Vowel coarticulation and undershoot in prosodically weakened positions

Conceição Cunha, Jonathan Harrington, Ulrich Reubold[1]

[1]*Institute of Phonetics and Speech processing, LMU Munich, Germany*

cunha|jmh|reubold@phonetik.uni-muenchen.de

Abstract

This paper focuses on the influence of vocalic context and prosodic weakening on the production and perception of the German fricatives /s-ʃ/. Many studies have shown coarticulatory influences of vowel quality on the preceding fricative and perceptual compensation for such coarticulatory relationships (Nittrouer & Studdert-Kennedy, 1987; Mann & Repp, 1980). However, very little is known about the perceptual parsing of coarticulated speech in prosodically weakened positions. The working hypothesis is that the degree of coarticulation may be similar in accented and deaccented positions, but that the variability associated with undershoot is greater in deaccented words, causing listener errors in attributing coarticulation to its source.

In order to explore this issue, we analyse the coarticulatory influences of German /ɪ, ʊ/ on /s, ʃ/. The present investigation involves physiological analysis of tongue position and movement. It addresses how this is related to perceptual judgments, and whether this relationship changes between accented and deaccented words.

Keywords: coarticulation, prosodic weakening, undershoot, German, prosody, perception

1. Introduction

The main aim of this study was to determine the extent to which vowel coarticulation and prosodic weakening affects the production and perception of /s/ and /ʃ/ in German. More specifically, the study investigates the articulatory implementation of coarticulation for /ɪ, ʊ/ on /s, ʃ/ in accented and deaccented positions and its parsing in perceptual judgements, with a special focus on perceptual compensation for prosodically weakened tokens.

The coarticulatory influence of vowel context on the preceding fricative has been extensively investigated in adult and child speech over the past decades (Recasens & Espinosa, 2007; Katz et al., 1991; Mann & Soli, 1991; Nittrouer et al., 1989; Nittrouer & Studdert-Kennedy, 1987; Repp, 1986; Mann & Repp, 1980). The raised tongue position in high vowels is assumed to facilitate the assimilation of the front-back location of the fricative's constriction, resulting in a more fronted constriction for fricatives in /i/ than in /u/ contexts (Soli, 1981). This anticipatory lingual coarticulation is acoustically detectable by a shift of F2 *loci*. Lip rounding for /u/ has also been associated with global shifts in the spectrum during the fricative (Nittrouer et al., 1989; Repp, 1986; Soli, 1981). The realization of lip rounding in post-alveolar fricatives – considered to be an enhancement strategy (Stevens and Keyser, 2010) – has been shown to be subject to great inter-speaker variability, since some speakers realise /s/ and /ʃ/ with the same lip configuration and absolutely no rounding even in /u/ context (Proctor et al. 2006, Fig. 6).

Many studies have shown perceptual compensation for anticipatory coarticulation. For example, the participants in Mann & Repp (1980) reported more /s/-responses before /u/ than in the unrounded context of /a/, meaning that listeners factored out the anticipatory spectrum-lowering of the lip rounding for /u/ on the preceding fricative (compensation).

The second aim of this study was to assess the influence of prosodic weakening on the degree of coarticulation of initial fricative vowel sequences and its influence on speech perception. More coarticulation has been attested in unaccented than in accented positions (Cho, 2004; Lindblom et al., 2009). In an acoustic and perceptual analysis of VCV-coarticulation in German, Harrington et al. (2013) suggested that the magnitude of coarticulation may be similar in both prosodic positions, but that the degree of variability resulting from target undershoot may be much greater in deaccented position. However, very little is known about the parsing of coarticulation under prosodic weakening in speech perception.

Listeners may have more difficulty parsing coarticulatory relationships in prosodically weak constituents (Harrington et al. 2013), not due to greater coarticulation but rather due to increased variability, which might mask coarticulation causing listener errors in attributing the coarticulation to the source that gives rise to it.

In order to explore this issue, we analyse the influence of German /ɪ, ʊ/ on /s, ʃ/ fricatives, building on earlier studies investigating similar materials (Fowler, 2006; Mann & Repp, 1980; Mann & Soli, 1991; Nittrouer & Studdert-Kennedy, 1987; Whalen, 1981). Assuming lip rounding in the production of the post-alveolar fricative (at least in some speakers) and no lip rounding in the alveolar counterpart, we predict the greatest amount of lip rounding and/or backmost tongue constriction for /ʃ/ in the /u/ context, since both segments are produced with rounding of the lips. Conversely, no lip rounding and the most fronted tongue position are expected in the production of the alveolar fricative with the front vowel (/sɪ/). In perception, we expect to replicate Mann & Repp's (1980) results: the greater the vowel's influence, the more /s/-responses in an /ʊ/-context and conversely more /ʃ/ in the context of /ɪ/. Following Harrington et al. (2013) we expect a similar degree of coarticulation in both prosodic positions, but undershoot and increased variability and less perceptual normalisation for context in deaccented positions.

2. Production Experiment

2.1. Methods

2.1.1. Experimental set-up

Physiological EMA data were recorded from eight speakers of southern German (four male, four female) using the 3D articulograph CARSTENS AG501. Two sensors were placed on the tongue: one on the midline 1 cm behind the tongue tip (TT) and the other on a level with the molar teeth at the tongue back (TB). Two sensors were placed on the upper and lower lip (the latter henceforth LL). Four additional sensors were fixed to the maxilla, the nose bridge, as well as to the left and right mastoid bones: these served as reference sensors to correct for head movement.

The speech material consisted of initial fricative-vowel sequences followed by /ɪ, ʊ/ in the 4 German lexical words Suppen ‘soups’, Schuppen ‘dandruff, hovel’, Sippen ‘clans’, Schippen ‘scoops’, supplemented with 14 distractor words. The target words were embedded in phrase-final position in the carrier sentence *Maria mag [target word]* (eng, ‘Maria likes [target word]’). Two of the target words contain voiced fricatives which usually become devoiced in Southern German when following a voiceless/devoiced context (as in [maʁiːamaːksʊpn̩]). In order to elicit either accented or deaccented position by shifting the focus between the initial and the target word in the carrier phrase, the participants were presented with questions designed to elicit a narrow focus on the target word for the accented context and a broad focus for the deaccented context: either *WAS mag Maria?* (‘WHAT does Maria like?’) or *WER mag [target word]?* (‘WHO likes [target word]?’). Thereafter, the stimulus was presented with the word carrying the nuclear accent in capital letters (e.g. *MARIA mag Schuppen* vs. *Maria mag SCHUPPEN*). If subjects made a mistake, they were instructed to repeat the sentence. In total each speaker produced 80 utterances containing one of the target words (2 accentuation conditions x 4 target words (=2 fricatives x 2 vowels) x 10 repetitions).

2.1.2. Data analysis

The acoustic data were digitized at 16 kHz and automatically segmented and labeled using the Munich Automatic Segmentation tool (MAuS, Schiel et al., 2011). The segment boundaries of the target words’ fricatives and the following vowels were manually corrected.

Post-processing of the physiological raw data was done semi-automatically in Matlab, whereas labeling and subsequent analyses of the physiological data were conducted using EMU and EMU/R (Harrington, 2010). The physiological annotation of the three sibilants was based on the vertical movement of the TT (in mm) and the TT tangential velocity (in mm/s).

Our articulatory analyses were all based on the same time-frame which was derived from the gesture trajectories of the vertical movement of TT measured between the velocity peak of the fricative closing gesture (v_{on}) and the acoustical vowel onset. We extracted in this time-frame the horizontal movement of TT and the horizontal movement of LL.

In order to quantify the articulatory trajectories of the *horizontal* TT and LL movement, we used discrete cosine transform (DCT) to reduce the trajectories and the spectral slices to a set of coefficients. The *m*th DCT-coefficient C_m (m = 0, 1, 2) was calculated with the formula in (1):

$$C_m = \frac{2k_m}{N}\sum_{n=0}^{N-1} x(n)\cos\left(\frac{(2n+1)m\pi}{2N}\right) \qquad (1)$$

These three coefficients C_m (m = 0, 1, 2) encode the mean, the slope, and curvature respectively of the signal to which the DCT transformation was applied (Harrington, 2010).

We used a relative measure, the log. Euclidean distance ratio, to quantify relative positions of tokens in relation to anchors. These anchors varied as follows: we averaged the corresponding parameters over all *Schuppen* and *Sippen* tokens per speaker and accentuation condition, as we expected the distance between these fricatives to be maximally distributed. To calculate a measure for both undershoot and coarticulation differences, we used only those tokens from the accented condition as anchors; to factor out undershoot we applied the same methodology, but using for each accentuation condition separate anchors, i.e. accented *Schuppen* and *Sippen* token for the accented, unaccented *Schuppen* and *Sippen* tokens for the unaccented condition. By doing so, only the effects of accentuation differences on the amount of coarticulation remain.

The Euclidean distances E_{Sippen} and $E_{Schuppen}$ were calculated in a space build up by the three DCT coefficients separately for each fricative token. The log-Euclidean distance ratio *d* was then calculated for each fricative, from (2):

$$d = \log(E_{Schuppen}/E_{\mathrm{Sippen}}) = \log(E_{Schuppen}) - \log(E_{\mathrm{Sippen}}) \qquad (2)$$

The log-Euclidean distance ratio *d* was calculated in order to obtain one value per fricative which is a relative measure: negative values denote a closer distance to the *Schuppen* centroid, whereas positive values are associated with distances nearer to the *Sippen* centroid, while a value of zero denotes that a given fricative is equidistant between the centroids.

We applied repeated measures ANOVAS with *d* as dependent variable, and with the within-subject factors ACCENTUATION (accented vs. unaccented), WORD (sʊ, sɪ, ʃʊ, and ʃɪ).

2.2. Results

The effect of accentuation on the fundamental frequency of the target vowels was verified by conducting a repeated measures ANOVA with F0 as dependent variable and accentuation (accented vs. deacccented) and vowel (/ɪ/, /ʊ/) as within-speaker factors. The results showed a significant effect only for accentuation ($F(1,7)= 6.8$, $p< 0.05$).

2.2.1. Horizontal tongue tip

For each speaker the average of the horizontal movement of the TT sensor, measured between the velocity peak of the closing gesture of the fricative and the vowel offset was calculated. Fig.1 show a clear degree of separation of the trajectories for /s/ and /ʃ/ with the alveolar trajectories of /sɪ/ and /sʊ/ presenting a lower value for anteriority (i.e. a more fronted position) relative to the post-alveolar counterparts. However, the effect of vowel context is visible in the second half of the fricative in the trajectories of some speakers, in which the trajectory of /sʊ/ won on posteriority and the /ʃɪ/ trajectory instead became more anterior.

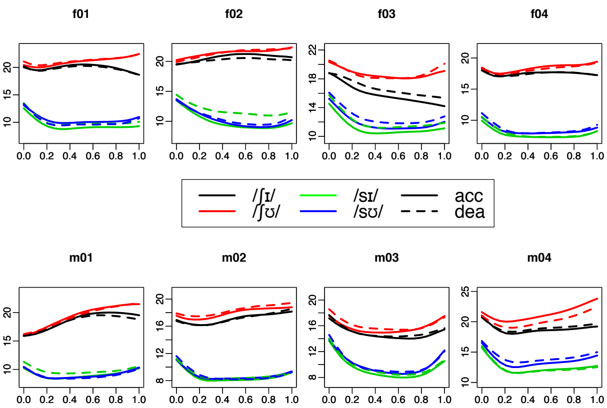

Figure 1: *Time normalized, averaged trajectories of the horizontal TT movement, measured between the velocity peak of the closing gesture of the fricative and the vowel onset.*

As evident in Fig. 2a, Log. Euclidean distance ratios are greater in /sɪ/ compared to /ʃɪ/, and in /ʃʊ/ compared to /sʊ/, confirming the impression given by the trajectories in Fig 1. The RM-ANOVA with Log. Euclidian distance ratios as dependent variable showed a significant effect of word

(F(3,21)= 30.4, p< 0.001) and a small effect of accentuation (F(1,7)= 6.0, p< 0.05), in which deaccented tokens are closer to zero than accented ones. Recall that this measure quantifies differences that come about mainly because of undershoot. Deaccented tokens show smaller distances than accented ones.

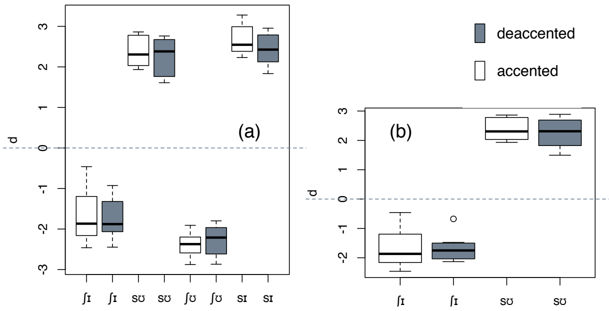

Figure 2: *TT: Log. Euclidean distance ratios calculated per subject for accented (white) and deaccented (grey) tokens of (a) /ʃɪ,sʊ,ʃʊ,sɪ/ in relation to anchors derived from accented /ʃʊ/ and /sɪ/, and of (b) /ʃɪ, sʊ/ in relation to accentuation-dependent anchors (/ʃʊ/$_{ACC}$ and /sɪ/$_{ACC}$ vs. /ʃʊ/$_{DEA}$ and /sɪ/$_{DEA}$).*

However, on the relative distances presented in Fig. 2b, we still find a main effect of word (F(1,7)= 122.5, p< 0.001), but no influence of accentuation. We interpreted this as being a consequence of the similar degree of coarticulation in both prosodic conditions (after the undershoot was factored out).

2.2.2. *Lower lip*

Fig. 3 displays the mean trajectories per speaker of the horizontal movement of the lower lip for the same time interval as in Fig 1.

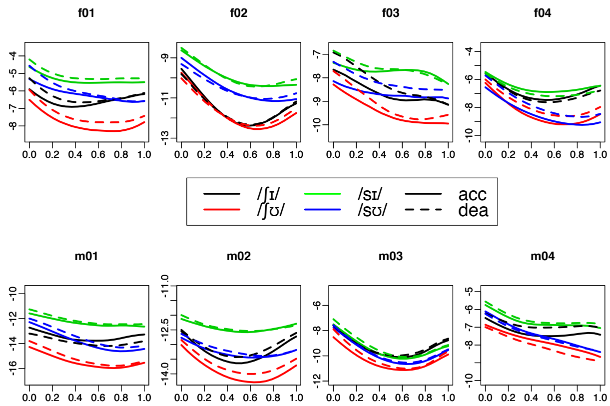

Figure 3: *Time normalized, averaged trajectories per speaker of the lower lip, measured between the velocity peak of the fricative closing gesture and the vowel onset.*

On the lower lip trajectories the degree of separation between fricatives is less evident. However, /ʃu/ showed as expected the more fronted position, which corresponds to the greatest amount of lip rounding. In /sɪ/ the lower lip showed the backmost position and /ʃɪ, sʊ/ displayed intermediate positions. The same trends could be confirmed in Fig. 4a, in which /sɪ/ displayed the greatest positive *d*-values and /ʃʊ/ displayed the most extreme negative values. Moreover, as expected, the mean distances of /sʊ/ and /ʃɪ/ present an intermediate value close to zero. This is the reason for the main effect of word on the *d*-values (F(3,21)= 29.8, p< 0.001).

Regarding accentuation, the trend for smaller distances in deaccented position visible in Fig. 4 was not statistically significant. The RM-ANOVA showed no effect of accentuation nor a significant interaction with word. For the relative distances shown in Fig. 4b we did not find any influence of word nor accentuation, meaning that the degree of coarticulation of the lower lip is also quite similar in both prosodic conditions.

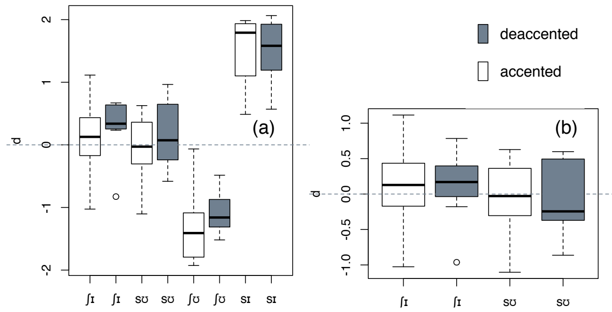

Figure 4: *LL: Log. Euclidean distance ratios calculated per subject for accented (white) and deaccented (grey) tokens of (a) /ʃɪ,sʊ,ʃʊ,sɪ/ in relation to anchors derived from accented /ʃʊ/ and /sɪ/, and of (b) /ʃɪ, sʊ/ in relation to accentuation-dependent anchors (/ʃʊ/$_{ACC}$ and /sɪ/$_{ACC}$ vs. /ʃʊ/$_{DEA}$ and /sɪ/$_{DEA}$).*

3. Perception Experiment

3.1. Method

For the perception experiment, we synthetized a 10-step continuum between /s/ and /ʃ/ by filtering white noise and prepended the resulting sounds to /ɪ, ʊ/ in the same words which we recorded in the production study. The target words were spoken in a carrier sentence (*Maria mag [target word]*) by a trained male phonetician with slight Southern German characteristics. The vowel transitions of the resulting stimuli were either appropriate to /s/ or /ʃ/ (see also Nittrouer & Studdert-Kennedy, 1987 for English). We manipulated the accentuation pattern by means of PSOLA in Praat by shifting the nuclear accent between *Maria* and the target word. Therefore, four different continua in two accentuation conditions could be tested:

- /s-ʃ/ continuum followed by /s/-transition and /ɪ/: **sɪ**.
- /s-ʃ/ continuum followed by /s/-transition and /ʊ/: **sʊ**.
- /s-ʃ/ continuum followed by /ʃ/-transition and /ɪ/: **ʃɪ**.
- /s-ʃ/ continuum followed by /ʃ/-transition and /ʊ/: **ʃʊ**.

These materials (10 step continuum x 2 vowels x 2 transitions x 2 prosodic conditions x 6 repetitions = 480 stimuli) were presented to 19 students of southern standard German aged between 21 and 29 years. Eight of them had taken part in the production experiment.

We ran a two alternative forced choice experiment in which participants were asked to choose between /s/ and /ʃ/ by clicking on buttons labelled with the corresponding German orthography for the sounds (either <S> or <Sch>).

For the statistics we ran repeated measures ANOVAS on the category boundary and slope with TRANSITION (sɪ, sʊ, ʃɪ, ʃʊ) and ACCENTUATION (accented vs, deaccented) as within-listener factors and, in case of an interaction, post hoc t-tests on each combination of vowel and accentuation.

3.2. Results

Fig. 5 shows mean psychometric functions fitted to the distributions of listeners' responses to the eight continua. The vertical lines correspond to the 50% cross over points.

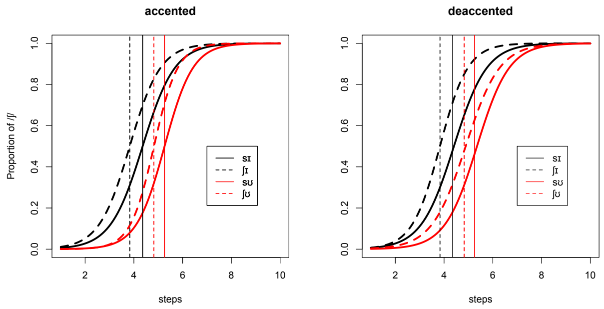

Figure 5: *Psychometric functions fitted to fricative (solid for /s/ and dashed for /ʃ/) and vowel context (/ɪ/ in black and /ʊ/ in red), shown separately for accented (left) and deaccented (right). The vertical lines correspond to the 50% cross over points.*

The psychometric curves presented in Fig. 6 and the corresponding 50% cross over boundaries showed a clear main effect of vowel on the category boundaries ($F(1,18)= 6.9$, $p< 0.001$): that is, listeners perceived more /s/ in the /ʊ/ context and more /ʃ/ in the /ɪ/ context, so that the vowels had the major influence on the category boundary. Accentuation had no effect on the cross over point nor on the slope.

4. Discussion and conclusion

This study analysed the coarticulatory influence of /ʊ, ɪ/ on /ʃ, s/ fricatives in two prosodic conditions (accented vs. deaccented). The analysis of the horizontal tongue tip (TT) and lower lip (LL) movement could confirm anticipatory lingual and labial coarticulation (Soli, 1981; Katz et al. 1991). As Fig. 3 suggests, there seems to be a considerable amount of inter-speaker variation in lip-rounding in the post-alveolar /ʃ/, irrespective of vowel context as in Proctor et al. (2006).

As can be seen in a comparison of Figs 2 and 4, the coarticulatory effect on the LL gesture was much more pronounced than the effect of anticipatory lingual coarticulation, which is reflected by the fact that the post-alveolar preceding unrounded /ɪ/ and the alveolar preceding rounding /ʊ/ share comparable amounts of lip rounding.

Our results concerning the influence of accentuation on the degree of coarticulation were consistent with Harrington et al. (2013) who had found a similar degree of coarticulation in both accented and deaccented positions: Our present results from the production data suggest that there is undershoot in deaccented position, but little change on the degree of vowel-on-fricative coarticulation in speech production across the prosodic conditions. However, the perceptual category boundaries were not flatter in deaccented position as expected, but listeners were found to compensate for coarticulation to about the same degree across the two prosodic contexts. These results suggest that listeners are highly attuned not only to coarticulatory variation, but also to the degree to which such variation is affected by prosodic context.

5. Acknowledgements

We would like to thank all participants of the experiments. This research was supported by the German Research Council grant number HA3512/9-1, awarded to the second author.

6. References

Cho, T. (2004). "Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English". In: *Journal of Phonetics*, 32, pp. 141–176.

Fowler, C. A. (2006). "Compensation for coarticulation reflects gesture perception, not spectral contrast". In: *Perception & Psychophysics*, 68, pp. 161–177.

Katz, W.F, C. Kripke and P. Tallal. (1991). "Anticipatory Coarticulation in the Speech of Adults and Young Children: Acoustic, Perceptual, and Video Data". In: *Journal of Speech and Hearing Research*, 34, pp.1222-1232.

Lindblom, B., H. Sussman and A. Agwuele (2009). "A duration dependent account of coarticulation for hyper-and hypoarticulation. In: *Phonetica*, 66, pp.188–195.

Harrington, J. (2010) *The Phonetic Analysis of Speech Corpora*. Wiley-Blackwell.

Harrington, J., F. Kleber, and U. Reubold (2013). "The effect of prosodic weakening on the production and perception of trans-consonantal vowel coarticulation in German". In: *Journal of the Acoustical Society of America*, 134, pp. 551–561.

Mann, V. A. and B. H. Repp (1980). "Influence of vocalic context on perception of the [ʃ]–[s] distinction". In: *Perception & Psychophysics*, 28, pp. 213-228.

Mann, V. and S. D. Soli (1991). "Perceptual order and the effect of vocalic context on fricative perception". In: *Perception & Psychophysics*, 49, pp. 399–411.

Nittrouer, S. and M. Studdert-Kennedy (1987). "The role of coarticulatory effects in the perception of fricatives by children and adults". In: *Journal of Speech, Language and Hearing Research*, 30, pp.73–93.

Nittrouer, S., & D. H. Whalen, (1989). "The perceptual effects of child-adult differences in fricative-vowel coarticulation". In: *Journal of the Acoustical Society of America*, 86, 1266–1276.

Proctor, M., C. H. Shadle, and K. Iskarous (2006). "An MRI study of vocalic context effects and lip rounding in the production of English sibilants". In: *Proccedings AICSST*, pp. 417–420.

Recasens, D. & A. Espinosa (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan, *Journal of the Acoustical Society of America*, 125, 2288-2298.

Repp, B. H. (1986). Some observations on the development of anticipatory coarticulation. In: *Journal of the Acoustical Society of America*, 79, 1616–1619.

Schiel, F., C. Draxler, and J. Harrington (2004). "Phonemic Segmentation and Labelling using the MAUS Technique" In: *Proceedings of the Workshop New Tools and Methods for Very-Large-Scale Phonetics Research*, Philadelphia, PA.

Soli, S. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. In: *Journal of the Acoustical Society of America*, 70, pp. 976–984.

Stevens K. N. and S. J. Keyser (2010) "Quantal theory, enhancement and overlap," In: *Journal of Phonetics*, vol. 38, pp. 10–19.

Whalen, D. H. (1981). "Effects of vocalic transitions and vowel quality on the English /s/-/ʃ/ boundary". *Journal of the Acoustical Society of America*, 69, pp. 275–282.