Adaptation strategies in perturbed /s/

Jana Brunner, Phil Hoole, Pascal Perrier

Adaptation strategies in perturbed /s/

Jana Brunner[1], Phil Hoole[2], Pascal Perrier[3]

[1]Institut für Germanistik, Universität Potsdam, Potsdam, Germany

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

[3]DPC/GIPSA -lab,UMR CNRS 5216, Grenoble INP, Grenoble, France

Correspondence: Jana Brunner
Institut für Germanistik
Universität Potsdam
Am Neuen Palais 10
14469 Potsdam, Germany.
Tel: 49 331977 1512.
Fax: 49 331972 163.
E-mail: jana.brunner@uni-potsdam.de

Abstract

The purpose of the work is to investigate the role of three articulatory parameters (tongue position, jaw position and tongue grooving) in the production of /s/. Six normal speakers' speech was perturbed by a palatal prosthesis. The fricative was recorded acoustically and via EMA in four conditions: (1) unperturbed, (2) perturbed with auditory feedback masked, (3) perturbed with auditory feedback available, (4) perturbed after a two week adaptation period. At the end of the adaptation speakers produced more high frequency noise while either having a higher jaw position or more grooving of the tongue or both. We discuss the potential clinical implications of the results with regard to the role of jaw height and tongue grooving in the treatment of impaired /s/.

KEY WORDS: articulation, palate, speech sound, speech intelligibility, perturbation

The alveolar fricative /s/ is a sound which is very complex and acquired late in speech acquisition. Robb and Bleile (1994) report an acquisition age of 19 months for /s/ in initial position for glossable and non-glossable speech, which is much later than for example /t/ which is already present at 8 months. Stoel-Gammon (1985), investigating glossable speech only, reports that /s/ was present at 24 months in initial and final position.

The fricative /s/ is also a sound which is frequently impaired (cf. e.g. Wilcox, Stephens & Daniloff, 1985; Gibbon & Hardcastle, 1987). There are several types of impairment, the most frequently occurring types being interdental, dentalized, lateral and palatal lisp. Most normally developing children take several years to acquire the correct production of the sound; however, there is still quite a high number of children needing treatment.

Treatment methods usually include perception training, which is followed by a production training (van Riper & Irwin, 1984) mainly based on a trial-and-error basis (Günther & Hautvast, 2009) with a number of explanations by the therapist, usually with regard to the place of articulation and the necessity of grooving the tongue. A number of therapists use visual aids comparing the spectral shape of the patient's production with that of a model speaker. Römer, Willmes & Kröger (2009) present a system which shows a spectrum of the patient's production and indicates the range of a 'correct' production so that the patient can judge his/her own production to be correct or incorrect. These methods which are based on acoustic analysis only, however, rely on the patient finding by himself an articulatory configuration which will produce the sound correctly.

Other systems offer more help with regard to the articulation by showing articulatory features of the patient's production. Gibbon & Hardcastle (1987), for example, describe the use of EPG in speech therapy. For the treatment of /s/ impairment this method

is quite helpful because it gives the patient feedback about two important characteristics of /s/: the constriction position and tongue grooving. Acoustically, the constriction position determines the frequency of the main spectral peak (Shadle, 1985) which is usually around 5kHz. Formation of a groove has an influence on the higher frequencies. Shadle, Berezina, Proctor, & Iskarous (2008) show in a modeling study that grooving of the tongue led to more energy in the high frequencies from 8 kHz upwards.

Production and perception of tongue grooving was investigated in a study by Fletcher and Newman (1991). Three speakers were asked to produce the two fricatives /s/ or /ʃ/ with different groove width (judged by the amount of EPG-tongue-palate contact) and different groove locations (fronted-retracted at the alveolar ridge). The speakers were not specially trained to perform this task, but they were provided with visual feedback about their productions from the EPG-palates they wore. Afterwards, listeners were asked to say whether the sound they heard was /s/ or /ʃ/. Sibilants which were produced more fronted and with a narrower groove were frequently judged as /s/ whereas sibilants with a more retracted and wider groove were judged as /ʃ/.

The importance of the constriction position and grooving of the tongue in /s/-production seems to be undisputed. Other work on normal speech, however, has shown that there is yet another characteristic of /s/ which is about as important for the acoustic outcome as the other parameters, namely the jaw position. Numerous studies have shown that the jaw has a very high position in sibilants (e.g. Amerman, Daniloff & Moll, 1970; Shadle, 1991; Lee, Beckman & Jackson, 1994; Mooshammer, Hoole & Geumann, 2006, 2007). The reason which is usually given is that in contrast to other fricatives such as /ç/ and /x/ the

sibilants /s/ and /ʃ/ have a different source mechanism. Stevens (1971) argues that sibilants are produced when moving air strikes the teeth to produce a turbulent wake. One study investigating this in depth is Shadle (1991). With the help of anatomic and acoustic data of a single speaker Shadle built various models of the vocal tract with built-in microphones at the constrictions of different fricatives and at the lower incisors. She then measured the acoustic signal at these different places in the vocal tract and compared it to the signal measured in the far-field. The vocal tract signal which was least similar to the one in the far-field was assumed to be measured at the noise source. For /ç/ and /x/ the source was identified at the constriction whereas for the two sibilants the noise source was detected at the incisors. Shadle concluded that /ç/ and /x/ have a 'wall' source, because the noise is created when air impinges on the palate, whereas /s/ and /ʃ/ have an 'obstacle' source, because most of the noise is created not at the exit of the constriction but when the sound impinges on the incisors which could be seen as an obstacle within the vocal tract.

Shadle (1990) compared outputs of a model with an obstacle and one without. The results show that the obstacle led to more energy in /s/ in the ranges from 6 to 10 kHz. Mooshammer et al. (2007) further stressed the importance of the jaw in /s/-production by showing that the high jaw position is not strictly related to the tongue position, which means that the jaw is not simply high because /s/ requires a high tongue position. If the jaw was simply supporting the tongue it should be higher when the tongue position is higher, and lower when the tongue is lower. A comparison with other coronal sounds (/l, n, t, d/), however, showed that although /s/ had a rather low tongue tip position it still had a high jaw position.

Howe & McGowan (2005) also discuss the role of the incisors in /s/-production but they go even further in stating that the acoustically relevant effect of a sibilant is caused not only by the air jet impinging at the upper or lower incisors but by the diffraction of the air at the edges of the upper and lower teeth caused by a small horizontal distance between them. Following from that, speakers not only need to have the jaw in a high position – and thus control the vertical distance between the upper and lower incisors - but they also need to control the horizontal distance between them.

To summarize, producing /s/ requires a very precise articulatory configuration. The following factors are important:

- The tongue has to form a vocal tract constriction at the alveolar ridge.
- The tongue has to be grooved.
- The jaw has to be in a rather high position so that the air jet hits the lower incisors.

Each of these factors contributes to the characteristic high frequency noise of this sound. The constriction position in /s/ has a strong influence on the location of the spectral peak; a high jaw and grooving are important for the production of large bandwidth noise with substantial energy above 6 kHz.

Although these articulatory characteristics and their influence on the acoustics of /s/ have been intensively studied, what is so far missing is an investigation of how these articulatory parameters interact in speech production. The first aim of the study is therefore to investigate which one of the three articulatory parameters speakers change when their articulation is perturbed by a palatal prosthesis lowering their palate and, as a consequence, their jaw. The second aim is to investigate the role of auditory feedback on the choice of the parameters when speakers adapt to the perturbation. The third aim is to investigate whether

adaptive behavior would differ depending on the shape of the prosthesis during the production of /s/.

The perturbation device, an artificial palate, was chosen because it has been shown to change the main acoustic characteristic of /s/: it reduces the high frequency energy (e.g. Hamlet & Stone, 1978; McFarland, Baum & Chabot, 1996; Baum & McFarland, 1997; Aasland, Baum & McFarland, 2006). Two types of artificial palate were used which induced different perturbations to the morphology. One of them moved the alveolar ridge posteriorily ('alveolar prosthesis') and one of them filled in the palatal arch and made the palate flat ('central prosthesis'). Each of the prostheses had a maximal thickness of 1 cm, meaning that for the alveolar palates the alveolar ridge was retracted by about 1 cm, and for the central palate the palate was lowered by about 1 cm at the highest point of the palate. Both prostheses resulted in some lowering of the whole inferior surface of the hard palate. Adaptation can be expected to involve a 'recalibration' of the interplay of tongue and jaw which might give new insights into how the articulatory parameters interact in the production of /s/.

Adaptation over a period of two weeks was investigated. At the onset of perturbation speakers' auditory feedback was masked in order to investigate adaptation without auditory feedback available.

Our hypotheses are:

(1) Acoustic adaptation hypothesis. At the onset of perturbation the high frequency energy, which is typical for /s/, will be lost. Over the adaptation speakers will try to find a way to produce high frequency energy (for example by using grooving of the tongue or a high jaw position).

(2) Somatosensory feedback hypothesis. When the speakers' auditory feedback is masked they will adapt differently for different prosthesis types by using somatosensory feedback. Somatosensory feedback (tactile and proprioceptive) will differ for the two prosthesis types: In order to retrieve something close to their usual sensations[1] speakers with an alveolar prosthesis will produce the constriction at a more retracted position, whereas speakers with the central prosthesis will keep the original position (on the anterior-posterior axis) of the constriction they used to have and just lower the tongue.

(3) Auditory feedback hypothesis. As long as no auditory feedback is available speakers will have a lower jaw position than in the unperturbed session in order to keep the constriction size and the tongue-jaw coordination unchanged as compared to the unperturbed condition. When auditory feedback becomes available speakers will notice that there is less high frequency energy (consistent with results for auditory feedback perturbation of /s/ presented in Shiller, Sato, Gracco & Baum, 2009). In order to adapt for that they will move the jaw up and/or use more tongue grooving.

Methods

Artificial palates. Two types of palatal prostheses were used, one which lowered the palate in the alveolar region only and thus moved the alveolar ridge to a more posterior position ('alveolar prosthesis'), and one which lowered the palatal surface by filling out the palatal vault evenly ('central prosthesis'). All prostheses had a maximal thickness of about one centimeter either at the alveolar ridge (alveolar prosthesis) or at the highest point of the

[1] Under perturbation speakers will never have the same sensations as in the unperturbed session simply because they are lacking sensory information from the palate.

natural palate (central prosthesis). From the place of maximal thickness the palates were tapered off towards the front, back and both sides so that there were no ridges at any end of the palates. Figure 1 shows an example for each of the palates from a sagittal perspective.

INSERT FIGURE 1 ABOUT HERE

Palates were made of acrylic and held in place by clamps between the first and second premolar and between the second premolar and the first molar. For the speakers with the alveolar palates the alveolar ridge was effectively moved towards a more posterior location (1.4 cm for speaker AM1, 1.7 for AM2, 1.1 for speaker AF1). For the speakers with a central palate the palate was lowered by 1.1 cm (CF1), 0.9 cm (CF2) and 1.0 cm (CF3). The speakers with a central palate thus had no alveolar ridge left.

Speakers. Six German subjects took part in the study, two males (AM1, AM2) and four females (CF1, CF2, CF3, AF1). Three of them, AM1, AM2 and AF1 were provided with an alveolar prosthesis, the other three, CF1, CF2 and CF3 had a central prosthesis. The speakers were between 25 and 40 years old and spoke Standard German with some regional influence. None of the speakers had a history of speech or hearing problems.

Experimental setup. The articulatory movements of the speakers were recorded via electromagnetic articulography (Carstens Medizinelektronik, Germany). Sensors were placed midsagittally, three on the tongue, one on the jaw, one on each lip. The most anterior tongue sensor was located at approximately 1 cm behind the tongue tip, the most posterior tongue sensor was located on the part of the tongue opposite the end of the hard palate

when the tongue is at rest in the mouth[2]. The remaining tongue sensor was halfway between these two sensors. Reference sensors were placed on the bridge of the nose and above the upper incisors. The reference sensors enabled a correction of the positional data for head movements during the recording. For the present purpose the data from the tongue and jaw sensors have been analyzed.

Acoustic recordings were carried out with a DAT recorder and a Sennheiser MKH 20 P48 microphone. The acoustic signal was downsampled to 24kHz using the Adobe Audition anti-aliasing filter at 11500 Hz with the standard settings resulting in -3dB at 11750 Hz. Since the most important frequencies in /s/ can be found in the range up to 10kHz (Shadle, 1985, Shadle & Scully, 1995, Shadle et al. 2008), data up to 10 kHz are used for the acoustic analysis.

Procedure. There were several recordings, three on the first day and one after a two weeks adaptation period. In the first session the speakers were recorded without the perturbation (henceforth termed unperturbed condition, UN). Then the artificial palate was inserted and the speakers' auditory feedback was masked with white noise (white noise condition, WN). Afterwards, the masking noise was removed and the speakers could adapt with auditory feedback (full feedback condition, FF). Then the speakers returned to their everyday life and were asked to wear the prosthesis all day except during eating and sleeping, to speak as much as possible with it and to make a serious effort to learn to speak 'normally'. Speakers were asked to write down the number of hours they had worn the prosthesis each day. They

[2]The location of the end of the hard palate relative to the tongue surface was determined by marking the posterior end of the hard palate with an oral disinfectant containing a strong purple colouring agent. The speaker was then asked to close the mouth and push the tongue gently against the hard palate, keeping as neutral a position of the tongue as possible. This resulted in the colour mark being transferred to the tongue surface.

were furthermore given a sheet with exercises and asked to read those aloud once a day. All subjects reported having worn the prosthesis between 10 and 16 hours on each day. After two weeks the speakers returned to the laboratory and were recorded once more with the prosthesis in place and full feedback (adapted condition, AD).

Auditory feedback masking. Auditory feedback masking in the second session of the first day was carried out via the presentation of bandpass-filtered white noise (100 Hz-10 kHz) over headphones. The reason for using auditory feedback masking was to investigate the extent to which speakers can adapt with somatosensory feedback only and how this adaptation differs for different palate shapes.

Speech material. The target sound /s/ was recorded in the nonsense word /'zasa/ spoken in a carrier phrase: Ich sah sassa an ('I looked at /'zasa/.'). There were 20 repetitions (randomised with other material, i.e. CVCV sequences consisting of all lingual sounds of German) in each session. Each session took about 20 minutes.

Acoustic analysis. The fricative /s/ was acoustically segmented (friction onset to friction offset) in each utterance. The segmentation was carried out with the software PRAAT (Boersma & Weenink, 1999).

In order to get a first impression of the spectral differences between normal and perturbed spectra, time-averaged power spectra were calculated for a 30 ms segment centered at the midpoint of the fricative using a series of 6 ms windows with 1 ms overlap with a preemphasis factor of 0.98 for each production. Additionally, an ensemble average

power spectrum (over the 30 ms segment) was calculated for each session from the 20 repetitions recorded in one session.

Inspection of power spectra of single productions and ensemble average power spectra showed that across sessions they differed inconsistently in the location of the main spectral peak. The main difference between the unperturbed and the early perturbed spectra consisted in the amount of high frequency energy. An example for one speaker is given in figure 2. This figure shows a mean bark transformed spectrum of the unperturbed session (thick solid line) and of the first perturbed session (thick dotted line). In order to show the variability within the sessions standard deviations are shown as thin lines. The perturbed spectrum has higher amplitudes in the region from 14 to 19 Bark and lower amplitudes in the region from 19 to 24 Bark.

In order to quantitatively assess this difference, the global shape of the bark transformed spectra was characterized by the fourth coefficient of a discrete cosine transform (DCT) of these spectra. This method was adapted from Watson & Harrington (1999), who used it to describe formant trajectories, and Guzik & Harrington (2007) who used it to classify fricative spectra. The method is explained in more detail in the appendix. Leaving aside the effect of energy in the lower frequencies (below 13 Bark) the fourth DCT coefficient will tend towards more positive values for spectra with an energy concentration in medium frequency ranges (about 13-20 bark, 2350-5500 Hz) and towards negative values for spectra with an energy concentration in very high frequency ranges of about 20-24 Bark (5500-10000 Hz). It should therefore tend towards more positive values for the early perturbed than for the unperturbed productions (note that a flat spectrum will give a coefficient of zero).

INSERT FIGURE 2 ABOUT HERE

Positional measurements. The horizontal position of the tongue tip sensor and the vertical position of the jaw sensor were measured in order to track the constriction position and the jaw height. In both cases analysis was performed at the temporal midpoint of the acoustically measured consonant interval.

Midsagittal concavity. Midsagittal EMA data in principle cannot provide information about the third articulatory parameter discussed so far, namely grooving. However, grooving can be inferred from the midsagittal tongue contour. Data presented in previous studies suggest that, if the tongue is grooved in the midsagittal plane the tongue dorsum is lower than the tongue tip. If it is not grooved, the tongue dorsum is higher than the tongue tip (e.g. Ladefoged & Maddieson, 1996: 147 (X-ray contour plots); Stone, Epstein, Li & Kambhanettu, 2006; Narayanan et al., 1995). To put it differently, if the tongue is bunched there is less grooving than if it is flat or even concave. Estimating grooving can thus be done with the help of a midsagittal contour as follows. The tongue shape was described as a quadratic function ($y=ax^2$) by calculating a quadratic interpolation between the three tongue sensor positions in each production. The coefficient a of this function, which is related to the curvature of the midsagittal tongue contour, was taken as a measure of grooving. If a is positive the midsagittal outline of the tongue is concave and there was grooving. If a is negative the outline is convex and there was little or no grooving.

A preliminary investigation of the data showed that the variability in a observed in the data is mainly due to variability in the vertical position of the tongue mid sensor. Tongue tip and tongue back sensor are higher in the unperturbed than in the perturbed

condition, but apart from this difference there is not much change over sessions. The tongue mid sensor position, however, varies in position and produces the variability in a. If it is high (usually as high or even higher than the tongue tip sensor), the tongue is convex, if is lower than tongue tip and tongue back, the tongue is concave.

The coefficient is related to the apical/laminal difference: higher a means that the production is rather apical, lower a means that it is rather laminal (Narayanan et al., 1995).

The coefficient a cannot be interpreted in absolute terms. a=0 does not mean that there is no grooving. Following earlier findings (e.g. Narayanan et al. 1995, Stone, 1991, Badin et al. 2002) we assume that all /s/-productions, even the laminal ones, have some grooving. However, if a is lower in one production than in another one we infer that there is less grooving in this production.

Estimating grooving in terms of midsagittal concavity is of course hypothetical, however, there is firm evidence that these two shape characteristics are linked. Badin et al. (2002) for example show that grooving of the tongue blade goes together with tongue tip raising. Stone et al. (2006) and Stone & Lundberg (1996) show that the tongue tends to be grooved if the middle of the tongue is low.

Statistical analysis. The following statistical analyses were carried out in R (The R Foundation for statistical computing, 2009):

- ANOVAs with Tukey multiple comparisons for data split by speaker for assessing the influence of the factor condition on (1) the fourth coefficient of the DCT (coefficient 4), (2) the horizontal tongue tip position, (3) the vertical jaw position and (4) midsagittal concavity,

- Pearson correlations between vertical jaw position and the parameter a measuring midsagittal concavity.

Results

DCT-coefficient 4. Figure 3 shows the values of the fourth coefficient of the DCT decomposition of the /s/ spectrum for the different sessions. In the left three subplots the results for the speakers with an alveolar prosthesis are given, in the right ones the results for the speakers with a central palate are shown. Within a subplot, each boxplot refers to one session as given in the axis labels (UP: unperturbed, WN: white noise, FF: full feedback, AD: adapted). Recall that the smaller the coefficient, the higher is the proportion of energy in the 20-24 Bark frequency band, and the larger the coefficient the higher is the energy in the 12-20 Bark frequency band.

The statistical analysis shows that the influence of the condition on coefficient 4 is highly significant ($p<.001$) for all speakers except CF2 (cf. table 1). In line with the results of earlier studies, lower values (meaning more high frequency energy) were found in the unperturbed (UP) condition and an increase of this coefficient (corresponding to an increase of spectral energy below 5kHz and/or a decrease above) was observed in the WN condition. Pairwise comparisons between the white noise session (WN) and the full feedback sessions (FF) show significant differences for only two subjects. This suggests that auditory feedback in general had no major immediate influence on the acoustic adaptation. However, in the cases where there is a significant difference between the WN and the FF-session (i.e. for speakers CF1 and CF3) coefficient 4 is lower in the FF-session than in the WN-session. In these cases there is thus an increase in energy in the 20-24 Bark frequency band from the WN to the FF session. In the adapted session (AD) almost all speakers had regained lower coefficient

values, close to the ones measured in the UP condition (no significant differences between UP and AD for 4 subjects). No differences between prosthesis types were observed.

There was substantial variability within conditions WN and AF. This could be because of adaptation within a session. Single repetitions were investigated for trends over a session; however, for none of the speakers were such trends observed.

To summarize the acoustic results, in line with earlier findings there was a decrease in high frequency energy at perturbation onset. In line with the acoustic adaptation hypothesis (1) speakers produce more high frequency energy towards the end of the adaptation. In contrast to the auditory feedback hypothesis (3) however, only a minority of speakers did produce more energy immediately when auditory feedback became available.

INSERT FIGURE 3 ABOUT HERE

INSERT TABLE 1 ABOUT HERE

Horizontal tongue tip position. Figure 4 shows the results for the horizontal tongue tip position at the midpoint of the acoustically measured interval during the four sessions. For all speakers there was a significant influence of the condition on the horizontal tongue position (cf. table 2). We expected speakers to try to reach the original position in order to keep the front cavity the same length from the FF-session onwards. Contrary to our expectation, however, only sometimes did the adaptation strategy seem to be directed towards reaching the original horizontal tongue position although this should have been possible with sufficient jaw opening. In contrast to the somatosensory feedback hypothesis (2) no consistent differences in tongue position between prosthesis types were found when speakers were adapting without auditory feedback.

INSERT FIGURE 4 ABOUT HERE

As for DCT-coefficient 4, trends over sessions were investigated. Speaker AM1 was the only one for whom trends over sessions were found. This speaker had a more and more fronted tongue during the WN session, and a trend for retraction of the tongue in the FF session. The retraction of the tongue towards the position the speaker had in the unperturbed condition might be influenced by auditory feedback.

INSERT TABLE 2 ABOUT HERE

Vertical jaw position. Figure 5 shows the results for the vertical jaw position. If the upper incisors are taken as a spatial reference, due to the remodeling of the palate the tongue should be lower in the constrictional region as compared to the unperturbed condition so that the air jet becomes lower, too. Therefore one could expect that the jaw should be lower as well so that the lower incisors have the same relative position to the air jet as in the unperturbed condition. However, as shown by Howe & McGowan (2005) the distance of the lower incisors relative to the upper incisors influences the acoustic output as well. Therefore, speakers can be expected to keep this distance unchanged by attempting to reach the jaw position they used to have in the unperturbed condition.

INSERT FIGURE 5 ABOUT HERE

Just after the insertion of the prosthesis (UP vs. WN) four speakers lowered the jaw, the two others kept it constant. When auditory feedback became available, all speakers moved back towards the position which they had in the unperturbed production (although the difference is not significant for speaker CF1, cf. table 3). After two weeks adaptation time all speakers with an alveolar prosthesis either returned to their high original jaw position (AM1 and AM2) or they even elevated the jaw more than in the UP condition (AF1). Two of the speakers with a central prosthesis (CF1 and CF2) had significantly lower positions in the AD than in the UP condition and the third speaker had no change in jaw position.

The adaptation in the full feedback condition could be explained by speakers' motor learning driven by the aim to produce spectral characteristics similar to those of the unperturbed condition. This is consistent with the acoustic adaptation hypothesis (1) and the auditory feedback hypothesis (3). However, the strategies developed to reach this acoustic aim are speaker dependent. All speakers with an alveolar prosthesis used a raised jaw, so that the lower incisors function as obstacles to the air jet. This is also the strategy used by one of the speakers with a central prosthesis (CF3). The other subjects with a central prosthesis, CF1 and CF2, however, seem to develop other strategies.

The process underlying the elaboration of the final compensatory strategy across sessions seems to be speaker-dependent, too. Speaker AM1 had a trend for jaw lowering during the WN session and for jaw raising during the FF session. The other speakers did not show trends over sessions.

In the WN session speakers lower the jaw by just a few millimeters. This is considerably less than the thickness of the prosthesis. A reason for this could be speakers' use of somatosensory feedback (cf. Lindblom & Lubker (1985) showing that speakers are

more aware of their jaw position than of their tongue position, and Nasir & Ostry (2008) suggesting that cochlear implant speakers have very precise somatosensory representations for jaw movements).

INSERT TABLE 3 ABOUT HERE

Midsagittal concavity. Figure 6 shows the results for midsagittal concavity. If coefficient a is positive, the tongue shape is considered to be concave and this is interpreted as evidence for more grooving. If the value is negative the tongue shape is considered to be convex and this is interpreted as evidence for less grooving. All except one speaker had higher values for midsagittal concavity (and thus presumably more grooving) when the artificial palate was first inserted (i.e. in the WN condition).

When auditory feedback became available, values of all speakers were at least as high as in the unperturbed session. In the last session all speakers except AM1 had higher values than in the unperturbed session. The influence of the condition on midsagittal concavity is significant for all speakers. Most pairwise comparisons between all the different levels of the factor condition are significant as well (cf. table 3). An increase in midsagittal concavity, which we interpret as evidence for more tongue grooving, should induce an increase of energy in the 20-24 Bark frequency band. Hence, for all speakers except one (AM1) this result is consistent with the acoustic adaptation hypothesis (1) and the auditory feedback hypothesis (3).

The investigation of trends over sessions showed that again only speaker AM1 had such trends. He had a more and more convex tongue during the FF session, possibly because he was adapting via jaw raising.

INSERT FIGURE 6 ABOUT HERE

Compensation between jaw height and midsagittal concavity. If one compares midsagittal concavity and jaw height one can see that over sessions the values of the two parameters tend to move in opposite directions. However, this general trend is associated with speaker dependent strategies. For speaker CF1, the jaw lowered progressively over sessions, while she had a more and more concave tongue shape (and hypothetically more grooving). A similar tendency can be found for speaker CF2. She lowered the jaw in the first perturbed session and changed this position only slightly afterwards. Also, she adjusted the tongue towards a less convex shape in the early perturbed session and varied this shape only slightly afterwards. The opposite behaviour was found for speaker AM1: in the final session this speaker had a very convex tongue shape but a very high jaw position. Acoustically, such a covariance would make sense: both grooving (hypothetically linked to a concave tongue shape) and a high jaw lead to high frequency energy, so it might be sufficient to have only one of the two.

Following these observations Pearson correlations between midsagittal concavity and jaw height were calculated for each speaker for all sessions combined. Table 4 shows the results. For four speakers a negative correlation was found, three of them were significant. For two speakers, AF1 and CF3, non-significant correlations were found. Half the speakers thus preferred a lower jaw position if they had higher values for midsagittal concavity (hypothetically linked to much grooving) and a higher jaw if they had lower values for midsagittal concavity.

Speakers AF1 and CF3 did not change the jaw position much, but only developed more and more grooving as measured via midsagittal concavity. This explains the absence of a significant correlation, but confirms the importance of the midsagittal concavity (and hypothetically grooving) in the adaptation strategy.

In general, the results for jaw position and midsagittal concavity both separately and in coordination with each other are compatible with the observed energy increase in the high frequency band.

INSERT TABLE 4 ABOUT HERE

Discussion

This study has dealt with a sound which is acquired late in speech and is often impaired, i.e. the fricative /s/. Impairment usually shows up as a lack of high frequency energy. Logopedic treatment of this sound usually consists of learning the correct place of articulation and learning to produce a groove of the tongue. The need of a high jaw position seems to be underestimated for the creation of high frequency noise.

The main aim of the present study was to investigate the role of these three articulatory parameters, constriction position, grooving and jaw height, in the production of /s/ in perturbed speech. Previous studies have shown that a palatal perturbation as it is used in the present study led to a lowering of the spectral centre of gravity. This could either be due to a backward displacement of the location of the constriction, or it could be due to a lowering of the jaw or to a change in tongue grooving at perturbation onset.

Acoustic measurements of our data support previous findings: There was less high frequency energy (above 5 kHz) at perturbation onset, but in the course of the adaptation speakers tended to develop more and more high frequency energy. This first finding is consistent with our initial hypothesis (1), that adaptation strategies aim at preserving the acoustic properties that are typical for /s/.

The articulatory results show that the horizontal tongue position cannot be the articulatory parameter which is primarily responsible for the differences in high frequency energy, since it varied inconsistently over the experiment. Adaptation of jaw position and midsagittal concavity, both of which have been shown to be involved in producing high frequency energy, seem to be more important. All the speakers used at least one of these parameters in order to adapt. Speakers AM1 and AM2, who wore an alveolar prosthesis seemed to prefer a high jaw position, and all the speakers with a central prosthesis used a higher midsagittal concavity as a means of adaptation. For three speakers a negative covariation across conditions of the two parameters was found.

The first aim of the study was to determine which articulatory parameter is most important in the adaptation process. The results show that the main adaptation parameters were jaw position and midsagittal concavity (hypothetically linked to grooving). Interestingly, our subjects tended to adapt the respective weights of the corrections to each parameter according to the shape of the palatal prosthesis (see below). They exploited the degree of freedom allowed by the trade-off between these two parameters.

The second aim of the study was to investigate whether adaptive behaviour would differ depending on the shape of the prosthesis, especially as long as no feedback is available. Our hypothesis was that with only somatosensory feedback available speakers with an alveolar prosthesis would adapt differently as compared to speakers with a central

prosthesis. This, however, was not observed. Instead, differences were found when auditory feedback was available. Speakers with an alveolar prosthesis then seem to have favoured a high jaw position as adaptation whereas speakers with a central prosthesis rather opted for a change in midsagittal concavity. This could be explained if one assumes that for these latter subjects high jaw positions might have rapidly induced contacts between the tongue and the prosthesis. For them the trade-off between jaw height and grooving may have represented the only viable route towards stable compensation.

The third aim of the study was to investigate the role of auditory feedback. Looking only at the differences between the spectral characteristics of the /s/ produced without vs. with auditory feedback (i.e the differences between the WN and FF conditions on the first day), it might be concluded that this feedback has little or no immediate influence. Indeed, for the majority of subjects no significant spectral change was observed between the two conditions. However, another interpretation could be that subjects needed time, even with auditory feedback, to find an appropriate compensation strategy that has an impact on the spectral characteristics. Evidence supporting this interpretation can be found in the variation of the important articulatory parameters across conditions. Indeed, a dependence of adaptive behaviour on the feedback available was found in the jaw height data. When auditory feedback became available, speakers moved back towards the jaw position they used to have in the unperturbed production. Similar observations were made for midsagittal concavity: for four speakers (AM2, AF1, CF1, CF3) a significant increase in this value was measured from the WN to the FF condition. When the difference is not significant (CF2) or when there was a decrease in the midsagittal concavity value (AM1), significant jaw elevation was observed. These observations are all consistent with the idea that auditory feedback was helpful to find an appropriate strategy in order to recover the original spectral

properties of /s/. However, more practice time was needed to get a noticeable effect in the acoustical domain. A significant improvement was often reached only after two weeks.

The results presented in this study have implications for the treatment of /s/-related impairments. Explanations of the articulation of /s/ should include all the three parameters, tongue position, tongue grooving and jaw position because all three of them are important for the production of /s/. Furthermore, new visualization devices should present data on the jaw height, in addition to data on tongue positioning and grooving. It would also be desirable to have not only information on the groove width and length (as presented by EPG) but also information on the groove depth. The therapist may also be able to help the patient to exploit the trade-off between grooving and a high jaw position.

Acknowledgements

Thanks to Olessia Panzyga, Susanne Waltl and Vivien Hein for acoustic segmentation, to Mark Tiede for a script for calculating average spectra, and to our subjects.

Declaration of interest

This work was funded by the German Research Council (PO 334/4-1 and HO 327/1-1), the Ministère délégué à l'Enseignement Supérieur et à la Recherche of France and the POPAART P2R programme from the CNRS, the French Foreign Office and the German Research Council.

References

Aasland, W., Baum, S. & McFarland, D. (2006). Electropalatographic, acoustic and perceptual data on adaptation to a palatal perturbation. *Journal of the Acoustical Society of America*, 119, 2372-2380.

Amerman, J., Daniloff, R. & Moll, K. (1970). Lip and jaw coarticulation for the phoneme /æ/. *Journal of Speech and Hearing Research,* 163, 147-161.

Baum, S. & McFarland, D. (1997). The development of speech adaptation to an artificial palate. *Journal of the Acoustical Society of America*, 102, 2353-2359.

Boersma, P. & Weenink, D. (1999). PRAAT 3.8, a system for doing phonetics by computer, www.praat.org.

Fletcher, S.G. & Newman, D.G. (1991). [s] and [ʃ] as a function of linguapalatal contact place and sibilant groove width. *Journal of the Acoustical Society of America,* 89(2), 850-858.

Gibbon, F. & Hardcastle, W. (1987). Articulatory description and treatment of 'lateral /s/' using electropalatography: a case study. *British Journal of Disorders of Communication*, 22, 203-217.

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

Howe, M. & McGowan, R. (2005). Aeroacoustics of [s]. *Proc. R. Soc. London*, Ser. A 461: 1005-1028.

Ladefoged, P. & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Lee, S., Beckman, M. & Jackson, M. (1994). Jaw targets for strident fricatives. *Proceedings of ICSLP*, Yokohama, 37-40.

Lindblom, B. & Lubker, J. (1985). The speech homunculus and a problem of phonetic linguistics. In: V.A. Fromkin (ed.): *Phonetic Linguistics*. London: Academic Press.

McFarland, D., Baum, S. & Chabot, C. (1996). Speech compensation to structural modifications of the oral cavity. *Journal of the Acoustical Society of America,* 100, 1093-1104.

Mooshammer, C., Hoole, P. & Geumann, A. (2006). Inter-articulator cohesion within coronal consonant production. *Journal of the Acoustical Society of America,* 120(2), 1028-1039.

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

Narayanan, S.S., Alwan, A.A. & Haker, K. (1995). An articulatory study of fricative consonants using magnetic resonance imaging. *Journal of the Acoustical Society of America*, 98(3), 1325-1347.

Nasir, S.M. & Ostry, D.J. (2008). Speech motor learning in profoundly deaf adults. *Nature Neuroscience*, 11(10): 1217-1222.

Robb, M.P. & Bleile, K.M. (1994). Consonant inventories of young children from 8 to 25 months. *Clinical Linguistics and Phonetics*, 8, 295-320.

Römer, A., Willmes, K. & Kröger, B.J. (2009). Erprobung des audiovisuellen Feedbackprogramms CoKo (computergestützter Sprechkorrektor) in der Sigmatismustherapie. *Sprache, Stimme, Gehör*, 33(4), 186-192.

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

Shadle, C.H., Berezina, M., Proctor, M. & Iskarous, K. (2008). Mechanical models of fricatives based on MRI-derived vocal tract shapes. In R. Sock, S. Fuchs & Y. Laprie (Eds.), *Proceedings of 8th International Seminar on Speech Production* (pp. 413-416).

Shadle, C. (1985). *The Acoustics of Fricative Consonants*. Technical Report 506, Massachusetts Institute of Technology, Research Laboratory of Electronics, Cambridge, Massachusetts.

Shadle, C. (1991). The effect of geometry on source mechanisms of fricative consonants. *Journal of Phonetics*, 19, 409-424.

Shiller, D.M., Sato, M., Gracco, V.L. & Baum, S.R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *Journal of the Acoustical Society of America*, 125(2): 1103-1113.

Schroeder, M. R., Atal, B. S., & Hall, J. L. (1979). Objective measure of certain speech signal degradations based on masking properties of human auditory perception. In B. Lindblom & S. E. G. Öhman (Eds.), *Frontiers of Speech Communication Research* (pp. 217–229). London: Academic Press.

Stevens, K.N. (1971). Airflow and turbulence noise for fricative and stop consonants: static considerations. *Journal of the Acoustical Society of America*, 50, 1180-1192.

Stoel-Gammon, C. (1985). Phonetic inventories, 15-24 months: A longitudinal study. *Journal of Speech and Hearing Research*, 28, 505-512.

Stone, M., Epstein, M., Li, M. & Kambhamettu, C. (2006). Predicting 3D tongue shapes from midsagittal contours. In J. Harrington, & M. Tabain (Eds.), *Speech Production: Models, Phonetic Processes, and Techniques* (pp. 315-330). New York: Psychology Press.

The R Foundation for Statistical Computing (2009). R version 2.9.0, www.r-project.org.

Van Riper, C. & Irwin, J.V.l (1958). *Voice and Articulation*. London: Pitman Medical Publishing Company.

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

Wilcox, K.A., Stephens, M.I. & Daniloff, R.G. (1985). Mandibular position during children's defective /s/ productions. *Journal of Communication Disorders*, 18, 273-283.

Appendix : Calculation of DCT-coefficients

This method for describing spectral contours follows Watson and Harrington (1999) and Guzik & Harrington (2007). A discrete cosine transform expresses a function (here a spectral envelope as a function of the frequency) in terms of a sum of cosine functions with different frequencies. The coefficients of a DCT thus give information about the weight of a cosine function with a certain frequency in this sum. The different coefficients of the spectra give information about global characteristics such as the mean amplitude of the spectrum or the tilt of the spectrum.

As described in detail in Watson and Harrington (1999), the first coefficient of the DCT describes the mean of the input vector, in our case the mean amplitude of the bark transformed spectrum. The second coefficient (corresponding to a half-cycle cosine over the width of the spectrum) is positive if the spectrum has more energy at low frequencies (as a half cycle cosine) and it is negative if it has more energy at higher frequencies. Thus, for our data a spectrum with a lower-frequency main spectral peak could be expected to have higher values of the second coefficient than a spectrum with a higher-frequency main peak. This coefficient describes broad characteristics of the complete spectrum, but not the rather subtle differences in the region above 6 kHz which were found between perturbed and unperturbed spectra. The third coefficient, corresponding to a whole-cycle cosine over the width of the spectrum describes the curvature of the spectrum. This coefficient is positive if the spectrum has a trough, negative if there is a peak and zero if the spectrum is flat.

For the present purpose we decided to use the fourth coefficient, even if this coefficient was not used by either Watson and Harrington or Guzik and Harrington. This coefficient describes the spectral envelope as a 1.5-cycle cosine (cf. figure A1, lower subplot). It thus allows us to analyse the energy distribution in six different portions of the

spectrum (for a 1.5 cosine the portions 0-0.5π- π-1.5π-2π-2.5π-3π). The fourth coefficient is strongly positive if there is much energy in the portions of the spectrum where a cosine function has positive values (0-0.5π, 1.5π-2.5π) and little energy in the other portions, where the cosine has negative values. For spectra plotted on a bark scale which covers a 0 to 24 Bark interval (figure A1, upper subplot) this means that the coefficient is high if there is much energy in the 0-4 Bark interval or in the 12 to 20 Bark interval. In the example plotted in figure A1, the unperturbed and perturbed spectra differed predominantly in the amount of energy below and beyond 5kHz (~20 Bark): the unperturbed spectra had more energy in the higher band (20-24 Bark), the early perturbed spectra had more energy in the lower band (16-20 Bark). This difference is described by the fourth coefficient: it is low if there is more energy in the 20 to 24 Bark band than in the lower frequencies and it is high if there is less energy in this frequency band.

INSERT FIGURE 1A ABOUT HERE

The DCT is applied to a bark-transformed spectrum. The bark transformation was carried out following Schroeder et al. (1979):

$$F_{Bark}=7 \cdot \sinh^{-1}(F_{Hz}/650)$$

The DCT-coefficients were calculated in Matlab.

Table 1: Results of ANOVA with Tukey Test for multiple comparisons for influence of *condition* on the acoustic parameters. Column 1: speaker, column 2: ANOVA-results, remaining columns: p values of post-hoc tests if the contrast was significant (UP: unperturbed, WN: white noise, FF: full feedback, AD: adapted). ↑: the mean of the second session in a pair is lower than the mean of the first. ↓: opposite.

coeff 4							
	ANOVA	Tukey-test					
		UP-WN	UP-FF	UP-AD	WN-FF	WN-AD	FF-AD
AM1	F(3, 75)=11.146, p<0.001	0.005↑	0.001↑	n.s.	n.s.	0.001↓	<0.001↓
AM2	F(3, 77)=12.822, p<0.001	<0.001↑	0.002↑	0.039↑	n.s.	0.006↑	n.s.
AF1	F(3, 76)=41.389, p<0.001	<0.001↑	<0.001 ↑	<0.001↑	n.s.	<0.001↓	<0.001↓
CF1	F(3, 76)=17.262, p<0.001	<0.001↑	n.s.	n.s.	0.003↓	<0.001↓	0.030↓
CF2	F(3, 76)=2.728, p=0.050	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
CF3	F(3, 77)=12.565, p<0.001	<0.001↑	n.s.	n.s.	0.040↓	<0.001↓	<0.001↓

Table 2: As table 1 for the parameter horizontal tongue position

horizontal tongue position							
AM1	F(3, 76)=12.554, p<0.001	n.s.	n.s.	<0.001↑	n.s.	<0.001↑	<0.001↑
AM2	F(3, 76)=75.92, p<0.001	<0.001↑	<0.001↑	<0.001↑	<0.001↑	n.s.	0.010↓
AF1	F(3, 76)=53.101, p<0.001	<0.001↑	<0.001↑	n.s.	<0.001↓	<0.001↓	<0.001↓
CF1	F(3, 76)=235.72, p<0.001	<0.001↑	<0.001↑	<0.001↑	<0.001↑	<0.001↑	<0.001↓
CF2	F(3, 76)=27.351, p<0.001	n.s.	n.s.	<0.001↑	n.s.	<0.001↑	<0.001↑
CF3	F(3, 76)=16.873, p<0.001	n.s.	<0.001↑	n.s.	<0.001↑	n.s.	<0.001↑

Table 3: As table 1 for the parameters jaw height and midsagittal concavity.

	ANOVA	Tukey-test					
		UP-WN	UP-FF	UP-AD	WN-FF	WN-AD	FF-AD
jaw height							
AM1	F(3, 76)=10.707, p<0.001	0.013↓	n.s.	n.s.	0.001↑	<0.001↑	n.s.
AM2	F(3, 76)=27.908, p<0.001	<0.001↓	n.s.	n.s.	<0.001↑	<0.001↑	0.002↑
AF1	F(3, 76)=13.482, p<0.001	n.s.	n.s.	0.004↑	<0.001↓	n.s.	<0.001↑
CF1	F(3, 76)=113.18, p<0.001	<0.001↓	n.s.	<0.001↓	n.s.	<0.001↓	<0.001↓
CF2	F(3, 76)=136.63, p<0.001	<0.001↓	<0.001↓	<0.001↓	0.007↑	<0.001↓	<0.001↓
CF3	F(3, 76)=2.119, p=.105	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
midsagittal concavity							
AM1	F(3, 76)=209.49, p<0.001	<0.001↑	<0.001↑	<0.001↓	<0.001↓	<0.001↓	<0.001↓
AM2	F(3, 76)=173.24, p<0.001	<0.001↑	<0.001↑	<0.001↑	0.001↑	0.001↓	<0.001↑
AF1	F(3, 76)=59.329, p<0.001	0.002↓	n.s.	<0.001↑	0.001↑	<0.001↑	<0.001↑
CF1	F(3, 76)=355.64, p<0.001	<0.001↑	<0.001↑	<0.001↑	<0.001↑	<0.001↑	<0.001↑
CF2	F(3, 76)=193.75, p<0.001	<0.001↑	<0.001↑	<0.001↑	n.s.	n.s.	n.s.
CF3	F(3, 76)=37.033, p<0.001	<0.001↑	<0.001↑	<0.001↑	<0.001↑	<0.001↑	n.s.

Table 4: Correlation coefficients and significance values (in parenthesis) for the correlation between jaw height and midsagittal concavity across sessions.

Speaker	**R (p)**
AM1	-0.4032 (<0.001)
AM2	-0.1760 (0.118)
AF1	0.1609 (0.154)
CF1	-0.6065 (<0.001)
CF2	-0.7419 (<0.001)
CF3	0.411 (0.151)

Figure captions

Figure 1: Left: example for an alveolar palate from a sagittal perspective. Right: example for a central palate from a sagittal perspective. Solid line: natural palatal contour, dashed line: artificial palatal contour.

Figure 2: Bark transformed mean spectra of the unperturbed session (solid lines) and white noise perturbed session (dotted lines) of speaker AF1. Thick lines show means over sessions, thin lines show standard deviations.

Figure 3: Mean fourth coefficient of the DCT. Each subplot refers to one speaker. Each boxplot within a subplot shows results for one session as given in the tick labels of the abscissa (UP: unperturbed, WN: white noise, FF: full feedback, AD: adapted). Boxplots show lower quartile, median, upper quartile, whiskers end at 1.5 quartiles. Higher values correspond to less energy in the 20-24 Bark frequency band. Alveolar prosthesis subjects are on the left hand side. Crosses: outliers.

Figure 4: As figure 2, but for horizontal tongue position. Higher values indicate a more retracted tongue position.

Figure 5: As figure 2, but for vertical jaw position. Higher values indicate a higher jaw position.

Figure 6: As figure 2, but for midsagittal concavity as expressed by coefficient *a*. Higher values mean the midsagittal tongue profile is more concave, which indicates more upstream grooving.

Figure A1: First subplot: mean unperturbed (solid line) and perturbed (dotted line, WN-condition) bark transformed spectra of speaker AF1. Second subplot: 1.5 cosine cycle.

Figure 1:

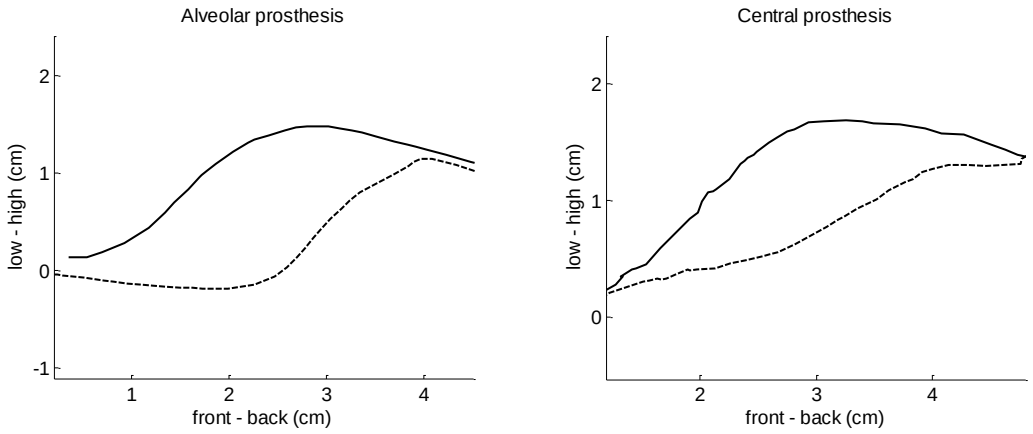


Figure 2

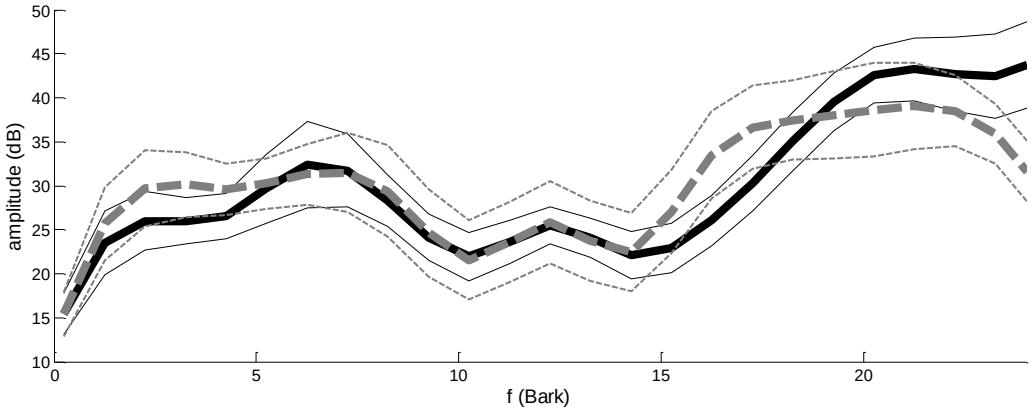


Figure 3

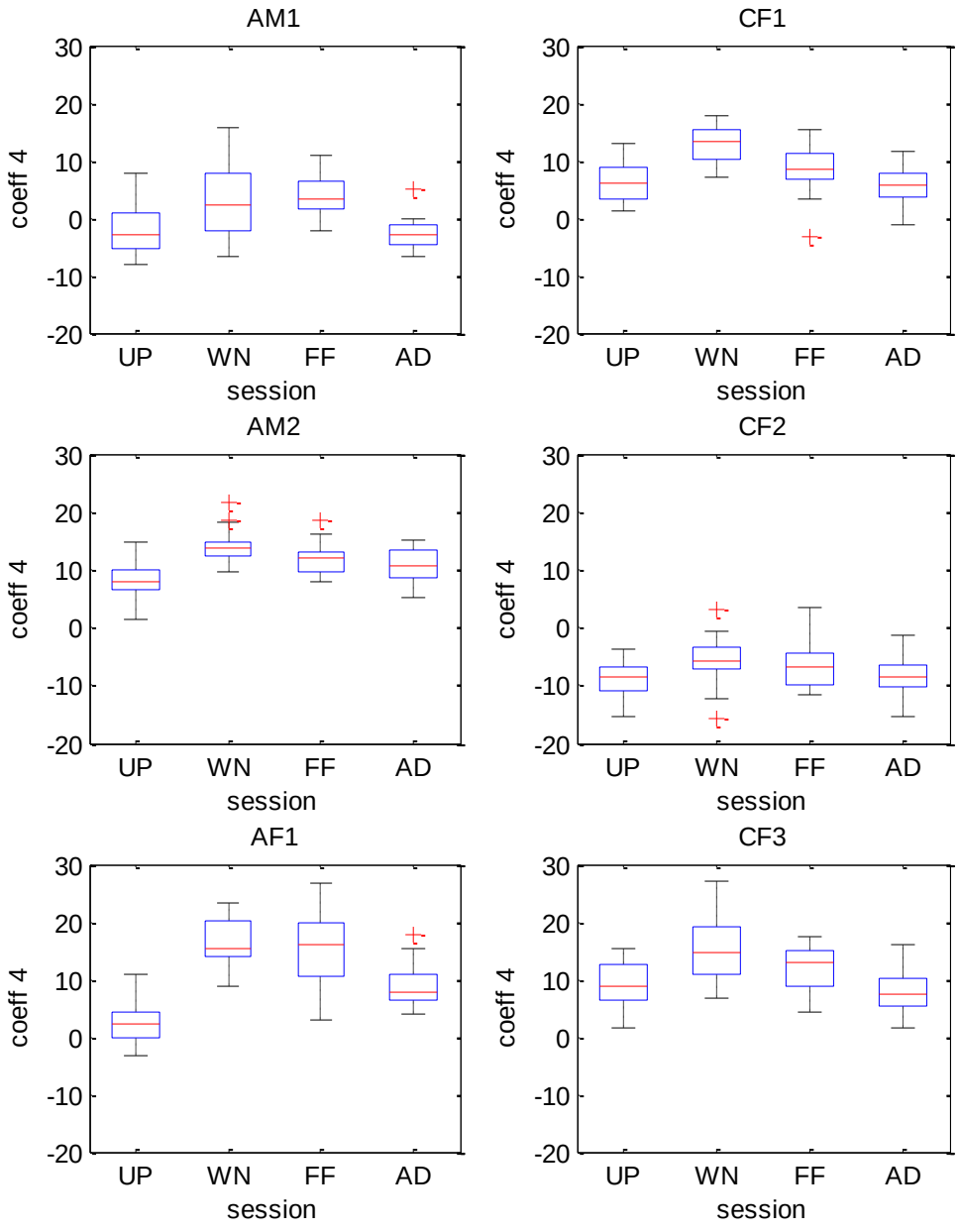

AM1
CF1
AM2
CF2
AF1
CF3
coeff 4
session
UP
WN
FF
AD
30
20
10
0
-10
-20

Figure 4

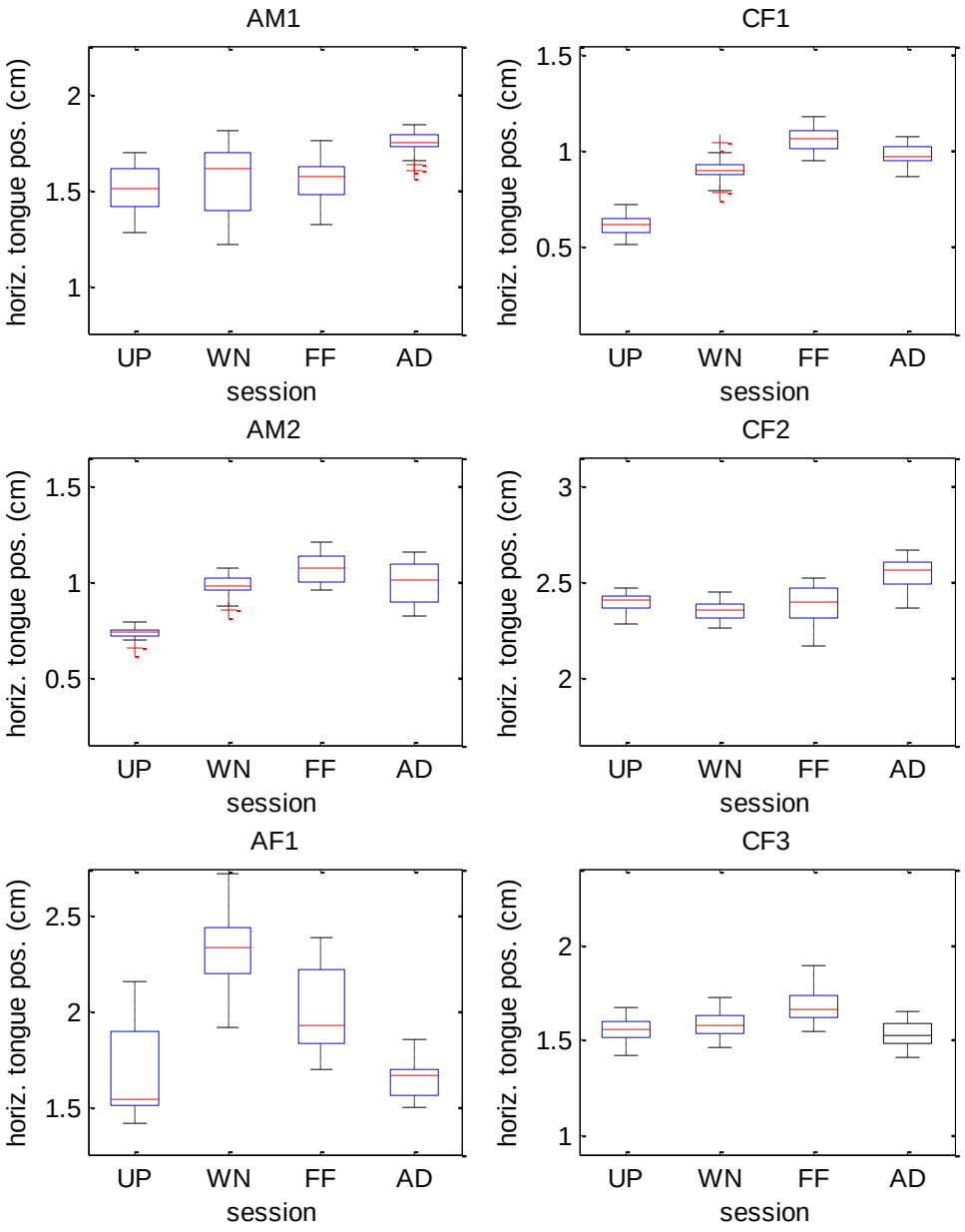


Figure 5

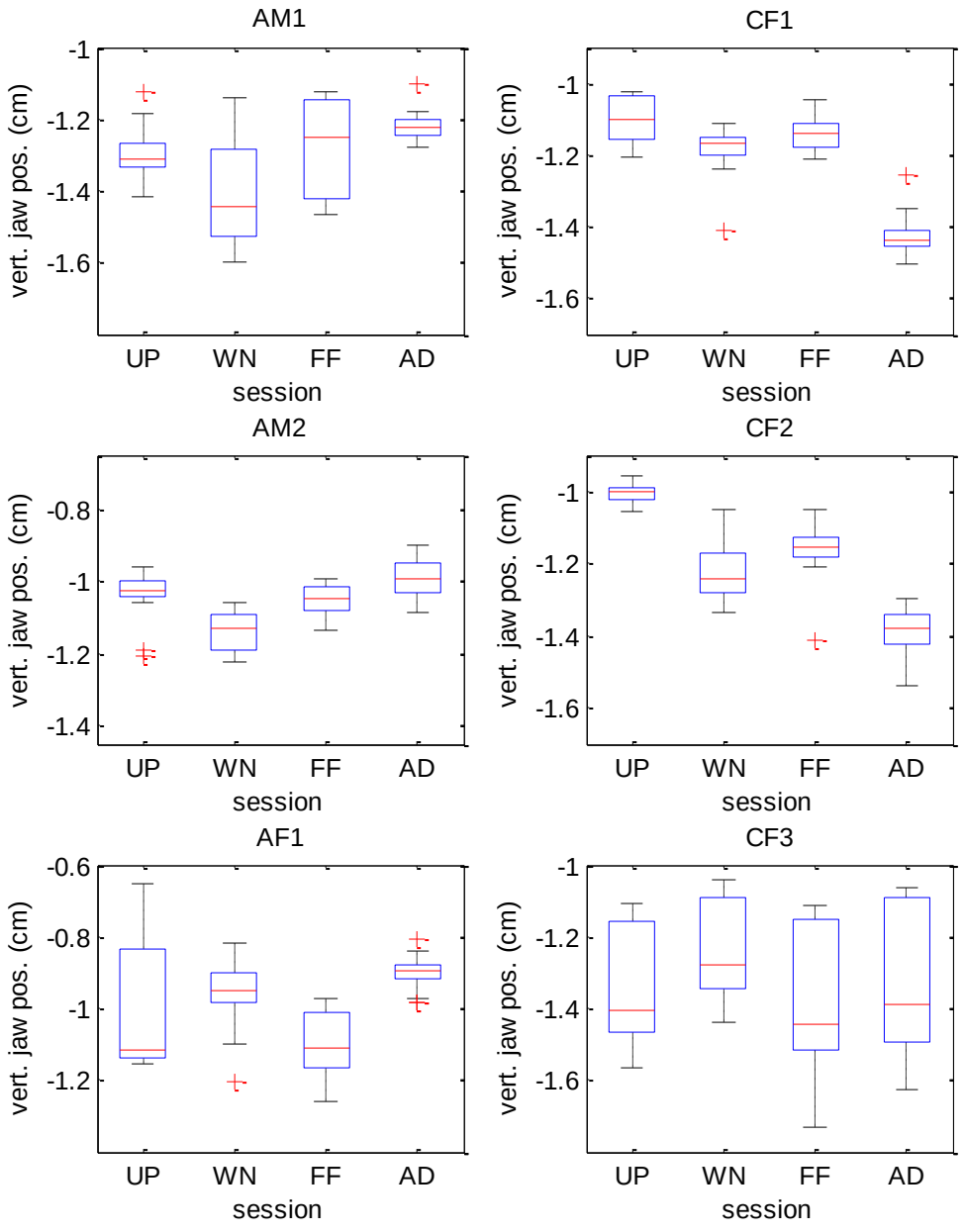


Figure 6

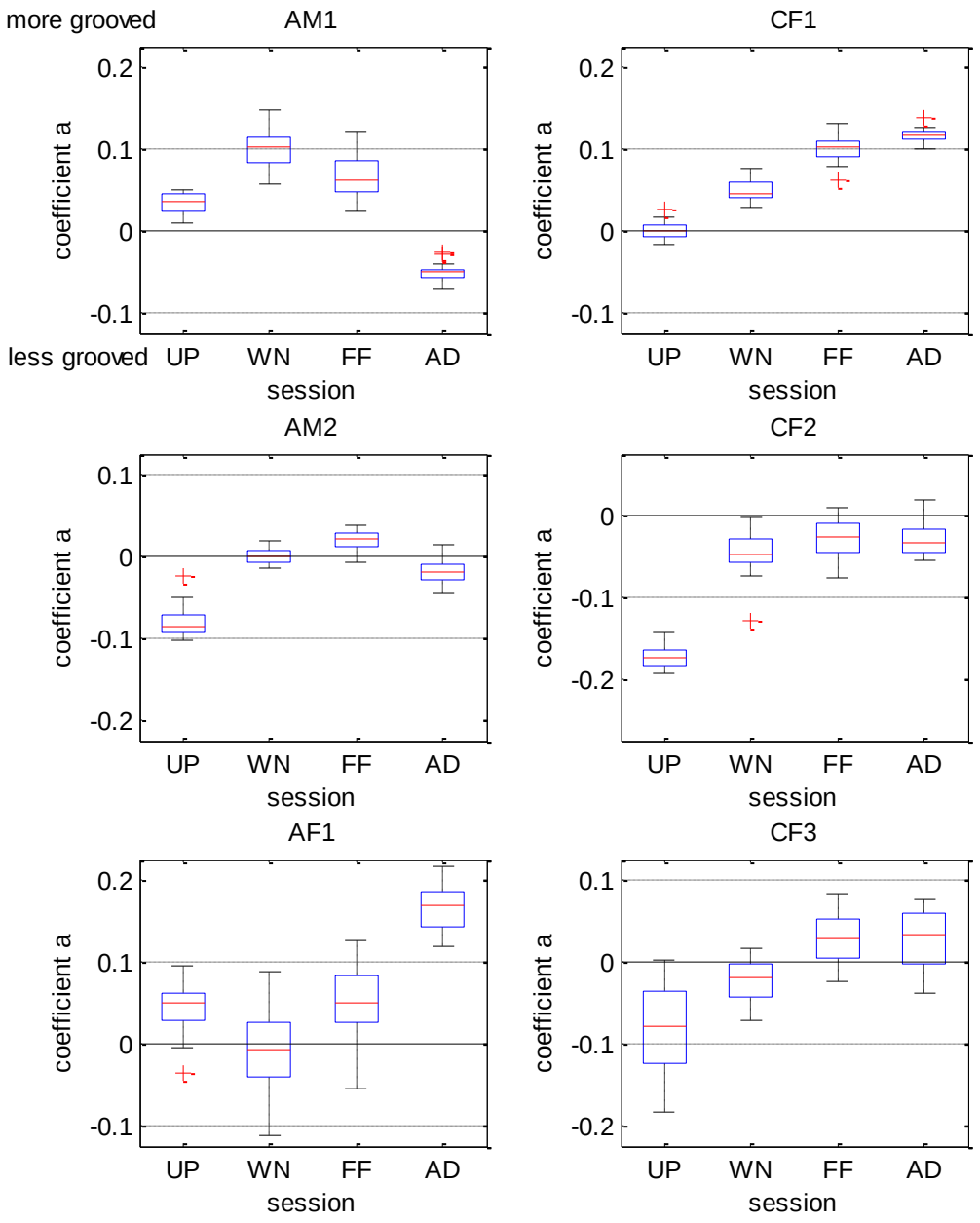
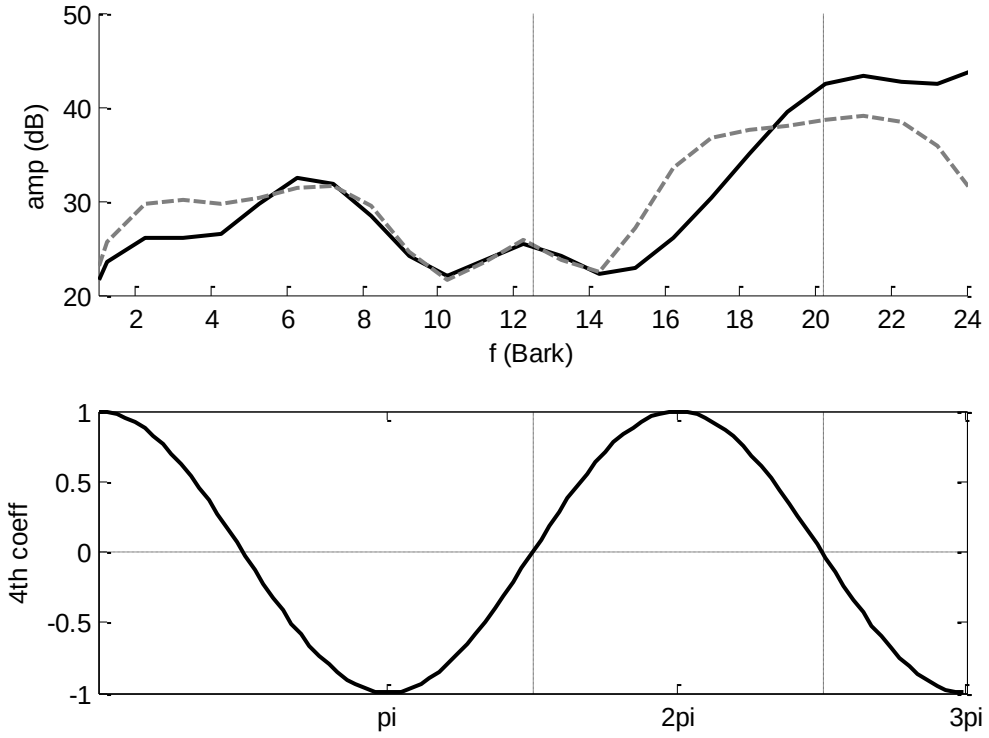

more grooved
less grooved
AM1
CF1
AM2
CF2
AF1
CF3
coefficient a
session
UP
WN
FF
AD

Figure A1


50
40
30
20
amp (dB)
2
4
6
8
10
12
14
16
18
20
22
24
f (Bark)
1
0.5
0
-0.5
-1
4th coeff
pi
2pi
3pi