ARTICULATORY-ACOUSTIC RELATIONS IN GERMAN VOWELS

Philip Hoole
Institut für Phonetik und Sprachliche Kommunikation
Munich University, Germany
hoole@phonetik.uni-muenchen.de

ABSTRACT

This work examines whether articulatory-acoustic relationships familiar from modelling studies are actually observable in speakers' utterances. Using electromagnetic articulography the relation between formant frequencies and constriction location and size was examined in /i/, /e/, /y/ and /ø/. F2 vs.constriction location: Correlations were close to zero. This applied to both unrounded and rounded vowels. F2 vs. constriction size: Correlations were moderately strong. However, the slope of the relationship was surprisingly flat, around 20 Hz/mm. F1 vs. constriction size: Correlations were very strong, but with flatter regression slope for rounded than unrounded vowels. Some articulatory compensation for the flatter slope in rounded vowels was observed. Implications for future work are discussed.

1. INTRODUCTION

The last half-century has seen the development of powerful acoustic models of the vocal tract. However, only in about the last 10 years has it become possible to routinely measure tongue position in large corpora. Consequently, there is still a clear lack of attempts to compare the predictions of models with the articulatory-acoustic relations actually observable in speakers' utterances (but see, for example, the “human articulatory synthesizer” approach followed by Ladefoged & Bladon [5]). A more relevant exception for present purposes is to be found in Beckman et al., [1] (picking up, in turn, on previous work in Perkell & Cohen [6]). These authors considered to what extent patterns of articulatory and acoustic variability in vowel production were consistent with expectations derived from Stevens' quantal theory [7], for example that point vowels are relatively insensitive to perturbations of constriction location. For the vowel /i/, in particular, it was indeed found that the correlation between F2 and constriction location was much weaker than that between F2 and constriction degree - even though there was much more articulatory variability along the constriction-location dimension. In the present paper we examine whether comparable results can be found for palatal vowels in German. One specific area of interest in German is the presence of both rounded and unrounded front (palatal) vowels. As Wood [8] has pointed out, the rounding opposition typically involves not just lip-rounding as such, but also differences in tongue and larynx position (lower tongue and larynx in /y/ than /i/). Wood suggests, rather in the spirit of quantal theory, that this combination of articulatory manoeuvers gives the rounded vowels a similar insensitivity to perturbations of constriction location as is to be found in the unrounded vowels. To our knowledge, however, a direct comparison of articulatory-acoustic relations in rounded and unrounded vowels based on synchronized articulatory and acoustic analyses of multiple-speaker, multiple-utterance datasets has not yet been performed.

2. PROCEDURE

Electromagnetic articulography (AG100, Carstens Medizinelektronik) was used to record articulatory data from four fleshpoints on the tongue, together with lower-lip and jaw (details in Hoole [3]). Seven speakers recorded all the monophthongal vowels of German in /pVp/, /tVt/ and /kVk/ consonantal contexts. Five randomized repetitions of each CV combination were collected. The target sequences were embedded in a constant carrier phrase. Synchronized audio data were recorded on DAT tape and digitally downsampled to 16 kHz.

For the present study the following subset of the data was chosen for analysis:

1) The long (tense) vowels /i/, /y/, /e/ and /ø/; as will be seen below they offer some fairly systematic variation in constriction location and size within a conveniently homogeneous group of vowels.
2) The six male speakers out of the total of seven speakers recorded to date.
3) The second tongue sensor from the front. Inspection of the complete dataset indicated that this was typically the sensor closest to the main constriction for these palatal vowels.

2.1. Preprocessing: Articulation

The articulatory data was corrected for head-movements and was mapped to a coordinate system based on each subject's occlusal plane. The data was also lowpass-filtered with a cutoff frequency of 35 Hz. The contour of the hard palate was derived from dental impressions for each subject. The sensor data of each subject was then rotated so that the average orientation of the palate was horizontal in the vicinity of the sensor data. The resulting rotated dimensions will be referred to as "constriction location" (parallel to the hard palate) and "constriction degree" (perpendicular to the hard palate).

2.2. Preprocessing: Acoustics

Estimates of the first 3 formant frequencies were calculated under interactive control using an autocorrelation-based LPC method (64ms Hamming window) applied to a frame centered as close to the midpoint of the vowel as was consistent with stable formant values (F3 will not be considered further here, but will be discussed in a more extensive report). Articulatory data was extracted at the time corresponding to the midpoint of the acoustic analysis frame.

Speaker C

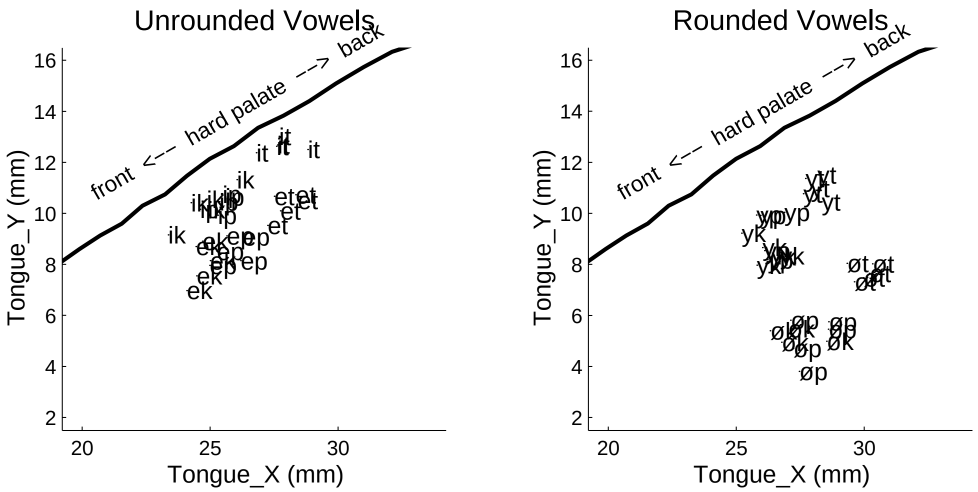

Figure 1. Raw tongue-sensor position data for all repetitions of the four target vowels for one speaker. Symmetric consonant context indicated by second letter at each data point.

Average of 6 male speakers

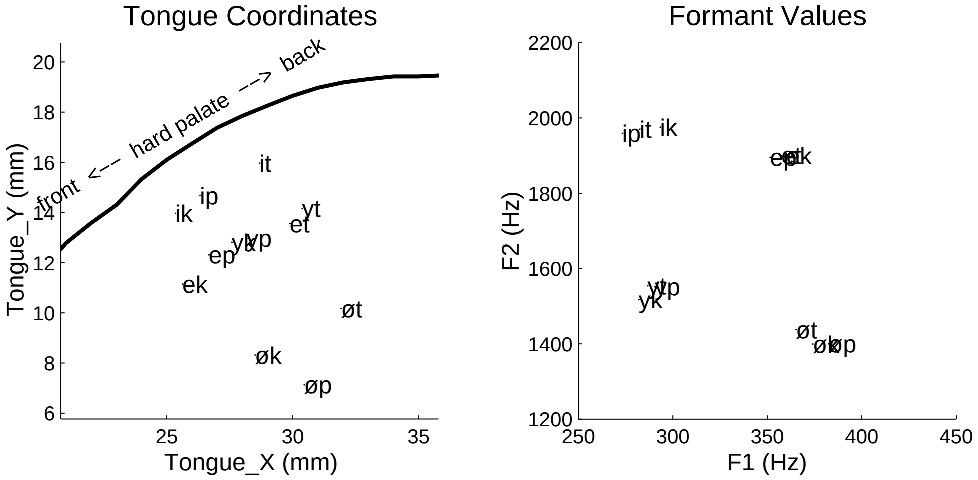

Figure 2. Tongue-position (left panel) and formant frequencies (right panel) averaged over all speakers for each CV combination.

3. RESULTS

Fig. 1 shows a scatter plot of the raw articulatory data (i.e before rotation) for a typical subject. Roughly speaking one can say that the variability of each individual vowel (over contexts and repetitions) is predominantly played out along the constriction-location dimension. These intra-phonemic differences are comparable in magnitude to the inter-phonemic differences, e.g between /i/ and /e/ or between /y/ and /ø/, that mainly involve the constriction-degree dimension. This general picture can also be seen in the left panel of Fig. 2 in which the data for each CV combination averaged over all speakers and repetitions are shown. The right panel of this figure shows the averaged formant values. The basic picture is that rounded/unrounded pairs differ strongly in F2, but very little in F1, while conversely the height pairs (/i/ vs. /e/, and /y/ vs. /ø/) differ strongly in F1 but little in F2. One immediate difference between the articulatory and acoustic results concerns contextual effects. These are very clear and systematic in articulatory terms (the tongue sensor shifts from front to back in the order /k, p, t/) but much less so in the F1/F2 space.

We will now look in detail at specific articulatory-acoustic relationships in terms of the correlation coefficient and of the gradient of the relationship, i.e estimated formant change in Hz for 1mm articulatory change. All analyses were first performed separately over subjects, and then averaged.

3.1. F2 vs. constriction location

This analysis, corresponding to the classic question as to whether palatal vowels are acoustically insensitive to perturbations in constriction location, was performed separately for each of the four vowel categories. (The sign of the relationship was set so that a positive correlation would correspond to an increase in F2 as the constriction moves away from the glottis and towards the mouth along the hard palate.) Fig. 3 shows the means and standard deviations over speakers for the 4 vowel categories. The traditional assumption is clearly confirmed: The correlation coefficients (and regression gradients) cluster around zero. In confirmation of Wood's hypothesis there is absolutely no evidence that rounded vowels tend to behave differently from the unrounded vowels.

3.2. F2 vs. constriction degree

For this analysis /i/ and /e/ were combined into a larger unrounded category, and /y/ and /ø/ into a rounded category. The sign of the results shown in Fig. 4 has been set such that a positive correlation would correspond to an increase in F2 with increase in size of the constriction (inverse of tongue height). As was to be expected from the work of Beckman et al. the correlation between F2 and constriction degre is clearly stronger than that between F2 and constriction location. The average absolute value of the correlation coefficient comfortably exceeds 0.5. However, the gradient of the relationship, which again is similar for rounded and unrounded vowels, is actually quite weak, being around -20 to -25 Hz/mm. (The size of the standard deviations indicates some interspeaker differences; these are currently under more detailed consideration). The speaker shown in Fig. 5 in Beckman et al. would have a value of about -45 Hz/mm for this slope. The weakness of this slope in our data is somewhat unexpected in the light of Gunnilstam's [2] interesting proposal (based on his interpretation of Fant's nomograms) to shift the emphasis for palatal vowels away from their insensitivity to constriction perturbations onto their simple linear response to changes in constriction size - this latter feature making them eminently learnable. It seems doubtful whether the weak slope found here would provide a reliable foundation on which to learn, for example, the /i/ vs. /e/ distinction.

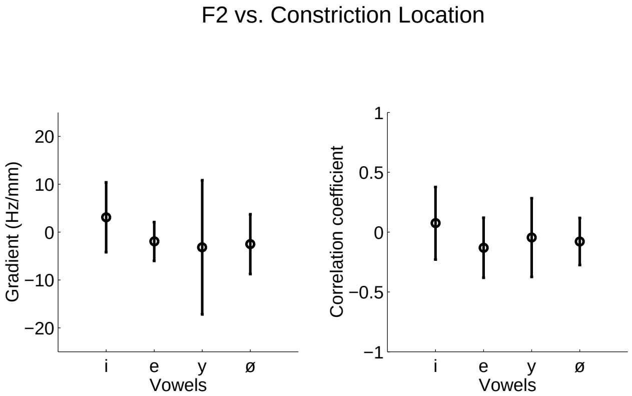

Figure 3.Mean and standard deviation of regression gradients (left panel) and correlation coefficients (right panel) for F2 vs. constriction location (averaged over subjects; n=6).

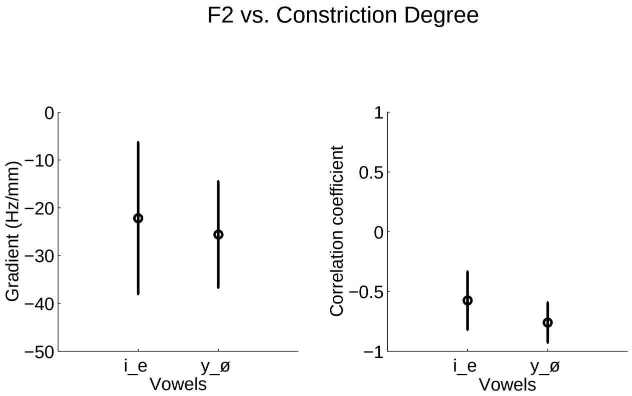

Figure 4. F2 vs. constriction degree (over details as in Fig. 3).

3.3. F1 vs. constriction degree

This analysis was intended to directly parallel the previous one (F2 vs. constriction degree). It gave by far the strongest and (over speakers) most consistent correlation coefficients of all the analyses presented here (cf. Fig. 5). This in itself is not surprising since the increase in F1 with increased mouth opening is probably the simplest of all known articulatory-acoustic relationships. However, two points ought to be made. Firstly, the gradients are very close in magnitude to those found for F2 vs. constriction degree, viz. about 20 Hz/mm. This serves to emphasize the point that modulation of F2 by constriction degree is probably not very salient perceptually since 20Hz corresponds to about 0.2 Bark at the frequency of F1 but to only about 0.07 Bark at the frequency of F2. The second point relates to the fact that there is a fairly consistent difference in gradient betwen rounded and unrounded vowels, with rounded vowels showing a shallower gradient. This is probably to be expected from acoustic theory, since rounded vowels exhibit more open (i.e closer to neutral) constrictions than unrounded. However, it can be observed from Fig. 2 that there is a greater difference in constriction degree between /y/ and /ø/ (about 5.2 mm on average) than between /i/ and /e/ (about 2.8 mm on average). Thus for the rounded pair the greater articulatory distance tends to compensate for the reduced acoustic sensitivity to changes in constriction size. This can be captured by comparing the following two ratios:

$$R1 = \frac{\left(\text{regression gradient for } /\,\text{i}\,/ \text{ and } /\,\text{e}\,/ \text{ tokens}\right)}{\left(\text{regression gradient for } /\,\text{y}\,/ \text{ and } /\,\text{ø}\,/ \text{ tokens}\right)}$$

$$R2 = \frac{\left(\text{constriction size difference } /\,\text{y}\,/ \text{ vs. } /\,\text{ø}\,/\right)}{\left(\text{constriction size difference } /\,\text{i}\,/ \text{ vs. } /\,\text{e}\,/\right)}$$

R1 and R2 amounted to 1.62 and 1.85 respectively, indicating quite a close trade-off (at least averaged over speakers) between articulatory-acoustic sensitivity and constriction size differences. The overall result of this was already to be seen in Fig. 2, i.e the difference in F1 between /i/ and /e/ is very similar to that between /y/ and /ø/. Thus it seems that speakers may plan their articulation to take the difference in the articulatory-acoustic slope into account.

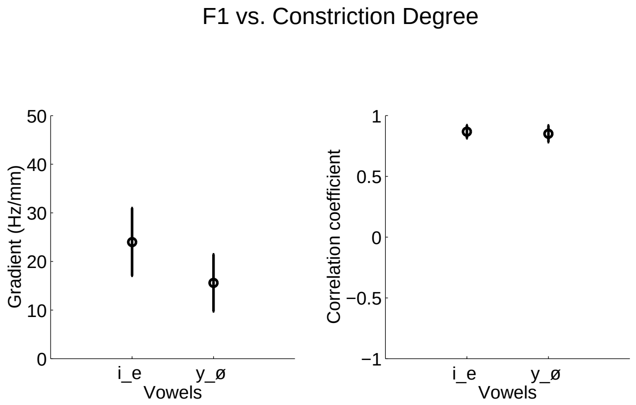

Figure 5. F1 vs. constriction degree (over details as in Fig. 3)

4. OUTLOOK

The analyses performed so far indicate that the articulatory-acoustic relationships found in this multiple-speaker, multiple-utterance dataset qualitatively conform with expectations from modelling studies, and suggest that speakers may indeed exploit those articulatory strategies that have been claimed to be acoustically advantageous. However, this conclusion is based on patterns observed when averaging over speakers. In fact, evidence of fairly substantial quantitative differences between speakers was found (e.g F2 vs. Constriction Degree). The present study will be expanded by considering to what extent individual speakers may depart from the supposedly optimal articulatory-acoustic strategy. To achieve this aim it will be necessary to examine more precisely the relationships between fleshpoint displacements and the resulting change in vocal tract cross-sectional areas. To this end comprehensive NMRI datasets are currently being acquired. In addition, we have also started to analyze speaker-specific patterns in vertical positioning of the larynx [4].

A further direction for future work will be to expand the analysis to the remaining vowels of the German vowel system. In particular, the lax vowels, being more strongly coarticulated than their tense counterparts examined here, should allow more detailed comparison of contextual effects in the articulatory and acoustic domains. For this aim, too, it will be necessary to move beyond the analysis of a single fleshpoint to include (at least) the behaviour of the tongue tip as an independent articulator.

ACKNOWLEDGEMENTS

Work supported by DFG Ti 69/29-3

REFERENCES

[1] Beckman, M., Jung, T-P., Lee, S-H., de Jong, K., Krishnamurthy, A., Ahalt, S., Cohen, K. and Collins, M. 1995. Variability in the production of quantal vowels revisited. *J. Acoust. Soc. Am.*, 97, 471-490.

[2] Gunnilstam, O. 1974. The theory of local linearity. *J. Phonetics*, 91-108.

[3] Hoole, P. 1996. Issues in the acquisition, processing, reduction and parameterization of articulographic data. *FIPKM*, 34, 158-173

[4] Hoole, P. and Kroos, C. 1998. Control of larynx height in vowel production. *Proc. 5th Int. Conf. Spoken Lang. Processing*, 2, 531-534

[5] Ladefoged, P. and Bladon, A 1982. Attempts by human speakers to reproduce Fant's nomograms. *Speech Communication*, 9(3), 231-238.

[6] Perkell, J.S. and Cohen, M.H. 1989. An indirect test of the quantal nature of speech in the production of the vowels /i/, /a/ and /u/. *J. Phonetics*, 17, 123-133.

[7] Stevens, K.N. 1989. On the quantal nature of speech. *J. Phonetics*, 17, 3-45.

[8] Wood, S. 1986. The acoustical significance of tongue, lip, and larynx maneuvers in rounded palatal vowels. *J. Acoust. Soc. Am.*, 80, 391-401.