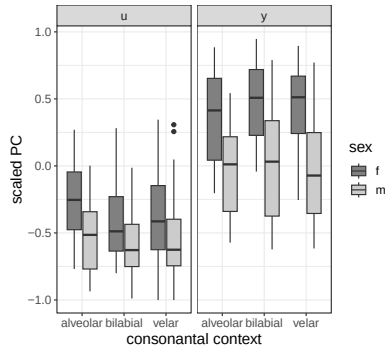

Figure 2: Scaled PC values of ultrasound images at the acoustically determined vowel midpoint, separately for consonantal contexts (alveolar, bilabial and velar), grouped by speaker sex (female in dark grey and male in light grey), left panel for /u/ and right panel for /y/.

3. RESULTS

3.1. Articulatory data

The scaled PC values of the ultrasound frames extracted at the acoustically determined vowel midpoint are illustrated in Figure 2 for the consonantal contexts (alveolar, bilabial, velar) and speaker sex, in the left panel for vowel /u/ and in the right panel for vowel /y/. Keep in mind that the speaker normalization of the PC values was carried out based on the ultrasound frames from the continua productions and the frames extracted at target vowel midpoints. Negative values are more /u/ like and positive values more /y/-like along the /u/-/y/ dimension. Figure 2 confirms this. In addition, /u/ in alveolar context has a slightly higher PC value than for bilabial or velar contexts. It seems that in word productions the difference between /u/ and /y/ was less extreme than when speakers were producing vowel continua, especially for /y/, where the value 1 was not reached in any context, neither by female nor by male speakers.

The full model revealed that the main effects of SPEAKER SEX ($F[1,19.0]=9.4$, $p<0.01$) and VOWEL ($F[1,19.3]=147.0$, $p<0.001$) were significant but the effect of CONSONANTAL CONTEXT ($p=0.8$) was not. In addition, the interactions between SPEAKER SEX and VOWEL ($F[1,19.3]=4.5$, $p<0.05$) and between VOWEL and CONSONANTAL CONTEXT ($F[2, 539.8]=10.2$, $p<0.001$) were significant.

3.1.1. Articulatory data of /u/

The linear mixed effect model for the back vowel /u/ revealed that the main effect of CONSONANTAL CONTEXT was significant ($F[2,19.0]=6.4$, $p<0.01$). There was only a trend effect for SPEAKER SEX ($F[1,19.0]=4.3$, $p=0.05$) and the interaction between CONSONANTAL CONTEXT and SPEAKER SEX was not significant. The post-hoc test revealed that, for females, the difference between /u/ in alveolar and bilabial contexts was significant ($p<0.01$), while the difference between alveolar and velar contexts ($p=0.09$) and between bilabial and velar contexts ($p=0.5$) were not. For male speakers, none of the differences were significant ($p>0.5$). In alveolar context, scaled PC values for females were significantly higher than for males ($p<0.05$), while the scaled PC values for females and males did not differ significantly between bilabial and velar contexts ($p=0.1$).

3.1.2. Articulatory data for /y/

For the front vowel /y/, the model revealed that the effect of SPEAKER SEX on the scaled PC values was significant ($F[1,19.0]=10.8$, $p<0.01$). The effect of CONSONANTAL CONTEXT ($F[2,19.3]=2.8$, $p=0.09$) and the interaction between CONSONANTAL CONTEXT and SPEAKER SEX ($F[2,19.3]=0.09$, $p=0.9$) were not significant. The post-hoc test confirmed that the difference between female and male speakers was significant for all consonantal contexts, while no paired comparisons between consonantal contexts were significant, neither for females nor for males.

3.2. Acoustic data

F2 values extracted at vowel midpoints are illustrated in Figure 3, for the consonantal contexts (alveolar, bilabial, velar) and speaker sex, in the left panel for vowel /u/ and in the right panel for vowel /y/. As expected, F2 values are lower for /u/ than for /y/, and a difference between female and male speakers can be seen for /y/ but not for /u/. In /u/, F2 is slightly higher for the alveolar context than for the bilabial and velar contexts, suggesting a fronting effect due to coarticulaton.

The full model revealed that the effects of VOWEL ($F[1,576.4]=8731.6$, $p<0.001$), CONSONANTAL CONTEXT ($F[2,32.5]=11.2$, $p<0.001$) and SPEAKER SEX ($F[1,19.0]=14.7$, $p<0.05$) were significant. In addition, the interactions between VOWEL and CONSONANTAL CONTEXT ($F[2,576.4]=15.7$, $p<0.001$) and between VOWEL and SPEAKER SEX ($F[1,576.43]=170.8$, $p<0.001$) were significant.

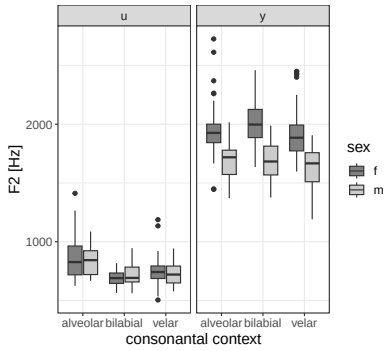

Figure 3: F2 values in Hz at the acoustically determined vowel midpoint, separately for consonantal contexts (alveolar, bilabial and velar), grouped by speaker sex (female in dark grey and male in light grey), left panel for /u/ and right panel for /y/.

3.2.1. Acoustic data for /u/

The model for the F2 values of the vowel /u/ revealed that the main effect of CONSONANTAL CONTEXT ($F[2,29.7]=18.3$, $p<0.001$) was significant, but the effect of SPEAKER SEX ($F[1,19.1]=0.01$, $p=0.9$) and the interaction effect between CONSONANTAL CONTEXT and SPEAKER SEX ($F[2,29.7]=2.3$, $p=0.1$) were not. The post-hoc test revealed that for female speakers, the differences between alveolar and bilabial context ($p<0.001$), alveolar and velar context ($p<0.05$) and bilabial and velar context ($p<0.05$) were all significant, while for male speakers, only the difference between alveolar and velar context ($p<0.05$) was significant, and the differences between alveolar and bilabial context ($p=0.06$) and between bilabial and velar context($p=1$) were not. F2 differences between female and male speakers were not significant regardless of consonantal context.

3.2.2. Acoustic data for /y/

For the F2 values of the vowel /y/, the model revealed that, the main effects CONSONANTAL CONTEXT ($F[2,18.8]=6.9$, $p<0.01$) and SPEAKER SEX ($F[1,19.0]=18.4$, $p<0.001$) were significant, but the interaction between CONSONANTAL CONTEXT and SPEAKER SEX ($F[2,18.8]=0.9$, $p=0.4$) was not. The post-hoc test revealed that F2 values differed significantly only between bilabial and velar contexts for female speakers ($p<0.05$).

4. SUMMARY AND DISCUSSION

The aim of this study was to explore differences in coarticulation between female and male speakers, which might also depend on segmental factors and which might reveal differences in acoustic-articulatory relations between female and male speakers. Although the interaction between CONSONANTAL CONTEXT and SPEAKER SEX was not significant in any of the models, a look at the post-hoc tests revealed that if the main effect of CONSONANTAL CONTEXT was significant, the differences between consonantal contexts were greater for female speakers than for male speakers. The results also corroborate findings from previous literature [9] that the coarticulatory effect in the front-back dimension is greater for /u/ than for /y/.

The articulatory data suggests that male speakers produce less fronted /y/ in words compared to when they are instructed to produce continua. This can be interpreted as evidence that indeed male speakers hypoarticulate more compared to females, in the sense that their articulation is more centralized. In contrast, female speakers produce more distinct vowels but at the same time the within category variation due to consonantal context is also clearer.

A visual inspection of the patterns across speakers revealed that the very low scaled PC values in /y/ for the articulatory data can be attributed to four of the eight male speakers. However, when we compared the F2 values in word productions with F2 values in continua productions separately for each speaker, no such differences could be found. Thus, while some speakers articulated the /u/-/y/ contrast differently in continua than in words, this was not reflected in their acoustic output. One possibility is that for these speakers variation other than only in the front-back dimension was captured in the PC, such as jaw opening or displacement of the probe due to different posture during the continua production. A second possibility is that in continua productions these speakers produced /y/ with a more fronted tongue, which however did not affect F2. In their attempt to reproduce Fant's nomograms with human speakers, [19] report a flattening of the F2 curve for anterior constriction locations for high vowels. They suggest that in the region of high front vowels the tongue body flattens with more anterior constriction location and the length of the back cavity, which F2 is associated with, does not increase.

In this study we showed that gender, vowel type and consonantal context all contribute to acoustic and articulatory variation differently or to a different degree, providing bias which can potentially lead to sound change. To better understand mechanisms of sound change, looking into fine-grained contexts and bringing together what has been shown in previous studies as we tried to do here seems fruitful.

5. ACKNOWLEDGEMENTS

The research was funded by the project SoundAct which has received funding from the European Research Council (ERC) under the European Union's Horizon Europe research and innovation programme (grant agreement No. 101053194). We thank Anna Ratzinger for her help with the post-processing, Jessica Siddins for proofreading, and colleagues from the IPS for various inputs at different stages.

6. REFERENCES

[1] A. Simpson and C. Ericsdotter, "Sex-specific differences in f0 and vowel space," in *16th International Congress of Phonetic Sciences*, 2007, pp. 933–936.

[2] G. Fant, "Non-uniform vowel normalization," *STL-QPSR*, vol. 16, no. 2–3, pp. 001–019, 1975.

[3] R. L. Diehl, B. Lindblom, K. A. Hoemeke, and R. P. Fahey, "On explaining certain male-female differences in the phonetic realization of vowel categories," *Journal of Phonetics*, vol. 24, pp. 187–208, 1996.

[4] D. Byrd, "Preliminary results on speaker-dependent variation in the TIMIT database," *The Journal of the Acoustical Society of America*, vol. 92, no. 1, pp. 593–596, 1992.

[5] H. Traunmüller, "Paralinguistic variation and invariance in the characteristic frequencies of vowels," *Phonetica*, vol. 45, no. 1, pp. 1–29, 1988.

[6] F. Herrmann, S. P. Cunningham, and S. P. Whiteside, "Speaker sex effects on temporal and spectro-temporal measures of speech," *Journal of the International Phonetic Association*, vol. 44, no. 1, pp. 59–74, 2014.

[7] A. P. Simpson, "Gender-specific articulatory–acoustic relations in vowel sequences," *Journal of Phonetics*, vol. 30, no. 3, pp. 417–435, 2002.

[8] H. van den Heuvel, B. Cranen, and T. Rietveld, "Speaker variability in the coarticulation of /a, i, u/," *Speech communication*, vol. 18, no. 2, pp. 113–130, 1996.

[9] J. Harrington, P. Hoole, F. Kleber, and U. Reubold, "The physiological, acoustic, and perceptual basis of high back vowel fronting: Evidence from German tense and lax vowels," *Journal of Phonetics*, vol. 39, no. 2, pp. 121–131, 2011.

[10] W. Labov, "The intersection of sex and social class in the course of linguistic change," *Language variation and change*, vol. 2, no. 2, pp. 205–254, 1990.

[11] Articulate Instruments Ltd, "Articulate Assistant Advanced user guide: Version 2.14," Edinburgh, UK, 2012.

[12] D. Derrick, C. Carignan, W.-R. Chen, M. Shujau, and C. Best, "Three-dimensional printable ultrasound transducer stabilization system," *The Journal of the Acoustical Society of America*, vol. 144, no. 5, pp. EL392–EL398, 2018.

[13] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language*, vol. 45, pp. 326–347, Sep. 2017.

[14] R. Winkelmann, J. Harrington, and K. Jänsch, "EMU-SDMS: Advanced speech database management and analysis in R," *Computer Speech & Language*, vol. 45, pp. 392–410, Sep. 2017.

[15] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022. [Online]. Available: https://www.R-project.org/

[16] P. Hoole and M. Pouplier, "Öhman returns: New horizons in the collection and analysis of imaging data in speech production research," *Computer Speech & Language*, vol. 45, pp. 253–277, 2017.

[17] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[18] R. V. Lenth, *emmeans: Estimated Marginal Means, aka Least-Squares Means*, 2022, R package version 1.8.3. [Online]. Available: https://CRAN.R-project.org/package=emmeans

[19] P. Ladefoged and A. Bladon, "Attempts by human speakers to reproduce Fant's nomograms," *Speech Communication*, vol. 1, no. 3-4, pp. 185–198, 1982.

[1] Sometimes two PC dimensions together seemed to best capture the contrast between /u/ and /y/, however, picking only one PC dimension simplified following speaker normalization and the PC dimension with the lower number explains the greater variance in the data.