Motor equivalent strategies in the production of German /ʃ/ under perturbation

Jana Brunner

Zentrum für Allgemeine Sprachwissenschaft, Berlin, Germany

Phil Hoole

Institut für Phonetik und Sprachverarbeitung, Ludwig-Maximilians-Universität München, Germany

Running title: Motor equivalence in /ʃ/

Acknowledgements

This study was supported by grants of the Deutsche Forschungsgemeinschaft (PO 334/4-1 to Bernd Pompino-Marschall, HO 3271/1-1 to Phil Hoole), by a grant of the Ministère délégué à l'enseignement superieur et à la recherche scientifique for a cotutelle de thèse, and a grant for post-doctoral research of the Deutscher Akademischer Austauschdienst to Jana Brunner. We thank Jörg Dreyer for carrying out the 2D EMA recordings and Olesya Rauch, Vivien Hein and Susanne Waltl for acoustic segmentation. Many thanks to Tine Mooshammer, Bernd Pompino-Marschall, Melanie Weirich, editor Jim Scobbie and reviewers Maria-Josep Solé and Megan McAuliffe for comments on earlier versions. Thanks also to Mark Tiede for providing scripts for the calculation of spectra and spectral moments.

Corresponding author

Jana Brunner, Zentrum für Allgemeine Sprachwissenschaft, Schützenstr. 18, 10117 Berlin, Germany

Abstract

The German sibilant /ʃ/ is produced with a constriction in the postalveolar region and often with protruded lips. By covarying horizontal lip and tongue position speakers can keep a similar acoustic output even if the articulation varies. This study investigates whether during two weeks of adaptation to an artificial palate speakers covary these two articulatory parameters, whether tactile landmarks have an influence on the covariation and to what extent speakers can foresee the acoustic result of the covariation without auditory feedback. Six German speakers were recorded with EMA. Four of them showed a covariation of lip and tongue, which is consistent with the motor equivalence hypothesis. The acoustic output, however, does not stay entirely constant but varies with the tongue position. The role of tactile landmarks is negligible. To a certain extent, speakers are able to adapt even without auditory feedback.

INTRODUCTION

Variation in the Production of /ʃ/

Producing /ʃ/ requires a very precise articulation. At least the following factors are important in articulating this sound:

- The tongue has to form a constriction in the postalveolar region (Shadle, 1985).
- In some languages such as German the lips may be protruded.
- The jaw has to be in a rather high position (Shadle, 1990, Mooshammer et al., 2007) so that the air jet is directed against the incisors.
- In contrast to fricatives such as /f/ and /x/ the tongue has to be grooved (e.g. Shadle, 1990, Shadle et al., 2008) within the constriction, but also at the more posterior region of the tongue behind the constriction (the "inlet").

A variation in the third and fourth parameter leads to changes in the higher frequency ranges of the spectrum beyond 4kHz (Shadle, 1990, Shadle et al., 2008). Both a change in lip protrusion and a change in constriction position have an influence on the length of the front cavity (region between constriction and lip opening), which in turn influences the frequency of the spectral peak (usually at about 3kHz). For a constant constriction position, more lip protrusion lengthens the front cavity and consequently lowers the frequency of the spectral peak. Less lip protrusion makes the cavity shorter and raises the spectral peak. For constant lip positions a more advanced constriction position raises the frequency of the spectral peak and a more retracted constriction position lowers it (Shadle, 1985).

Speakers can thus influence the length of their front cavity by two parameters: lip protrusion and constriction position. Speakers who have a rather fronted constriction position have to use more lip protrusion than speakers who have a rather retracted constriction position in order to produce similar acoustic results. Since there is a many-to-one relationship between articulation and acoustics (cf. e.g. Guenther, 1994, Jordan, 1996, Laboissière, 1992), even a single speaker can use several articulatory configurations while keeping the acoustic result constant. This study aims at investigating this kind of *motor equivalence* (i.e. different articulations but similar acoustic outputs) in single speakers.

Motor Equivalence

Within the speech field, the term *motor equivalence* is used to denote different articulatory strategies to produce similar acoustic outputs, for example by using different contributions of several articulators to a vocal tract shape or by using different vocal tract shapes resulting in similar acoustic outputs.

A very well-known example for motor equivalence is bite block speech. Bite block studies have shown that, if the jaw movement was blocked the tongue managed to compensate for that and the speaker could still produce nearly normal speech by forming constriction sizes that are similar to the ones formed in unperturbed speech (Lindblom et al., 1979, Kelso & Tuller, 1983). Some studies, however, found that there were small acoustic differences between the bite-block and the normal condition (McFarland & Baum, 1995) in vowels and consonants.

Other studies suggesting the existence of motor equivalent strategies are the lip perturbation studies by Folkins & Zimmermann (1982) and Gracco & Abbs (1985). Folkins & Zimmermann carried out a perturbation experiment where speakers' lower lip was pushed down unexpectedly by an electrical stimulation. The three speakers investigated in the study showed active compensatory behavior in bilabial stop production by moving the upper lip further down and by moving the jaw up. Gracco & Abbs (1985) perturbed the movement of the lower lip by pushing it down with a paddle unexpectedly during bilabial stop production. Speakers compensated via an increase in movement amplitude, velocity and movement time of both upper and lower lip in order to produce the closure they intended to produce without the perturbation. In all these studies speakers were able to produce a similar vocal tract shape while the contributions of single articulators (jaw, lips, tongue) varied.

However, motor equivalence has also been reported in unperturbed speech, for example in American English /u/. This sound can be produced for example (1) with protruded lips and a constriction in the velar region, (2) with open lips and a constriction in the velo-pharyngeal region. Perkell et al. (1993) investigated motor equivalence patterns between lip rounding and tongue body raising and -retraction in American English /u/. Both strategies lead to a lower F2 and could therefore be covaried while keeping the acoustic output constant. This EMA study discussed productions of four speakers each producing /u/ 300 times in different contexts. Three of the four subjects showed weak correlations thereby giving some support to the motor equivalence hypothesis.

Another study investigating different articulations of /u/ is the one by Savariaux et al. (1995 and 1999). In this study speakers' lip movement was perturbed by a 2.5 cm diameter tube

which held the lips open. The opening that was enforced by this lip tube is considerably larger than the one observed in unperturbed productions. Perkell et al. (1993) report a maximal opening of 0.86 cm^2 for their speakers which would, for a perfectly circular lip opening, correspond to a diameter of slightly more than 1 cm. Speakers were asked to produce /u/ in this condition. The nomograms by Fant (1960) suggest that a way to compensate for this perturbation is to produce the constriction at a more posterior place. In fact, the Xray-data recorded from the subjects showed that some of the speakers retracted the tongue further than in the unperturbed condition, with one of them producing a constriction in the velo-pharyngeal region resulting in similar formant frequencies of the produced sound as compared to the productions with protruded lips and a velar constriction.

Perkell et al. (2000) investigated motor equivalence in /ʃ/. The acoustic characteristics of this sound are determined by the length of the front cavity (from the constriction to the lip opening). This length of the front cavity, and thereby the acoustic output, can be kept similar by covarying lip protrusion and constriction position. Motor equivalence for lip rounding vs. tongue tip fronting was investigated for eight speakers. The results were mixed in that some speakers showed motor equivalence (i.e. covariation between the two gestures) but others did not.

Another very well-known case of motor equivalence is found in the various articulations of American English /r/. This sound can be produced by several markedly different articulatory configurations, namely with a bunched tongue or with a retroflexed tongue (Delattre & Freeman, 1968). In both cases a constriction is created which leads to a low F3, a very salient acoustic characteristic of this sound. Westbury et al. (1998) found for a sample of 53 speakers that the two

production types, bunched and retroflex, rather form the edges of a continuum and productions are often somewhere between the two extremes. The production type furthermore depends not only on the speaker but also on the context. Zhou et al. (2007, 2008) found small acoustic differences between the two production types. The retroflexed variant has a larger difference between F4 and F5 than the bunched variant. The authors explain this with the fact that the resonances come from different cavities in the two production types. This finding shows that the vocal tract shapes of the two sounds indeed differ, although the acoustic output, at least for the lower formants, is similar.

There is some evidence that motor equivalence is a phenomenon that is language specific. Lawson et al. (in press) investigated the /r/-productions of working class and middle class Scottish speakers and found that there was no motor equivalence within a speaker. Rather, working class speakers tended to use variants with the tongue tip up or the tongue front up whereas middle class speakers tended to have bunched variants. Furthermore, acoustic differences were found: middle class speakers produced audibly rhotic variants of the sounds whereas the productions of working class speakers were often r-less or weakly rhotic. The authors note that the acoustic difference is merely due to a different timing (raising gesture delayed beyond the offset of voicing for the working class speakers), however, the results suggest that for these Scottish speakers the two articulations are not equivalent since they are used in sociolectal variation.

It is also possible to see a link between motor equivalence and cue trading[a]. In trading relations also, different articulations can lead to a similar *perceptual* result. Diehl & Kingston

[a] We are grateful to editor Jim Scobbie for drawing this to our attention.

(1991), for example, show that [+voice] judgments in stops increase not only for longer voicing durations but also for low F1 and low f0 of the following vowel. Diehl & Kingston therefore suggest that these three parameters are merged into the perception of only one characteristic[b]. So, maybe the concept of motor equivalence should be broadened to produce not only something acoustically similar but rather something perceptually similar.

One could wonder why speakers should use motor equivalent strategies if they could just as well use a single articulatory strategy. There are several possibilities. One possibility is reduction of articulatory effort. In some contexts it might be easier to use bunched /r/ than in others, where the production of retroflex /r/ might involve less effort (cf. discussion in Brunner et al. in press a). A similar example would be utterances of the same sound in the same context but with differences in emphasis on the word (Scobbie, pers.comm.). A syllable bearing emphasis can be expected to have a slightly different articulation than the same syllable not bearing emphasis. As a result of this the production of a particular sound in that syllable might be easier using a different articulatory strategy in the emphasis case than in the no-emphasis case.

In perturbed speech motor equivalent strategies are often used because the speakers are blocked from using their usual strategies (e.g. bite block). Another reason for using motor equivalence strategies could be to find either the most efficient articulatory strategy or the one giving the best acoustic result by trying out several strategies, for example when adapting to a dental device.

What seems to be clear is that speakers only use motor equivalent strategies if there is some change in some domain (e.g. context, emphasis, perturbation). If speakers produce the same

[b] But see Löfqvist et al. (1989) and Hoole & Honda (2011) for alternative approaches to the relationship between voicing and f0.

utterance in the same situation over and over again they will use the same articulatory strategy over and over again. Consequently, in all motor equivalence studies speakers were brought to produce a lot of articulatory variability, for example by perturbing their speech (e.g. Savariaux et al., 1995) or by producing sounds in different contexts (e.g. Westbury et al., 1998). Inducing sources of variability is essential for these studies since, if speakers are producing the same sound several times under exactly the same conditions articulatory strategies will be so similar that motor equivalence cannot be observed.

The present study uses a perturbation in order to increase articulatory variability and thus the use of motor equivalent strategies. If speech is perturbed speakers have to vary their articulation to adapt. During this process they are likely to try out different articulatory strategies in order to find an efficient way to produce the sound under perturbation.

Auditory and tactile feedback in normal and perturbed speech

A number of studies have shown that during speech acquisition, but also when adapting to a perturbation speakers use auditory and tactile feedback. Within the framework of their speech production model DIVA Guenther et al. (1998) and Guenther (1995, 2003) propose that during speech learning speakers set up various mappings between orosensory, articulatory and auditory characteristics of sounds. Once set up these mappings can be used in feedforward control to find an articulatory configuration for a desired acoustic output.

Within the framework of this model, as long as the vocal tract does not change, it is not necessary to have auditory or tactile feedback available in order to speak. This assumption is in

agreement with the observation that postlingually deaf speakers are able to maintain intelligible speech for years (e.g. Waldstein, 1990), possible due to the mapping between orosensory targets onto speech sounds. Further support regarding the role of tactile feedback after speech learning comes from Linke (1980) who discusses a speaker suffering from loss of oral sensibility. Although this subject had problems controlling many actions involving the oral cavity (for example eating and smoking) he did not experience problems during speech, even with auditory feedback masked, possibly due to the mapping from articulatory actions to acoustic outputs. However, when his speech was disturbed (stimulation of the M. orbicularis oris), he had problems adapting. Similarly, Hoole (1987) discusses adaptation to a bite block by a speaker suffering from loss of oral sensibility with and without auditory feedback available. This speaker did not have problems speaking without perturbation. In the absence of auditory feedback the speaker was not able to compensate sufficiently although he showed clear compensatory behavior when auditory feedback was available.

The examples show that in unperturbed speech feedback is not necessary. However, when speech is perturbed feedback becomes essential in order to carry out remappings between orosensory, articulatory and auditory information. An example for such a reset using auditory feedback is described in a study by Jones & Munhall (2003). Jones and Munhall extended the upper incisors of speakers and asked them to produce /s/, at first with auditory feedback masked and afterwards with auditory feedback available. The extension increased the size of the front cavity (from the constriction to the mouth opening). Acoustically, this resulted in a lower center of gravity. As long as there was no auditory feedback available the speakers did not adapt and possibly used their learned orosensory-auditory mappings. When auditory feedback became

available speakers adapted the tongue position so that the acoustic output became more similar to that of the unperturbed condition.

Honda et al. (2002) also investigated the role of auditory feedback in adaptation. In this study speech was perturbed by a palatal prosthesis with an inflatable balloon at the alveolar ridge. The speakers were asked to produce the sounds /ʃa/ and /tʃa/. On some trials the balloon was blown up. Auditory feedback was temporally masked. Identification scores in the subsequent perception experiment were in general high except for the first trial after inflation. The productions without auditory feedback were more often misidentified than the productions with auditory feedback available, showing that auditory feedback is used in adaptation.

In a follow up study (Honda & Murano, 2003) adaptation to the inflatable palate was investigated in four different feedback conditions, (1) with tactile and auditory feedback available, (2) with auditory feedback masked, (3) with tactile feedback masked via an anaesthetic, and (4) with both auditory and tactile feedback masked. The results show that speakers adapt best with both kinds of feedback available and worst with both kinds of feedback masked. When only one kind of feedback was available the error scores for the perturbed trials were comparable. These results suggest that both kinds of feedback are used in adaptation.

Aims of the Study and Hypotheses

Whereas some studies have found that some speakers use motor equivalent strategies in /r/ and /u/, so far no clear results have been found for /ʃ/. A possible reason for that could be that /ʃ/ has a very stable articulation due to the great amount of linguo-palatal contact. This might reduce the articulatory variability in this sound so that motor equivalence is simply not found because there

is not enough variability. If the production of /ʃ/ was perturbed this should lead to a change in articulation so that motor equivalence could possibly be observed. The primary aim of this study was to investigate the use of motor equivalence strategies in adaptation. In order to do this, speakers' speech was perturbed by a palatal prosthesis. The first hypothesis, related to this first aim, is:

- *Motor equivalence hypothesis.* If speakers use motor equivalence strategies there should be a positive correlation between the horizontal position of the tongue tip and the lip position: The more retracted the tongue, the less lip protrusion. Furthermore, there should be no correlation between an acoustic parameter such as the center of gravity and either one of the articulatory parameters because the aim of motor equivalence strategies is to keep the acoustic output constant. Alternatively, if there were no motor equivalence strategies but just random variation in articulation, each articulatory parameter should correlate with the acoustic parameter, and the two articulatory parameters should not correlate.

The second aim of the study was to investigate whether a certain kind of palate shape has an influence on the amount of articulatory variability (and possibly motor equivalence) observed in perturbed speech. Some studies have shown that tactile feedback is used in adaptation (cf. e.g. the sessions with auditory feedback masking in Honda et al, 2002, Honda & Murano, 2003). Related to that, it is possible that speakers use certain landmarks such as a pronounced alveolar ridge as points of orientation for the tongue so that the articulatory variability (and possibly motor equivalence) is reduced. In contrast to that, it is possible that a very flat palate increases the articulatory variability because there is no landmark and the tongue can slide along the palate. In

order to investigate this question two different artificial palates were used, one with a pronounced alveolar ridge ("alveolar palate") and a flat one ("central palate"). Related to this aim the second hypothesis is set up:

- *Articulatory landmark hypothesis.* When speech is perturbed with a palatal prosthesis speakers will probably lower their tongue in order to prevent a closure. Furthermore, since the prosthesis changes not only the palatal height but also the palatal contour, they should try to find a new place of articulation for /ʃ/. While doing this speakers should make use of auditory feedback, but they could also use tactile feedback of articulatory landmarks, such as the alveolar ridge. Following on from this, speakers with an alveolar prosthesis (with a pronounced alveolar ridge), should produce /ʃ/ at a certain point behind the alveolar ridge. In contrast to that, speakers with a central palate should find it more difficult to find the new place of articulation because there is no landmark. As a result of this, within the perturbed productions, there should be more variability in tongue position (and, provided that the motor equivalence hypothesis is correct, more motor equivalence) for the speakers with a central palate as opposed to the speakers with an alveolar palate.

The third aim of the study was to investigate in how far speakers can adapt to an artificial palate while using tactile feedback only. Honda et al. (2002) suggest that adaptation with tactile feedback only is possible to some extent. In the present study this will be further investigated. We are interested in the question whether speakers, if they adapt without auditory feedback, use a motor equivalent strategy. Therefore, speakers' auditory feedback was masked at perturbation onset. Related to this aim the following hypothesis was set up:

- *Auditory feedback hypothesis.* If auditory feedback is absolutely essential in adaptation speakers should not be able to adapt when their auditory feedback is masked. The acoustic result of their productions should differ from the productions with auditory feedback available. Furthermore, the productions without auditory feedback should not "fit in" with the covariation of tongue and lip (if there is such covariation). Alternatively, if auditory feedback is not essential in adaptation there should be no clear acoustic difference in the productions with auditory feedback masked and auditory feedback available and the productions without auditory feedback should fit in with the covariation between tongue and lip.

METHODS

Speakers

Six volunteers took part in the study, two males and four females. They were between 25 and 40 years old. None of the speakers reported any history of speech or hearing problems, although hearing was not screened prior to the experiment. All speakers spoke Standard German with some regional influence. All the subjects reported in this study had worn dental devices for 1-3 years.[c] For each person a dental cast was made by a dental technician in order to be able to exclude speakers with malocclusions and to make sure that the speakers had a rather domed palate so that there was enough space for the insertion of the prosthesis. As a result of this

[c] We were actually interested in recruiting subjects without any experience in wearing dental prostheses. However, it was found that there are hardly any young people in Germany without such experience. In order to have a homogeneous group of subjects, we therefore decided to recruit speakers with such experience.

procedure, one speaker was excluded because her palate was too flat, and another speaker was excluded because of malocclusion and a very small palate.

Perturbation

A two-week perturbation experiment was carried out. The speakers' articulation was perturbed by a palatal prosthesis. There were two types of custom made 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 arch ("central palate"). The palates had a maximal thickness of 1cm. They were thus considerably thicker than a standard EPG-palate. An example of each palate type is shown in figure 1. The plots show the natural palatal contour (bold solid line) and the artificial palate (bold dashed line). The tongue contour of one production of /ʃ/ is shown as a linear interpolation between three EMA sensors on the tongue tip, tongue dorsum and tongue back (see "experimental setup"). The position of the upper lip sensor is shown as well.

The prosthesis in each case lowers the palate so that one can expect a closure during /ʃ/ if speakers do not adapt. If speakers compensate just by lowering the tongue, this will, assuming a rotational movement of the jaw, lead to a post-laminal constriction (thin dashed lines in figure 1). The change will be much more severe for speakers with the alveolar palate than for the ones with the central palate. Since it is difficult to form the medial groove postlaminally one can expect the speakers to adapt further, for example by a retraction of the tongue. This will lead to a change in the length of the front cavity which has to be compensated for by less lip protrusion.

If the speakers with the alveolar prosthesis have found a way to produce /ʃ/ with the prosthesis, it should be easy for them to remember the tongue position and use it over and over

again because of the alveolar ridge, which can be used as a point of orientation. Thus, for speakers with an alveolar palate the strategy should change at perturbation onset and perturbation offset, but stay rather constant in between.

For the speakers with a central palate there should be less retraction of the tongue because the changes are less severe. However, since for these speakers the palate is completely flat, it should be difficult to keep a constant place of articulation since the tactile feedback is similar over a wide range of articulatory positions. In contrast to the alveolar palate the articulation might therefore vary over the time of adaptation.

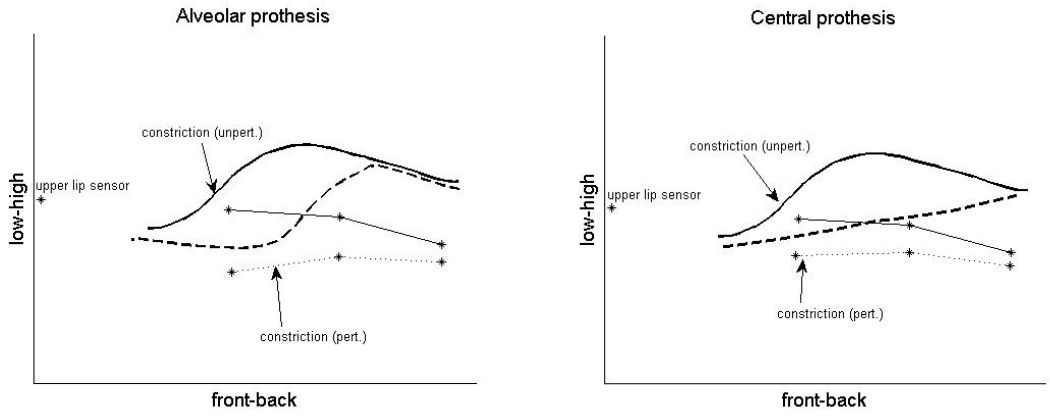

Figure 1: Examples for an alveolar prosthesis (left subplot) and a central prosthesis (right subplot), midsagittal view. Front is left. Bold solid line: natural palatal contour. Bold dashed line: artificial palatal contour. Thin solid line: estimation of unperturbed tongue contour (linear interpolation between sensor positions) during a production of /ʃ/. Thin dashed line: estimation of tongue contour under perturbation when tongue is lowered without further adaptation. Asterisks: sensor positions.

Three speakers (AM1, AM2 and AF1) were recorded with an alveolar prosthesis and three (CF1, CF2 and CF3) with a central prosthesis. Allocation of the speakers to one of these groups was done according to the shape of their natural palates. Speakers without a pronounced alveolar

ridge were provided with an alveolar palate, the others were provided with a central palate. This procedure unfortunately led to a biased gender distribution. The central palate group consisted of females only, whereas there were two men in the alveolar palate group. Speakers were asked to wear the prosthesis all day for two weeks and to make a serious effort to improve their speech.

Experimental Setup

Speakers' articulator movements were recorded via electromagnetic articulography. Sensors were placed midsagittally, one at approximately 1 cm behind the tongue tip (*tongue tip sensor*), one as much retracted as possible (*tongue back sensor*) and one in between the two (*tongue mid sensor*), one below the lower incisors (*jaw sensor*) and one on each lip. Reference sensors were placed on the bridge of the nose, and above the upper incisors. After the recordings, the articulatory data were corrected for head movements and rotated to the occlusal plane. For the present purpose the data of the tongue tip and the upper lip sensor were analyzed. The upper lip was analyzed rather than the lower lip in order to avoid the difficulties involved in decomposing the lower-lip sensor signal into its jaw-related and intrinsic lip components (cf. Westbury et al., 2002) .

Acoustic recordings were carried out with a DAT recorder and a Sennheiser MKH 20 P48 microphone. The distance between the microphone and the speaker’s lips was about 30 cm. The acoustic signal was downsampled to 24kHz.

The experiment consisted of several recording sessions which are summarized in table I. On the first day of the experiment three different sessions were recorded. First, speakers were recorded without the prosthesis (session *1np*, meaning "first session, no perturbation"). In the

second session, the artificial palate was inserted and speakers' auditory feedback was masked with white noise (session *2wp*, "second session, white noise - perturbed"). In the third session speakers were recorded with auditory feedback available (session *3pe*, "third session, perturbed"). The subjects were instructed to wear the prosthesis all day and practice speaking. They were asked to read aloud an exercise sheet once a day and to write down the number of hours they had worn the prosthesis each day. All speakers reported to have worn the prosthesis between 12 and 18 hours per day. After one week adaptation time speakers were recorded with the prosthesis in place (session *4pe*, "fourth session, perturbed"). A final perturbed session was recorded after two weeks (session *5pe*, "fifth session, perturbed"). Then speakers removed the prosthesis and were recorded without the perturbation (session *6np*, "sixth session, no perturbation).

Table I: Recording sessions

Recording name	**Adaptation time**	**Auditoy feedback masking**	**Perturbation**
1np ("no perturbation")	0	no	No
2wp ("white noise, perturbed")	0	yes	Yes
3pe ("perturbed")	duration of session 2wp	no	Yes
4pe ("perturbed")	1week	no	Yes
5pe ("perturbed")	2 weeks	no	Yes
6np ("no perturbation")	2 weeks	no	No

Speech Material

The target sound /ʃ/ was recorded in the nonsense word /'ʃaxa/ spoken in a carrier phrase: *Ich sah Schacha an* ("I looked at /'ʃaxa/."). There were 20 repetitions in each session, randomized with 10

other CVCV sequences including all lingual sounds of German, each repeated 20 times in the same carrier phrase giving 220 sentences per session and a total of 120 repetitions of /'ʃaxa/ per speaker for the six sessions. The recording time of each of the six sessions was about 20 minutes.

Auditory Feedback Masking

In order to investigate how speakers adapt when no auditory feedback is available their auditory feedback was masked in the first perturbed session with white noise (100 Hz-10 kHz) presented over head phones.

Acoustic Analysis

Segmentation. The consonant /ʃ/ was acoustically segmented (friction onset to friction offset) in each utterance. The segmentation was carried out in PRAAT. All further analyses, articulatory and acoustic, were carried out at the temporal midpoint of the segments.

Band-pass filtering. As has been shown in previous palate perturbation studies, in fricatives speakers adapt not only the position of the constriction (influencing the frequency of the spectral peak) but also the tongue shape (e.g. Hamlet & Stone, 1978, Brunner et al., in press b for /s/). This influences the amplitudes of the higher frequencies (Shadle et al., 2008). Visual analyses of our spectra confirmed this. During session 2wp there was less energy in the higher frequencies than in the unperturbed session. For both prosthesis types, only with time did speakers manage to produce tongue shapes which resulted in the production of high frequency noise. Furthermore, although /ʃ/ is phonologically unvoiced, for some speakers there were remnants of voicing. Both the changes in high frequency energy and the voicing would have been problematic for the

calculation of spectral parameters since we were primarily interested in the frequency of the main spectral peak at around 3kHz because this is the acoustic characteristic which is influenced by the position of the constriction and lip protrusion.

In order to get information about the frequency of this peak without much influence of other characteristics of the spectrum the data were band-pass filtered with cut-off frequencies of 700 Hz and 6 kHz. The lower cut-off frequency was chosen so that voicing was excluded from the analysis. The higher cut-off frequency was chosen so that the influence of tongue shape changes during adaptation was excluded from the analysis.

Calculation of spectral parameters. As shown in a modeling study by Shadle (1985) a lengthening of the front cavity leads to a downwards shift of the main concentration of energy of the spectrum. This lengthening can be reached by either retracting the constriction or by using more lip protrusion. Three parameters which have been used for the description of shifts in the main concentration of spectral energy before, i.e. center of gravity, skewness and the second coefficient of a discrete cosine transform (Watson & Harrington, 1999, Guzik & Harrington 2007) have been tested for their suitability to describe the relation between acoustics and a change in front cavity length (cf. appendix). For a longer front cavity, the center of gravity should become lower whereas the skewness value and the DCT coefficient 2 should become higher. Similarly, for a shorter front cavity due to either less lip protrusion or a more advanced constriction position, leading to an upwards shift of the main energy concentration in the spectrum, the center of gravity should become higher, whereas the skewness and DCT coefficient 2 should become lower. As is shown in the appendix, the DCT coefficient 2 described this relation best and was therefore selected for the following analyses.

Articulatory Analysis

Data exclusion. On each recording day the sensors had to be glued to the tongue anew. We tried to position the sensor at exactly the same location in each recording session. To do so, photos of the tongue were taken for each session and anatomical landmarks on the tongue were noted in order to be able to position the sensor at the same place in the following sessions. In order to be able to judge the comparability of the sensor positions the speaker was asked in each session to put his/her tongue against the palate in rest position. Comparisons of the photos and the rest position recordings showed that in session *4pe* of speakers AM2 and CF2 the tongue tip sensor had been glued to a slightly different location. These sessions were removed from the data. Unfortunately, it is not possible to completely exclude the possibility that there were small differences in sensor positioning in the other sessions as well, but they should be in the range of a millimeter maximum. Also, the first three sessions and the last two were recorded with the same sensor glueing. Comparisons of positions with different and same glueings showed no greater differences in sensor positions for sessions with different glueings than for sessions with the same glueing.

For the sessions of the first day of speaker CF1 no data were available because of technical problems with the upper lip sensor. Also, some of the acoustic measurements for speakers AM1, CF1 and CF3 resulted in outliers (below or beyond the mean ±2 standard deviations). These data were removed as well.

Lip position. Lip protrusion was estimated as the horizontal position of the upper lip sensor. As a consequence of the experimental arrangement lower values mean more lip protrusion.

Tongue tip position. The constriction position was estimated as the horizontal position of the tongue tip sensor. The higher this value, the more retracted is the tongue.

Relationships Between Articulatory and Acoustic Parameters

According to the motor equivalence hypothesis, speakers should covary the horizontal upper lip position and the horizontal tongue tip position. Therefore, Pearson correlations were calculated for these two articulatory parameters, pooling data from all sessions.

Furthermore, correlations between the acoustic parameter DCT-coefficient 2 and each of the articulatory parameters were calculated. According to the motor equivalence hypothesis, there should be no correlations between the acoustic parameter and any of the articulatory parameters. Statistical analyses were carried out in R.

RESULTS

At the beginning of this section the measurement results are presented briefly. Afterwards they are discussed in relation to the three hypotheses.

Relationship between tongue tip position and lip position. Figure 2 shows the relations between the two articulatory parameters (horizontal tongue tip position and horizontal upper lip position). Each subplot shows the results for one speaker. The speakers with an alveolar palate are shown on the left, the speakers with a central palate are shown on the right. The numbers in the plots refer to single productions and give the session in which these productions were recorded (1: 1np,

2: 2pe…). Numbers are given in different grey shades in order to make sessions visually better distinguishable. Regression lines are also given. In each subplot lip position is shown on the abscissa and tongue position is given on the ordinate. Low values on either axis mean that an articulator was at a more anterior position, high values mean that it was at a more posterior position. Table III (second column) gives the correlation coefficients and p-values for the correlations between tongue tip position and lip position. A positive correlation between tongue tip position and lip position would be in agreement with the motor equivalence hypothesis since protruding or retracting both tongue and lip at the same time would keep the front cavity about the same size.

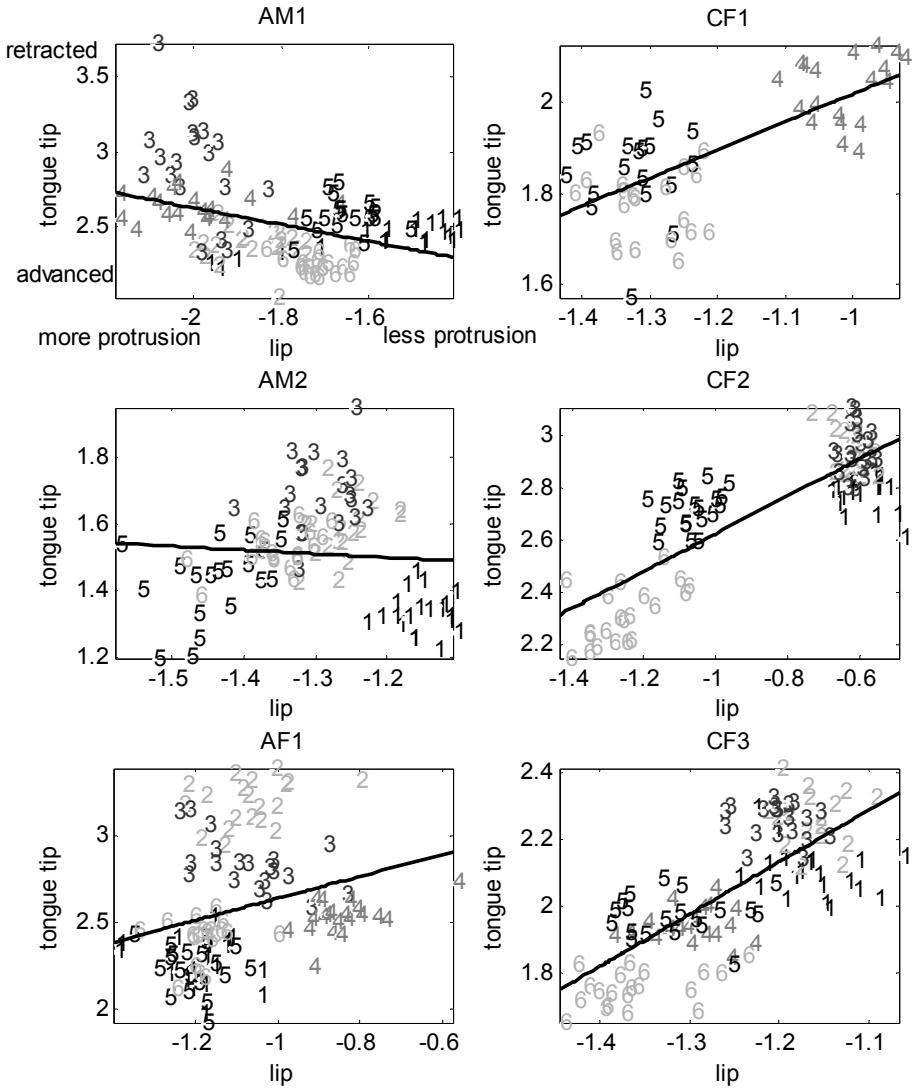

Figure 2: Lip position versus tongue position. If speakers use motor equivalent strategies there should be a positive correlation. Each subplot shows the results for one speaker. The abscissa gives results for horizontal upper lip position in cm, the ordinate gives results of the horizontal tongue tip position in cm. Smaller values refer to more advanced positions. Numbers in different grey shades refer to the session in which the production was recorded (1: session 1np, no perturbation, no masking, 2: session 2wp, perturbed, with masking, 3: session 3pe, perturbed, no masking, 4: session 4pe: perturbed, no masking, 5: session 5pe, perturbed, no masking, 6: session 6np, no perturbation, no masking). Straight lines show regression.

Table III: Correlation between articulatory and acoustic parameters. First column: speaker. Second column correlation coefficients and p-values for relation between tongue position and lip position, third column: relation between lip position and coefficient 2, fourth column: relation between tongue position and coefficient 2, last column: number of repetitions taken into account.

Speaker	**tongue – lip**	**lip - coeff2**	**tongue - coeff2**	**N**
AM1	-0.40 (.000)	-0.509 (.000)	0.386 (.000)	119
AM2	-0.08 (.439)	-0.200 (.046)	0.648 (.000)	100
AF1	0.27 (.003)	0.143 (.121)	0.821 (.000)	120
CF1	0.69 (.000)	-0.090 (.505)	0.181 (.174)	58
CF2	0.83 (.000)	0.336 (.001)	0.682 (.000)	100
CF3	0.74 (.000)	0.620 (.000)	0.787 (.000)	118

Relationship between Lip and DCT-coefficient 2. Figure 3 shows the lip position on the abscissa and the DCT coefficient 2 on the ordinate. Table III (third column) gives the results for the correlation between lip position and coefficient 2. No correlation between these two parameters would be in line with the motor equivalence hypothesis. A negative correlation would mean that there was more energy in the lower frequencies if there was more lip protrusion (and thus a longer front cavity). This would suggest that there is no motor equivalence. A positive correlation would mean that there was more energy in the lower frequencies if there was less protrusion. This is only possible if, at the same time, the tongue was retracted.

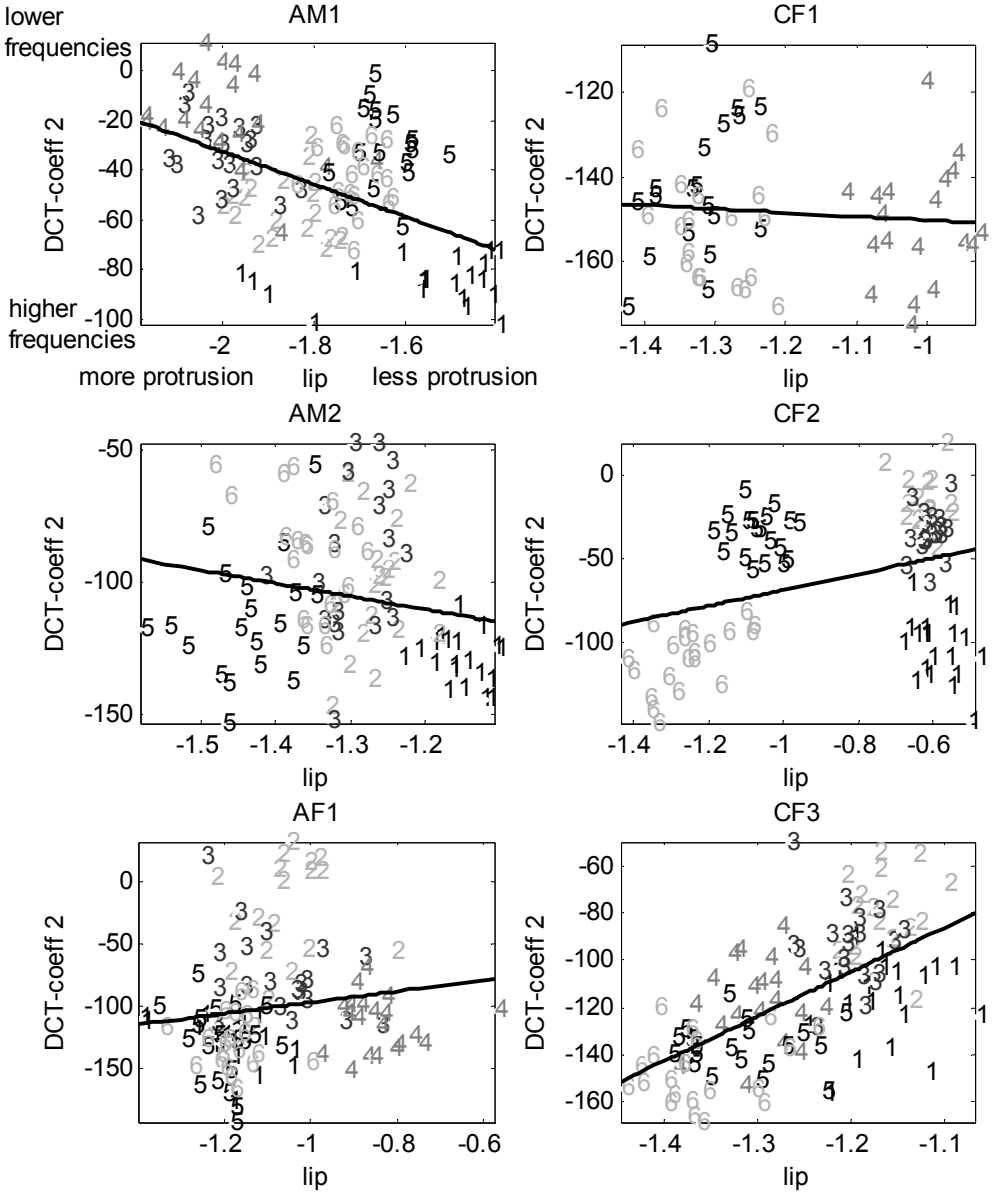

Figure 3: As figure 2 but for lip position vs. DCT-coefficient 2.

Relationship between tongue position and DCT-coefficient 2. Figure 4 and the fourth column in table III give the results of the correlation between tongue position and DCT-coefficient 2. No correlation between the two parameters suggests that there was motor equivalence. A positive correlation means that there was more energy in the lower frequencies when the tongue was retracted and the lip did not compensate (sufficiently) for it. A negative correlation would mean that there was more energy in the lower frequencies when the tongue was more advanced. This is only possible if the lip overcompensates.

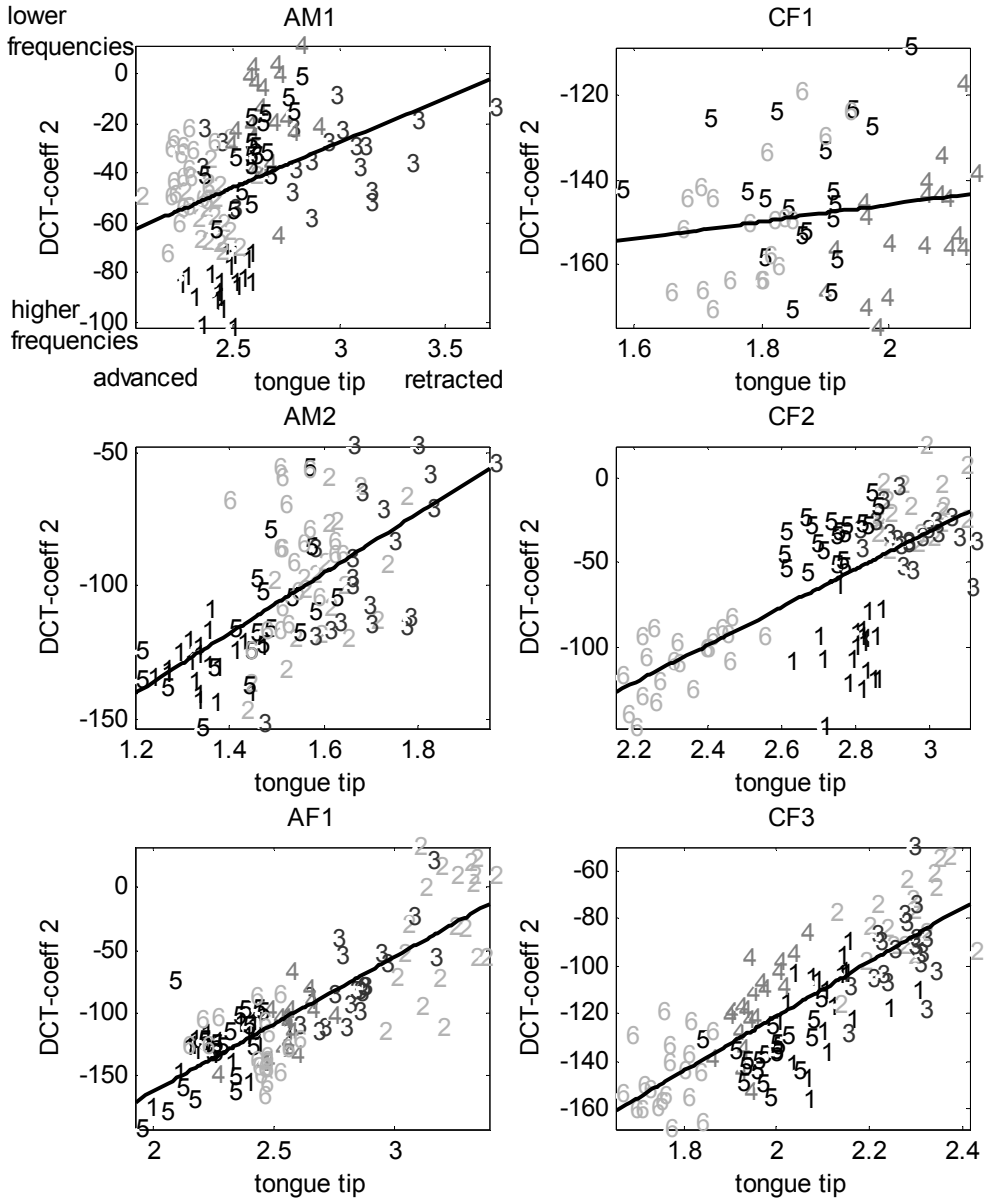

Figure 4: As figure 2 but for tongue tip position vs. DCT-coefficient 2.

Motor equivalence hypothesis. According to the motor equivalence hypothesis there should be a positive correlation between tongue tip position and lip position. Furthermore, there should be no correlation between either one of the articulatory parameters and DCT-coefficient 2. As can be seen in figure 2 and Table III (second column), there is a significant positive correlation between tongue tip position and lip position for four speakers (AF1, CF1, CF2 and CF3), i.e. the tongue was more advanced when the lip was protruded and more retracted when the lip was retracted. Speaker AM1 has a significant negative correlation, speaker AM2 has a nonsignificant negative correlation, suggesting that these two speakers do not use motor equivalence strategies in adaption.

The second part of the motor equivalence hypothesis concerns the relation between the articulatory parameters tongue tip position and lip position and the acoustic parameter DCT-coefficient 2. According to the hypothesis, if speakers are using motor equivalence strategies, the acoustic output should stay similar. Thus, there should be no correlation between either one of the articulatory parameters and DCT-coefficient 2. Speaker CF1, who had a positive correlation between tongue and lip position, has no correlation between articulatory and acoustic parameters (cf. table III, columns 3 and 4, figures 3 and 4), which is perfectly in agreement with the motor equivalence hypothesis. Two of the speakers with a positive correlation between tongue position and lip position, CF2 and CF3, have a significant positive correlation between lip position and coefficient 2 meaning that there was more energy in the higher frequency bands when there was more lip protrusion (i.e. a longer front cavity). This is only possible with a positive correlation between tongue tip and DCT-coefficient 2, which in fact, was found (cf. table III, fourth column, figure 4). These two speakers, CF2 and CF3 thus have insufficient adaptation via the lip for their great variability in tongue position.

The fourth speaker with a positive correlation between lip and tongue, AF1, does not have a significant correlation between lip position and coefficient 2, but a positive correlation between tongue position and coefficient 2. Similar to speakers CF2 and CF3, this speaker thus also has more variability in tongue than in lip position, but to a lesser degree than speakers CF2 and CF3.

The remaining two speakers, AM1 and AM2, who do not have positive correlations between tongue and lip, have negative correlations between lip and coefficient 2 and positive correlations between tongue and coefficient 2, as would have been expected for no motor equivalence.

Summarizing the results, the speakers can be classified as follows:

(1) No motor equivalence: speakers AM1 and AM2, no positive correlation between articulatory parameters but correlations between articulatory and acoustic parameters.

(2) Perfect motor equivalence: speaker CF1, positive correlation between tongue and lip, no correlation between articulatory and acoustic parameters.

(3) Partial motor equivalence: speakers CF2, CF3 and AF1, positive correlation between tongue and lip, but more variability in tongue than in lip position leading to a positive correlation between tongue position and coefficient 2.

Articulatory landmark hypothesis. According to the articulatory landmark hypothesis there should be more variability in tongue position for speakers with a central palate than for speakers with an alveolar palate. As a consequence there should also be more motor equivalence in those speakers.

The two speakers for whom there is no motor equivalence are speakers with an alveolar prosthesis. In addition to that, the correlation is not very strong for the third speaker with this type of prosthesis. At first sight it seems that this is in line with the articulatory landmark hypothesis. However, according to this hypothesis the landmark should reduce the articulatory variability of the tongue tip during the perturbed sessions and this should be the reason for a reduction of the motor equivalence. Looking at the variability in tongue tip positions in the perturbed sessions (2wp, 3pe, 4pe, 5pe), this does not seem to be true. For speakers AM1 and AF1 the tongue tip position varied about 15mm and for speaker AM2 about 8 (figure 2). The speakers with the central palates had between 5 and 10mm variability. So, although there is more motor equivalence for speakers with a central palate than for speakers with an alveolar palate this cannot be attributed to the presence or absence of an articulatory landmark.

Auditory feedback hypothesis: According to the auditory feedback hypothesis, if auditory feedback is absolutely essential in adaptation speakers should not be able to adapt when their auditory feedback is masked. The acoustic result of their productions should differ from the productions with auditory feedback available. Also, the productions without auditory feedback should not “fit in” with the covariation of tongue and lip. Comparing session 2wp to all other sessions, this session usually does not fall out of the correlation between tongue tip and lip (if there is one, cf. figure 2). Some speakers had a higher DCT-coefficient in this session than in the other sessions (CF2, AF1, CF3, cf. figures 3 and 4). Except for speaker AF1, this acoustic difference between sessions 1np and 2wp was, however, quite small, especially if one compares it to the overall acoustic variability in all sessions together. This suggests that adaptation is to a high degree possible even without auditory feedback.

For three speakers (AM2, CF2, CF3), the productions in session 2wp are very similar to the ones in 3pe. So, in contrast to the auditory feedback hypothesis, adaptation is not impossible without auditory feedback even if it is mainly restricted to keeping the articulatory strategy unchanged as much as possible. Generally, it is not possible to say whether changes between session 2wp and 3pe are solely due to the lack vs. presence of auditory feedback or whether they are at least in part due to practice gained in the course of session 2wp.

DISCUSSION

To summarize the results, in line with the hypothesis the majority of the speakers in this study showed the expected covariation of tongue and lip (cf. correlation of tongue and lip sensors in figure 2). Contrary to the hypothesis, however, three of these speakers had more tongue movement than lip movement resulting in a correlation between lip movement and DCT2 (cf. figure 3) so that the acoustic output was not entirely independent of the two articulatory parameters but depended on the tongue position. So the motor equivalence hypothesis might be reformulated: Speakers covary their articulatory positions not in order to keep the acoustic output entirely stable, but to reduce variability in the acoustic output. This is also in line with studies such as Zhou et al. (2008) showing that in motor equivalence in American English /r/ there is acoustic variability in the higher formants. The acoustic variability is kept within a range which makes it possible for a potential listener to recognize the sound.

The lip compensated only partly for changes in tongue position. This led to some acoustic variability which, given that the degree of compensation did not increase any further towards the end of the adaptation phase, seemed to be tolerated by the speakers. Speaker CF3 shows this very

clearly. She had a clear correlation between tongue and lip position, moreover a positive correlation between tongue position and coefficient 2 as well as between lip position and coefficient 2. The tongue position varied by about 8mm whereas the lip position varied by 4mm only. Thus, if the tongue was fronted the lip was fronted as well but not as much as would be necessary to keep the cavity the same length.

Especially for the alveolar prosthesis one could have expected that the speakers produce the constriction at a more posterior place of articulation. This would have resulted in a very clear contrast between the postalveolar and the alveolar fricative. However, the data show that at least some of the speakers produce a more advanced constriction position for the postalveolar which is compensated for by more lip protrusion by some of the speakers. This result is in line with the assumption that speakers do not simply aim at maximizing the contrast between phonemes but that fine phonetic detail matters and the speakers consequently try to produce a fricative which they regard as typical for a German speaker.

A further observation was that motor equivalence was clearer for speakers with a central artificial palate than for speakers with an alveolar artificial palate. There was only one speaker with an alveolar palate showing a positive correlation between the two articulatory parameters, and this speaker actually had the weakest of the significant positive correlations between the articulatory parameters. The articulatory landmark hypothesis suggested that speakers with an alveolar palate would use the alveolar ridge as a landmark, which would reduce the variation in tongue position in the perturbed sessions. This was not observed. The lack of motor equivalence for these speakers can therefore not be ascribed to too little variability in tongue position due to

the use of the alveolar ridge as a landmark, possibly because the speakers noticed that the alveolar ridge was at a more posterior place and therefore did not regard it as a useful landmark.

Looking at the results of session 2wp in figures 2 to 4 it seems that, except for speaker AF1, missing auditory feedback did not have much of an influence on the relationship between tongue and lip position or between either of the articulatory parameters and the acoustics. There was a usually small change in tongue position when no auditory feedback was available which was not compensated for by a change in lip position. However, the speakers did not compensate with auditory feedback either (session 3pe) suggesting that the acoustic change induced by the tongue position change was negligible. Thus, although the results do not show that speakers use motor equivalence strategies without auditory feedback, they suggest that speakers can adapt up to a certain degree even without auditory feedback. This is in line with a finding in Perkell et al. (2004) suggesting that the goals for sibilants are to a high extent somatosensory rather than auditory.

Although the results for single sessions are not given here, it is evident from figure 2 that there are no correlations within a session except for some few exceptions (e.g. sessions 5pe and 6np of speaker CF2). One could wonder why, if speakers were using motor equivalence to a certain degree, they did not do this within a session. We can think of two reasons for that, a methodological one and one concerning speech motor learning. The first reason for not finding correlations between the two articulatory parameters could be that there are too few repetitions within a session. As demonstrated by Perkell et al. (1993) for motor equivalent strategies in American English /u/ in unperturbed speech about 300 repetitions are needed until a correlation between lip protrusion and tongue body retraction can be found. In the present study, only 20

repetitions per session were available. A second possibility would be that speakers optimize their articulatory patterns over the adaptation process (Schulz et al., 2001, Tiede et al., 2009). An optimization process would necessarily reduce the variability so that no correlation will be found. As can be seen in our data, the development over sessions is not linear, i.e. speakers do not for example move from a configuration with little lip protrusion and a retracted tongue towards one with much lip protrusion and an advanced tongue. Rather, they seem to try out another strategy on each adaptation day. It is possible that they have already optimized the strategy they are using in each session, so that the articulatory variability within the session is reduced.

The results are partly in agreement with studies investigating adaptation to electropalatography palates, and they can expand the knowledge gained by these earlier studies. A study by McAuliffe et al. (2008) shows that not all adult speakers are able to adapt to an EPG palate. In that study adaptation to the presence of an EPG palate was investigated. It is shown that within three hours two of the three subjects were able to adapt /s/, /ʃ/ and /t/ perceptually, but the third subject was not. In a similar study McAuliffe et al. (2007) found that speakers were able to adapt within 45 min to 3 hours. Vowel durations and vowel formant frequencies were found to be virtually unaffected by the palate whereas especially /s/ (but not so much /ʃ/) was severely affected immediately after the insertion of the palate. All speakers had adapted to some extend after 3h, most of them earlier (at 45 min).

The results of the present study are in agreement with these earlier studies by showing that there are speaker specific differences. Some speakers adapt better than others. As we suggest in a follow up study of the present one (Brunner et al. in press a) this could be due to differences in the auditory acuity of these speakers.

Although our palatal prostheses were much thicker than standard EPG palates, the acoustic measurements show that speakers adapted quite quickly (in less than 45 minutes). This difference in adaptation time is probably due to the fact that we focused on the location of the main spectral peak and ignored the higher frequencies. As our results for /s/ from a previous perturbation study (Brunner et al., in press b) suggest, the adaptation of the higher frequencies might take up to two weeks for thick palates.

Our results might help to explain findings from some other earlier studies. Timmins et al. (2009), for example, investigated the production of /ʃ/ in young people with Down's syndrome. They show that many of those speakers, who differ from typically developing children in their palate shape, managed to produce perceptually acceptable productions of the fricative even if the articulation differed as compared to typically developing children and adults. The speakers often produced a retracted groove. Our results suggest that these speakers might have succeeded in producing these sounds by changing the degree of lip protrusion. In a similar study Cleland et al. (2009) report two young people with Down's syndrome who were able to produce perceptually adequate productions of /ʃ/ with fronted articulations. Following our results these speakers can be assumed to use more lip protrusion.

The measurement for front cavity size (horizontal difference between tongue tip and upper lip sensor) is fairly rudimentary in the present study. Future studies could use additional techniques, for example MRI, to provide more exact information on front cavity dimensions.

Literature

Brunner, J., Ghosh, S.S., Hoole, P., Matthies, M., Tiede, M. & Perkell, J. (in press a). The influence of auditory acuity on acoustic variability and the use of motor equivalence during adaptation to a perturbation. *Journal of Speech, Language, and Hearing Research.*

Brunner, J., Hoole, P. & Perrier, P. (in press b). Adaptation strategies in perturbed /s/. *Clinical Linguistics and Phonetics.*

Cleland, J., Timmins, C., Wood, S.E., Hardcastle, W.J. & Wishart, J.G. (2009). Electropalatographic therapy for children and young people with Down's syndrome. *Clinical Linguistics and Phonetics*: 23(12):926-939.

Delattre, P. & Freeman, D. (1968). A dialect study of American R's by x-ray motion picture. *Linguistics,* 44: 29-68.

Diehl, R.L. & Kingston, J. (1991). Phonetic covariation as auditory enhancement: The case oft he [+voice]/[-voice] distinction. In: O. Engstrand & C.Kylander (eds.). *Current Phonetic Research Paradigmsn: Implications for Speech Motor Control*, PERILUS, volume 14, University of Stockholm, 139-143.

Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.

Folkins, J.W. & Zimmermann, G.N. (1982). Lip and jaw interaction during speech: Responses to perturbation of lower-lip movement prior to bilabial closure. *Journal of the Acoustical Society of America*, 71(5): 1225-1233.

Forrest, K., Weismer, G., Milencovic, P. & Dougall, R.N. (1988). Statistical analysis of word-initial voiceless obstruents: Perliminary data. *Journal of the Acoustical Society of America*, 84(1): 115-123.

Gracco, V.L. & Abbs, J.H. (1985). Dynamic control of the perioral system during speech: Kinematic analyses of autogenic and nonautogenic sensorimotor processes. *Journal of Neurophysiology*, 54(2): 418-432.

Guenther, F.H. (1994). A neural network model of speech acquisition and motor equivalent speech production. Biological Cybernetics, 72: 43-53.

Guenther, F.H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, 102: 594-621.

Guenther, F.J. (2003). Neural control of speech movements. In: A. Meyer and N. Schiller (eds.), Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities. Berlin: Mouton de Gruyter.

Guenther, F.H., Hampson, M. & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, 105: 611-633.

Guzik, K. & Harrington, J. (2007). The quantification of place of articulation assimilation in electropalatographic data using the similarity index (SI). *Advances in Speech-Language Pathology*, 9: 109-119.

Hamlet, S.L. & Stone, M. (1978). Compensatory alveolar consonant production induced by wearing a dental prosthesis. *Journal of Phonetics*, 6: 227-248.

Honda, M., Fukino, A. & Kaburagi, T. (2002). Compensatory responses of articulators to unexpected perturbation of the palatal shape. *Journal of Phonetics* 30(3), 281-302.

Honda, M. & Murano, E. (2003). Effects of tactile and auditory feedback on compensatory articulatory response to an unexpected palatal perturbation. In: Palethorpe, S. & Tabain, M. (eds). Proceedings of the 6th Speech Production Seminar, Sydney, 97-100.

Hoole, P. (1987). Bite-block speech in the absence of oral sensibility. *Proceedings of the 11th International Congress of Phonetic Sciences*, 4. 16-19.

Hoole, P. (2006). *Experimental studies of laryngeal articulation*. University of Munich, Habilitation.

Hoole, P. & Honda, K. (2011). Automaticity vs. feature-enhancement in the control of segmental F0. In: N. Clements & R. Ridouane (eds.) *Where do phonological contrasts come from? Cognitive, physical and developmental bases of phonological features*. John Benjamins (Series . *Language Faculty and Beyond (LFAB): Internal and External Variation in Linguistics*), pp.131-171.

Jain, A.K. (1989). *Fundamentals of Digital Image Processing*. Englewood Cliffs, New Jersey: Prentice Hall.

Jones, J.A. & Munhall, K.G. (2003). Learning to produce speech with an altered vocal tract: The role of auditory feedback. *Journal of the Acoustical Society of America*, 113(1):532-543.

Jordan, M.I. (1996). Computational aspects of motor control and motor learning. In: H. Heuer & S. Keele, (eds.). *Handbook of Perception and Action: Motor Skills*. New York: Academic Press.

Kelso, J.A.S. & Tuller, B. (1983). "Compensatory articulation" under conditions of reduced afferent information: A dynamic formulation. *Journal of Speech and Hearing Research*, 26: 217-224.

Kingston, J. & Diehl, R.L. (1994). Phonetic knowledge. Language, 70: 419-454.

Laboissière, R. (1992). *Préliminaires pour une robotique de la communication parlée: inversion et contrôle d'un modèle articulatoire du conduit vocal.* Doctoral dissertation. Institut National Polytechnique de Grenoble.

Lawson, E., Scobbie, J.M. & Stuart-Smith, J. (in press). The Social Stratefication of Tongue Shape for Postvocalic /r/ in Scottish English. *Journal of Sociolinguistics.*

Lindblom, B., Lubker, J. & Gay,T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics*, 7: 147-161.

Linke, D. (1980). Vorprogrammierung und Rückkopplung bei der Sprache. In: M. Spreng (ed.) *Interaktion zwischen Artikulation und akustischer Perzeption*. Stuttgart: Thieme. 50-57

Löfqvist, A., Baer, T., McGarr, N.S. & Story, R.S. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America*, 85: 1314-1321.

McAuliffe, M.K., Lin, E., Robb, M.P. & Murdoch, B.E. (2008). Influence of a standard electropalaography artificial palate upon articulation. A preliminary study. *Folia Phoniatrica Logop.* 60(1):45-53.

McAuliffe, M.J., Robb, M.P. & Murdoch, B.E. (2007). Acoustic and perceptual analysis of speech adaptation to an artificial palate. *Clinical Linguistics and Phonetics*, 21:885-894.

McFarland, D.H. & Baum, S.R. (1995). Incomplete compensation to articulatory perturbation. *Journal of the Acoustical Society of America*, 97(3): 1865-1873.

Mooshammer, C., Hoole, P. & Geumann, A. (2007). Jaw and order. *Language and Speech*, 50(2):145-176.

Perkell, J.S., Matties, M.L., Svirsky, M.A. & Jordan, M.I. (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, J., Guenther, F., Lane, H., Matthies, M., Perrier, P., Vick, J., Wilhelms-Tricarico, R. & Zandipour, M. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics*, 28(3): 233-272.

Perkell J. S., Guenther F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America*, *116*, 2338–44.

Savariaux, C., Perrier, P. & Orliaguet, J.-P. (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(5): 2428-2442

Savariaux, C., Perrier, P., Orliaguet, J,-P. & Schwartz, J.-L. (1999). Compensation strategies for the perturbation of French [u] using a lip tube II. Perceptual analysis. *Journal of the Acoustical Society of America*, 106(1): 381-393.

Schulz, G.M., Stein, L. & Micallef, R. (2001). Speech motor learning: preliminary data. *Clinical Lingustics and Phonetics*, 15 (1&2):157-161.

Shadle, C. (1985). The acoustics of fricative consonants. *Technical Report* 506, Research Laboratory of Electronics, MIT Cambridge, MA.

Shadle, C. (1990). Articulatory-acoustic relationships in fricative consonants. In W.J. Hardcastle and A.Marchal (eds). *Speech Prodution and Speech Modelling*. Kluwer Academic Publishers: 187-209.

Shadle, C., Berezina, M., Proctor, M. & Iskarous, K. (2008). Mechanical Models of Fricatives Based on MRI-derived Vocal Tract Shapes. *Proceedings of the 8th International Seminar on Speech Production*, Strasbourg: 417-420.

Tiede, M., Mooshammer, C., Goldstein, L., Shattuck-Hufnagel, S., Perkell, J. & Matthies, M. (2009). Optimization of articulator trajectories in producing learned nonsense words. *Journal of the Acousticial Society of America*, 125(4): 2499.

Timmins, C., Cleland, J., Wood, S.E., Hardcastle, W.J. & Wishart, J.G. (2009). A perceptual and electropalatographic study of /S/ in young people with Down's syndrome. *Clinical Linguistics and Phonetics*. 23(12):911-925.

Waldstein, R.S. (1990). Effects of postlingual defness on speech production: Implications for the role of auditory feedback. Journal of the Acoustical Society of America, 88: 2099-2114.

Watson, C. & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *Journal of the Acoustical Society of America*, 106(1): 458-468.

Westbury, J., Hashi, M. & Lindstrom, M. (1998). Differences among speakers in lingual articulation of American English /r/. *Speech Communication*, 26: 203-226.

Westbury, J.R., Lindstrom, M.J. & McClean, M.D. (2002). Tongues and lips without jaws: A comparison of methods for decoupling speech movements. Journal of Speech, Language, and Hearing Research, 45: 651-662.

Zhou, X., Espy-Wilson, C., Tiede, M., Boyce, S. (2007). Acoustic cues of "retroflex" and "bunched" American English rhotic sound. *Journal of the Acoustical Society of America*, 121(5): 3168.

Zhou, X., Espy-Wilson, C., Boyce, S., Tiede, M., Holland, C. & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/. *Journal of the Acoustical Society of America*, 123(6): 4466-4481.

Appendix: Acoustic analysis

Three parameters that have been used previously to describe spectral tilt were tested for their suitability to describe the relation between spectral tilt and size of the front cavity, i.e. center of gravity, skewness and the second coefficient of a discrete cosine transform.

Center of gravity and skewness were calculated for a segment of 30 ms around the sibilant midpoint (means over 6ms windows with 1ms overlap and a preemphasis factor of 1 according to the formulae given in Forrest et al., 1988).

As a third spectral parameter, the second coefficient of a discrete cosine transform (DCT) was calculated. This method was adapted from Watson & Harrington (1999), who used it to describe formant trajectories, Guzik & Harrington (2007) who used it to classify fricative spectra and Brunner et al. (in press b) who used it to analyze spectra of perturbed and unperturbed /s/. The method is described in detail in Watson & Harrington (1999)[i]. A discrete cosine transform has a number of similarities with a discrete Fourier transform: The signal (here a spectrum) is decomposed into a number of sinusoids of different frequencies. However, in contrast to the DFT the sinosoids of the DCT are phaseless and they are at half integer cycles (k=0, 0.5, 1…) instead of full cycles. Each coefficient of the DCT describes a certain characteristic of the spectrum. The first coefficient (k=0), for example, gives the mean amplitude of the spectrum. The second coefficient (k=0.5) gives the tilt of the spectrum and therefore appeared suitable for the present purpose. Very basically one can say that if the spectrum has more energy in the lower frequencies, the second coefficient is high, if it has more energy in the higher frequencies, it is low.

If speakers are covarying lip protrusion and tongue position this should result in a stable front cavity length leading to a stable acoustic output. It is this relation which is of interest for this study, therefore the acoustic parameter to be used should be suitable to describe this relation. In order to find out which of the three spectral measurements, center of gravity, skewness or DCT-coefficient 2, is best suited for the present purpose correlations between each acoustic parameter and the front cavity length were calculated. The length of the front cavity was estimated as the horizontal distance between the sensor at the tongue tip and the one at the upper lip at the temporal midpoint of the acoustic segment. In order to find out which acoustic parameter is best suited for describing the acoustic consequences of variation in front cavity length Pearson correlations between each of the three acoustic parameters (center of gravity, skewness, DCT-coefficient 2) and the length of the front cavity were calculated.

Table II gives the correlation coefficients and p-values (in parenthesis) for the relation between the length of the front cavity and the three acoustic parameters. The first column gives the speaker, the last column the number of productions taken into account. The bold figures give the strongest one of the three correlations for a given speaker.

The correlation coefficients for the center of gravity (second column) are negative: The center of gravity was lower for longer front cavities. All correlations for the COG are significant. The correlations between skewness and length of front cavity are not as clear, presumably due to the low upper cut off frequency. The correlations should be positive (higher skewness values for longer front cavities) but they are not for all speakers. Also, they are not significant for all speakers. For most speakers the correlation coefficients are highest for the DCT-coefficient 2 (see

bold numbers giving best correlations for a speaker). All correlations between front cavity length and coefficient 2 are positive (higher coefficients for long front cavities) and significant.

Table II: Correlations between length of front cavity and the acoustic parameters. First column: speaker, second to fourth column: correlation coefficients and significances (in parentheses) of the correlations between front cavity length and center of gravity, skewness and coefficient 2, respectively, fifth column: number of productions. Highest correlation coefficients for a speaker are bold.

Speaker	**COG**	**Skewness**	**Coeff2**	**N**
AM1	-.384 (.000)***	-.095 (.304)	**0.520 (.000)** ***	119
AM2	**-.686 (.000)** ***	.494 (.000) ***	0.623 (.000) ***	100
AF1	-.641 (.000) ***	.162 (.077)	**0.781 (.000)** ***	120
CF1	**-.492 (.000)** ***	.426 (.001) *	0.338 (.010) *	58
CF2	-.237 (.017) *	-0.243 (.015) *	**0.443 (.000)** ***	100
CF3	-.538 (.000) ***	.282 (.002) **	**0.699 (.000)** ***	118

Given these results all further analyses are based on the DCT-coefficient 2 because it describes the acoustic result of front cavity variation best. As stated above, a lengthening of the front cavity leads to a downward shift in the main energy concentration of the spectrum, whereas a shortening leads to an upwards shift. A downwards shift of the energy concentration is mirrored by a higher DCT coefficient 2, an upwards shift is mirrored by a lower DCT coefficient 2 (Guzik & Harrington, 2007). The correlations are also in line with the assumption that this parameter, applied to the filtered signal, predominantly describes the relation between front cavity size and acoustics and not the acoustic effects of other articulatory maneuvers which speakers might have carried out during adaptation.

[i] In contrast to Watson & Harrington and Guzik & Harrington, our Hz-spectra were not transformed into Bark because the length of the front cavity is mirrored primarily in the location of the central peak which can be estimated well from band-pass filtered Hz-spectra. Furthermore, we did not use the formula given in Watson & Harrington (1999) but the MATLAB implemented dct-function which is based on a formula given in Jain (1989). The two methods differ only in the absolute values of the coefficients that are in constant ratios of sqrt(N/2) with N being the number of points in the spectrum. Furthermore, as indicated, we used a different spectral range (700 Hz to 6kHz).