coustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Motor equivalent strategies in the production of /u/ in perturbed speech

Jana Brunner[a], Phil Hoole[b] and Pascal Perrier[c]

[a]Humboldt-Universität Berlin / Gipsa-lab, INP Grenoble, ZAS, Schützenstr. 18, 10117 Berlin, Germany
[b]Institut für Phonetik und Sprachverarbeitung, Schellingstr. 3, 80799 München, Germany
[c]ICP/GIPSA-lab, INPG, 46 Avenue Félix Viallet, 38031 Grenoble Cedex 01, France
brunner@zas.gwz-berlin.de

Several articulatory strategies are available during the production of /u/, all resulting in a similar acoustic output. /u/ has two main constrictions, at the velum and at the lips. A perturbation of either constriction can be compensated at the other one, e.g wider constriction at the velum by more lip protrusion, wider lip opening by more tongue retraction. This study investigates whether speakers use this relation under perturbation.
Six speakers were provided with palatal prostheses which were worn for two weeks. Speakers were instructed to make a serious attempt to produce normal speech. Their speech was recorded via EMA and acoustics several times over the adaptation period. Formant values of /u/-productions were measured. Velar constriction width and lip protrusion were estimated. For four speakers a correlation between constriction width and lip protrusion was found. A negative correlation between lip protrusion and F1 or F2 could sometimes be observed, but no correlation occurred between constriction size and either of the formants.
The results show that under perturbation speakers use motor equivalent strategies in order to adapt. The correlation between constriction size and lip protrusion is stronger than in studies investigating unperturbed speech. This could be because under perturbation speakers are inclined to try out several strategies in order to reach the acoustic target and the co-variability might thus be greater.

1 Introduction

Some sounds such as rounded back vowels or the American English /r/ allow for several possibilities to produce them. In rounded back vowels, for example, speakers can vary lip protrusion, lip opening, larynx height and tongue height in dependence on each other. Several vocal tract configurations can thus lead to the same acoustic output. Lowering the larynx, for example, lowers the formants. This effect can also be reached by protruding the lips. Similarly, raising and retracting the tongue leads to a lower first and second formant, which can be compensated for by less lip protrusion. "Rounded" back vowels can even be produced with open lips and no rounding at all when the velar constriction is retracted.

To give another example, the American English /r/ can be produced with a bunched tongue shape or a retroflex tongue. Both ways to articulate the sound result in similar acoustic outputs (e.g. (2), (9), but see (1) for slight acoustic differences).

Different articulatory strategies which lead to the same acoustic output, as the ones discussed for rounded back vowels and /r/ are called motor equivalent strategies. A number of studies show that speakers in fact use these motor equivalent strategies in their speech. The study presented in (3) shows that during several productions of /u/ speakers vary in their application of tongue position and lip rounding. Three of the four speakers presented in this study moved the tongue to a higher and more retracted position when they had little lip protrusion. When they had more lip protrusion they produced a wider constriction by lowering and fronting the tongue. The acoustic output of all these productions was similar.

The study presented in (3), however, shows a problem when investigating motor equivalent strategies: Even if there are clear interspeaker differences in the way speakers produce /u/, the effect is normally not as obvious within a speaker. Speakers seem to have a preference for a certain lip protrusion in /u/. Similarly, investigating /r/ one often finds speakers using a bunched OR a retroflex strategy exclusively, or one finds speakers alternating between the two according to context (cf. (2)). A reason for the fact that different strategies can often only be observed across speakers or at least across contexts, but hardly ever within a given speaker and context might be that speakers are used to a certain articulatory strategy involving little articulatory effort in a given context.

Perturbation experiments have shown that if one forces speakers to explore other articulatory possibilities by confronting them to a perturbation, the articulatory changes carried out in order to keep the acoustic output constant can be quite remarkable even if the context is not changed. For example, in the perturbation experiment presented in (6) and (7) speakers' lips were hold open and speakers were asked to produce the vowel /u/ which is usually rounded in unperturbed speech. Some of the speakers managed to do so by retracting the constriction.

To give another example, results presented in (8) for a short term experiment on North American English /r/, where speakers vocal tract shape was changed by a palatal prosthesis, show that some speakers who used to have bunched tongue shapes switched to retroflex shapes under perturbation.

Thus, speech perturbation seems to be a useful means to explore motor equivalent strategies. The study presented here shows results of a long term perturbation experiment where speakers' palate shape was changed in a similar way as it was done in (8). The articulation of /u/ was investigated. As in the two perturbation experiments discussed above speakers in our experiment were thus encouraged to use other strategies leading to the same acoustic result, either because the old strategy did no longer lead to the correct result or because it involved too much articulatory effort.

Our hypothesis was that being confronted with a palatal prosthesis lowering the palate speakers should at first lower their tongue in order to keep the same constriction size. The degree of lip protrusion should stay about the same. Later they might notice that the acoustic output does not match with the one they used to have in the unperturbed condition, or they might look for a strategy involving less articulatory effort. Thus, they should use different degrees of lip protrusion and vary the constriction size correspondingly. Over sessions one should therefore find a positive correlation between lip protrusion and constriction size. Furthermore, if speakers make efficient use of motor equivalence, no correlation between either of the two articulatory parameters and one of the formants should be found.

2 Methods

A two-week perturbation experiment was carried out with six German speakers. Their articulation was perturbed by a palatal prosthesis. There were two types of prostheses, the first one moved the alveolar ridge posteriorily ("alveolar palate"), the second one made the palate flatter and lower by filling out the palatal arc ("central palate"). The thickness of the palates differed from speaker to speaker and depended on the anatomical conditions of the speaker. However, they all had a maximal thickness of around 1 cm. Four speakers (A1 to A4) were recorded with an alveolar prosthesis and two (C1 and C2) with a central prosthesis. Speakers were asked to wear the prosthesis all day for two weeks and to make a serious effort to improve their speech.

Speakers' articulator movements and the resulting acoustic signal were recorded via electromagnetic articulography. Four speakers were recorded with the AG 100 at the ZAS Berlin and two speakers were recorded with the AG 500 at the IPS Munich. On the first day of the experiment three different sessions were recorded. First, speakers were recorded without the prosthesis (session *nopert*, meaning "no perturbation") in order to record their habitual articulation. In the second session, the artificial palate was inserted and speakers' auditory feedback was masked with white noise (session *wnpert*, "white noise - perturbed"). In the third session speakers were recorded with auditory feedback available (session *pert*, "perturbed"). After one week adaptation time speakers returned to the laboratory and were recorded with the prosthesis in place (session *1wpert*, "one week, perturbed"). A final perturbed session was recorded after two weeks (session *2wpert*, "two weeks, perturbed"). Then speakers removed the prosthesis and were recorded with their normal vocal tract shape (session *2wnopert*, "two weeks, no perturbation") in order to investigate possible after effects.

Since the experimental setup, especially for the first session, is extremely complex and difficult to coordinate the session with auditory feedback masking (*wnpert*) was left out for one speaker who was recorded towards the beginning of the study (speaker A2).

Eight sensor coils were attached to the subject. Three were glued to the tongue, one around a centimeter behind the tongue tip, one at the part opposite the border between hard and soft palate and the third one in the middle between these two. Another sensor was placed below the lower incisors in order to track jaw movements. In order to record lip movements, two further sensors were glued to the upper and the lower lip. Two sensors at the upper incisors and the bridge of the nose served as reference sensors to compensate for head movements.

Since the sensors had to be glued to the articulators anew on each recording day, the positions of the tongue sensors differed slightly for different recording days. Whereas it was easy to find approximately the same place for the sensor at for example the lower incisors (midsagittaly below the incisors), it was hard to find landmarks on the tongue which could serve as points of orientation for finding the same position in a further recording. Photos of the tongue with the sensors on it were taken and the distance between the sensors was measured, in order to find the positions again, however, this method turned out to be not completely satisfactory.

After the recording a number of preprocessing steps were carried out which included correction algorithms for head movement, filtering of the data, rotation and translation of the position data and synchronisation with the acoustic data. The acoustic data were downsampled to 24 kHz.

The target vowel /u/ was recorded in the nonsense word /'tu:ta/ which was embedded in the carrier phrase *Ich sah ... an.* (I looked at ...). The sentences were in general repeated 20 times per session in randomised order with other items. Thus, over sessions around 120 productions of /u/ were recorded.

As a first analysis step an acoustic segmentation of the material was carried out. /u/ was segmented from the onset of the second formant to the offset of the second formant. These segmentations were later used in order to carry out semiautomatic formant measurements of F1 and F2 in each /u/-production. In these measurements the formants were at first calculated fully automatically by just adapting the LPC order according to the sex of the speaker. Afterwards, the measurements were corrected manually by determining the frequency ranges for each formant in each production and calculating formants for these ranges. Measuring the formants of one speaker (A4) turned out to be problematic since she had a rather breathy voice.

Lip protrusion was measured by calculating the difference in the horizontal dimension between the upper lip sensor and the upper incisor sensor at the middle of the interval for which the formants had been measured. By doing this it was assumed that the upper incisor sensor and the upper lip sensor were glued in about the same location on different recording days, which seemed to be basically true.

The constriction size was measured as the shortest Euclidean distance between the palate and the tongue. In order to carry out this measurement, the palatal contour for each session was estimated by plotting all positional data of the complete speech material recorded in the session. Then the palatal contour was estimated at the upper border of these tongue positions. In order to estimate the tongue contour, spline functions going through the three tongue sensors were calculated at the acoustic target position. A comparison of the position of just the tongue back sensor (as was done in (3)) was not possible since the sensor position on the tongue varied between different experimental sessions. Afterwards, the constriction size was estimated as the smallest Euclidean distance between tongue contour and palate contour.

Bivariate Pearson correlations for the four parameters (F1, F2, lip protrusion and constriction size) were calculated with SPSS 15.0.

3 Results

At first the results for the formant measurements will be discussed. Afterwards, results of the measurements of the articulatory parameters will be presented. In section 3.3 the results of the correlation measurements can be found. Section 3.4 deals with articulatory parameters measured within a single session. Section 3.5 discusses the development across sessions.

3.1 Acoustic parameters: F1 and F2

Figure 1 shows mean values of F1 (upper subplot) and F2 (lower subplot) for each session (abscissa). Different line styles and grey shades represent different speakers. The speakers are given either on the left or on the right of each graph in the same height and in the same colour as the graph. If one looks at sessions *unpert, wnpert* and *pert* first, one can see that both formants deviate somewhat in the session when no auditory feedback is there (*wnpert*) as compared to the unperturbed session. In the session thereafter (*pert*), however, the speakers regain the original values.[1] An exception is speaker A1 with more deviation in F1 in session *pert* than *wnpert*. In the later sessions the speakers do not stay with these values measured for session *pert* but the values vary in a rather inconsistent manner. The values reached in the last perturbed session (*2wpert*), which presents the end of the practice period and should therefore give the "best possible" adaptation, are not more similar to the original values than the ones already measured in session *pert*.

Repeated measures ANOVAs for data split by speaker showed a significant influence of the session on both formants except for the second formant of speaker C2. However, even if the influence is significant, it might not be perceptually relevant.[2] In general, the variation over sessions within a speaker (about 100 Hz around the F1 of the initial session and about 250 Hz around the F2 of the initial session) is much lower than the variation across speakers. Furthermore, the variation in F1 is not much higher than the "compensation threshold" of 60 Hz found in a study by Purcell & Munhall (4) and 64 Hz in a later study (5). An exception is speaker A4 with large deviations in both F1 and F2. However, an alternative explanation for the inconsistent results could be the measurement problems for this speaker discussed above.

3.2 Lip protrusion and constriction size

Before discussing the results of the articulatory parameters two problems which arose during the analysis will be described. First, for one speaker (A4) the data of the upper lip sensor were unreliable in the sessions recorded on the first day, so these data could

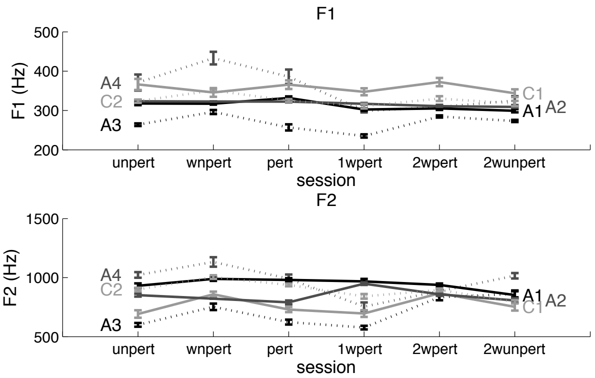

Figure 1: Mean values of formants of /u/ over sessions. Abscissa: Sessions in chronolgical order. Upper subplot: results for F1, lower subplot: results of F2. Each graph refers to one speaker. Speakers are given on the left or the right of the graph in the same colour as the corresponding graph. Error bars show standard error.

not be used for further analysis and they are missing in the figure below as well as in the calculation of the correlation.

The second problem was that sometimes the constriction size could not be measured because it was in the velar region. For one speaker (C2) the constriction was in all the sessions so far in the velar region that the palatal contour above the tongue could not be estimated with certainty and the constriction size could consequently not be measured either. The data for this speaker are therefore missing in the figure below. A similar problem occurred for speaker A4, but only in one session, session *1wpert*. In all the other sessions the constriction was in the palatal region. For this speaker constriction size was measured normally for all the sessions except *1wpert*. For the problematic session the most retracted measureable point at the end of the hard palate was taken as constriction location.

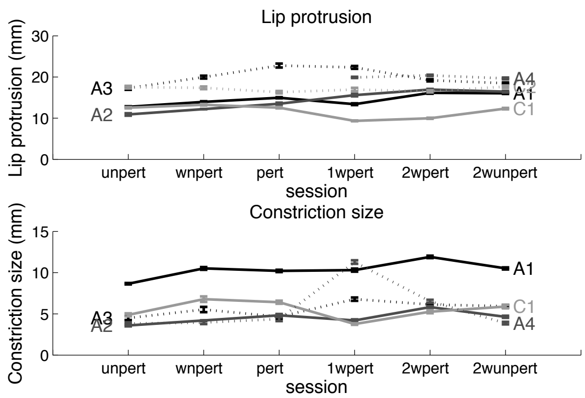

Figure 2: Mean values of lip protrusion (upper subplot) and constriction size (lower subplot) of /u/ over sessions. Abscissa gives sessions. Lip protrusion data of the first three sessions for speaker A4 and constriction data of speaker C2 are missing. Error bars show standard error.

[1]This cannot be seen for speaker A2 of course since there was no session *wnpert*.

[2]This is being tested at the moment.

Looking at the upper subplot showing lip protrusion one can see that it varies over session. It is difficult to find a tendency for all speakers. Whereas speakers A1 (black solid line) and A2 (dark grey solid line) in general increase lip protrusion until the end of the adaptation time (session *2wpert*), speaker C1 (light grey solid) rather has an oscillating pattern (increase - decrease - increase). The other speakers keep the degree of lip protrusion about equal.

Looking at the constriction size (lower subplot) one can see that there is indeed a measurement problem for speaker A4 (dark grey dotted) in session *1wpert* where the constriction was assumed to be behind the end of the hard palate. Probably it was not the constriction which was measured here but a point in front of the constriction so that the distance measured is too large. Apart from this one problematic case, the constriction sizes go parallel with the results for lip protrusion: There is an increase in lip protrusion for speakers A1 (black solid) and A2 (dark grey solid) and an increase-decrease-increase pattern for speaker C1. Repeated measures ANOVAs showed that the influence of the session on both parameters is significant.

3.3 Correlations

Results were gained for five speakers (all except C2 because no constriction could be measured). For four speakers significant correlations for at least one of the parameter pairs could be found. For speaker A4 no significant correlations could be found. This can be assumed to be due to methodological problems (one sensor missing in the three first sessions, difficulties while measuring formants).

Table 1 gives the results of the correlations between lip protrusion, constriction size and the two acoustic parameters for the four remaining speakers. The first line gives the speaker, the second line shows the correlation coefficients for the correlation between lip protrusion and constriction size. If speakers use motor equivalent strategies this correlation should be positive: For more lip protrusion the constriction should be larger. The other lines give the correlations of the two articulatory parameters with F1 and F2. If speakers use motor equivalent strategies effectively in order to keep the acoustic output constant, there should be no correlations between the articulatory and the acoustic parameters.

A significant positive correlation between lip protrusion and constriction size could be found for four speakers. For three speakers (A1, A2 and A3) there are also significant correlations between lip protrusion and one or both of the formants. These could be seen as cases of ineffective compensation: Either there is too much lip protrusion or a too narrow constriction.

speaker	A1	A2	A3	C1
lip-const	0.568***	0.537***	0.235*	0.660***
lip-F1	-0.126	-0.330**	-0.351***	0.006
lip-F2	-0.220*	0.065	-0.313**	0.118
const-F1	-0.173	-0.197	-0.014	0.037
const-F2	-0.033	-0.108	0.042	0.093

Table 1: Correlation coefficients and significance levels for four speakers (first line) of the correlation between lip protrusion and constriction size (second line), lip protrusion and F1 (third line), lip protrusion and F2 (fourth line), constriction size and F1 (fifth line), constriction size and F2 (sixth line). Significance levels: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

3.4 Relation within a session

In order to show more details the results for lip protrusion vs. constriction size for speaker A1 are shown in figure 3. The abscissa shows lip protrusion, the ordinate constriction size. A correlation between the two parameters can clearly be seen: If a production has more lip protrusion it also has a wider constriction. Numbers represent different sessions in chronological order. The figure is meant to illustrate something which could be seen for all the four speakers: Within single sessions the correlation between the parameters either does not exist or it is weaker or it is even negative. For the speaker shown here a positive correlation exists for session 4 (*1wpert*) but for example not for session 1 (*nopert*). For session 5 (*2wpert*) there is a clear negative correlation.

The reasons for this result are so far not entirely clear. It is possible that the measuring accuracy of our method which can, for constriction size be assumed to be around 1 mm is too low in order to see a relation within a session. The correlation between lip protrusion and constriction size might be blurred when both parameters change only little (as they do within a session) but it might become clear when they change more (as is the case across sessions).

The fact that the parameters change more across sessions than within sessions shows in any case that within each session the speaker has a preferred strategy: a preferred constriction size and a preferred degree of lip protrusion. In session *unpert* ('1'), for example the lips are only slightly protruded and the constriction is rather small. In the last perturbed session ('5') there is a lot of lip protrusion, and the constriction is larger.

3.5 Development across sessions

There is an initial increase in constriction size in session *wnpert* as compared to session *unpert* (cf. figure 2, lower subplot). This could be because the speakers try to avoid having too much linguo-palatal contact and to thus produce a fricative rather than a vowel. Some of the speakers compensate for this lower tongue position by more lip protrusion even in the initial perturbed session. Afterwards, with some practice, speakers try out different strategies by varying both

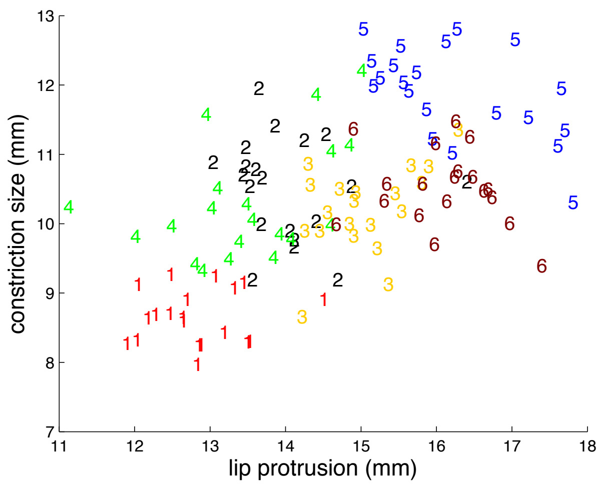

Figure 3: Lip protrusion and constriction size measurements for speaker A1. Different numbers represent different sessions in chronological order: *unpert*: 1, *wnpert*: 2, *pert*: 3, *1wpert*: 4, *2wpert*: 5, *2wunpert*: 6.

lip protrusion and constriction size. The development over sessions is quite individual and might be due to speaker specific morphological differences which lead to preferences for certain strategies.

4 Conclusion

This study has dealt with motor equivalent strategies in /u/. The results show that for all speakers for whom there were no methodological problems motor equivalent strategies could be found. Speakers vary the tongue position and in order to compensate for that they produce more or less lip protrusion. However, the compensation is not complete: For three of the four speakers there is a negative correlation between one of the articulatory parameters and one or even both of the acoustic parameters. The constriction is thus too small, or there is too much lip protrusion in order to keep the acoustic output completely stable.

An alternative explanation for the negative correlations between one of the articulatory parameters and the acoustic parameters could be that other articulatory parameters are changed in order to keep the acoustic output constant. As discussed above, the lip opening area, for example, can vary in dependence on the constriction location. Neither of these additional parameters has been investigated here.

The correlations found here are a little stronger (higher correlation coefficients) than the ones presented in (3). A reason for that could be that, in contrast to this earlier study, in the present study the speakers are confronted with a perturbation. They are thus more inclined to try out several articulatory strategies. Judging from the sensor plots presented in (3), lip protrusion in all cases varies considerably less than a centimetre. The ranges for the speakers presented here are between 7 and 12 mm.

References

[1] C. Espy-Wilson and S.E. Boyce, "Acoustic differences between "bunched" and "retroflex" variants of American English /r/", *J. Acoust. Soc. Am.* 95 (5), 2823 (1994)

[2] F. Guenther, C. Espy-Wilson, S. Boyce, M. Matthies, M. Zandipour and J. Perkell, "Articulatory Tradeoffs Reduce Acoustic Variability During American Englisch /r/ Production", *J. Acoust. Soc. Am.* 105 (5), 2854-2865 (1999)

[3] J.S. Perkell, M.L. Matthies, M.A. Svirsky and Jordan, M.I., "Trading Relations Between Tongue-body Raising and Lip Rounding in Production of the Vowel /u/: a Pilot 'Motor Equivalence' Study", *J. Acoust. Soc. Am.* 93 (5), 2948 - 2961 (1993)

[4] D. Purcell and K. Munhall, "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation", *J. Acoust. Soc. Am.* 120 (2):966-977 (2006)

[5] D. Purcell and K. Munhall, "Psychophysical detection threshold of formant manipulations during speech production", *J. Acoust. Soc. Am.* 122 (5):3068-3069 (2007)

[6] C. Savariaux, P. Perrier and J.-P. Orliaguet, "Compensation strategies for the perturbation of the rounded vowel [u] using a lip-tube: A study of the control space in speech production", *J. Acoust. Soc. Am.* 98 (5), 2428-2442 (1995)

[7] C. Savariaux, P. Perrier, J.-P. Orliaguet and J.-L. Schwartz, "Compensation strategies for the perturbation of the rounded vowel [u] using a lip-tube II: Perceptual analysis", *J. Acoust. Soc. Am.* 106 (1), 381–393 (1999)

[8] M.K. Tiede, V. Gracco, D.M. Shiller, C. Espy-Wilson and S.E. Boyce, "Perturbed palatal shape and North American English /r/ production", *J. Acoust. Soc. Am.* 117 (4), 2568–2569 (2005)

[9] X. Zhou, C. Espy-Wilson, M.K. Tiede and S. Boyce, "Acoustic cues of "retroflex" and "bunched" American English rhotic sound", *J. Acoust. Soc. Am.* 121 (5), 3168 (2007)

Acknowledgements

This study was supported by grants of the Deutsche Forschungsgemeinschaft (PO 334/4-1 and HO 3271/1-1) to Bernd Pompino-Marschall and Phil Hoole as a project in the framework of the German-French network POPAART, and a grant of the Ministère délegué à l'enseignement et à la recherche of France. We would like to thank Jörg Dreyer from the Zentrum für Allgemeine Sprachwissenschaft for technical support during the recordings, and Olessia Panzyga, Vivien Hein and Susanne Waltl from the Humboldt Universität zu Berlin and the Ludwig-Maximilians-Universität München for acoustic segmentation and formant measurements.