STRESS DISTINCTION IN GERMAN: MODELING KINEMATIC PARAMETERS OF TONGUE TIP GESTURES

Christine Mooshammer[1] and Susanne Fuchs[1,2]
[1]ZAS - Research Centre for General Linguistics, Berlin, Germany
[2]Speech and Language Science Department, Queen Margaret University College, Edinburgh, UK
timo@zas.gwz-berlin.de, s.fuchs@sls.qmced.ac.uk

ABSTRACT

Unstressed syllables are produced by a reduction of movement amplitudes and durations compared to stressed syllables. In order to investigate the effect of deaccentuation on tongue tip gestures in /tVt/ sequences with all vowels of German we recorded tongue tip movements of five speakers by means of EMMA. The movement paths of stressed items were manipulated to simulate kinematic parameters of unstressed items in three different ways: truncation, rescaling and combined truncation and rescaling. The following parameter of simulated movements were compared to measured unstressed items: durations and peak velocities of opening and closing movements, overall distance during syllable, the interval between velocity peaks in percent of syllable duration and the number of acceleration peaks between velocity peaks. The results indicate that no type of simulation can account for all parameters in all vocalic contexts. The combined simulation resembled most closely the kinematic parameters of unstressed items but could not generate the smaller amplitudes of lax unstressed syllables, since durational reduction of lax vowels due to deaccentuation was very small, i.e. the spatial reduction was not proportional to the temporal reduction for lax items.

1. INTRODUCTION

The supralaryngeal production of stressed syllables involves an increase of jaw movement amplitudes and movement durations compared to unstressed syllables (e.g. Kent and Netsell 1971). The reduction of movement duration for unstressed syllables was attributed to an increase of gestural stiffness by Kelso et al. (1985). However, changing this dynamic parameter couldn‘t account for reduced movement amplitudes.

Harrington et al. (1995) suggested two alternatives to the control by gestural stiffness: rescaling and truncation (cf. Fig. 2, upper and middle panel). Linear rescaling involves a proportionally scaled change of gestural spatial and temporal extent, i.e. a reduction of syllable duration changes the spatial extent of the gesture proportionally. Truncation is defined as a temporal overlap between the opening and the closing gesture in a CVC sequence, i.e. the opening gesture is truncated by the closing gesture. A closer phasing of the opening and closing gesture was also observed by Beckman et al. (1992) and Edwards et al. (1991).

These two alternatives were tested in Harrington et al. (1995) by truncating and rescaling jaw movement traces of stressed /bæb/ sequences. These simulated sequences and measured jaw movements of unstressed sequences were compared along the following parameters: durations, amplitudes and peak velocities of opening and closing movements, the interval between velocity peaks as a percentage of syllable duration (hereafter P2P ratio) and the number of acceleration peaks between velocity peaks. In case of a closer phasing between the opening and the closing gesture the P2P ratio decreases. The number of acceleration peaks also depends on the degree of gestural overlap: for untruncated sequences there is one acceleration peak for the deceleration phase towards the vowel and one peak for the acceleration phase towards the final consonant. With further gestural overlap the acceleration peaks mingle into one peak.

The comparison of the parameters of unstressed, truncated and rescaled sequences showed that the P2P ratio and the number of acceleration peaks for unstressed syllables most closely resembled the simulated truncated sequences. However, the reduction of movement amplitudes showed better agreement with rescaled items. Therefore the authors concluded that a combination of both would model best the kinematic parameters of deaccentuation.

Most articulatory studies on the accented/deaccented distinction only investigated CVC syllables with a single vowel type. One aim of our study is to extend the simulations carried out by Harrington et al. (1995) to the complete vowel inventory of Standard German. The second aim was to test whether a combination of both, rescaling and truncation, improves the results.

The third aim is to test whether the assumed predominance of truncation for the production of stress distinction also holds for apical gestures. The studies cited above analysed mainly mandibular movements. This is in agreement with the jaw expansion model proposed by Macchi (1985), i.e. stress is produced by a greater amount of jaw movement. However, as was found by de Jong (1995), not all subjects use the jaw for the production of stressed syllables. Therefore we analysed tongue tip movements in CVC sequences with the pre- and postvocalic apical stop /t/.

2. DATA AQUISITION

Tongue tip movements of five German speakers were recorded by means of EMMA at a sampling rate of 400Hz. The speech material consisted of nonsense syllables containing /tVt/with the vowels of German /i-ɪ,y-ʏ,e-ɛ,ɛː, ø-œ,aː-a,u-ʊ,o-ɔ/. Stress alternations were fixed by morphologically conditioned word stress and contrastive stress. Thus the first test syllable in the sentence "Ich habe /ˈtVtɐ/, nicht /tVˈtal/ gesagt" ("I said /ˈtVtɐ/, not /tVˈtal/") was always stressed and the second unstressed. All 15 sentences were repeated six (two speakers) or ten times (three speakers).

Movements of the tongue, lower lip and jaw were monitored by EMMA (AG100, Carstens Medizinelektronik). Four sensors were attached to the tongue (1 to 5 cm behind the tongue tip), one to the lower incisors and one to the upper lip. Two sensors on the nasion and the upper incisors served as reference coils to compensate for helmet movements during the recording session. The speech signal was recorded simultaneously on a DAT recorder.

The signal of the tongue tip sensor was smoothed by a low pass filter and its axis rotated to the main movement direction of the test sequence. Kinematic parameters of opening and closing gestures were determined by using a 20% threshold criterion of the tangential velocity signal (cf. Hoole et al. 1994, see Fig. 1). The movement amplitude of the whole sequence was computed as the integral of the tangential velocity between the beginning of the opening movement and the end of the closing movement.

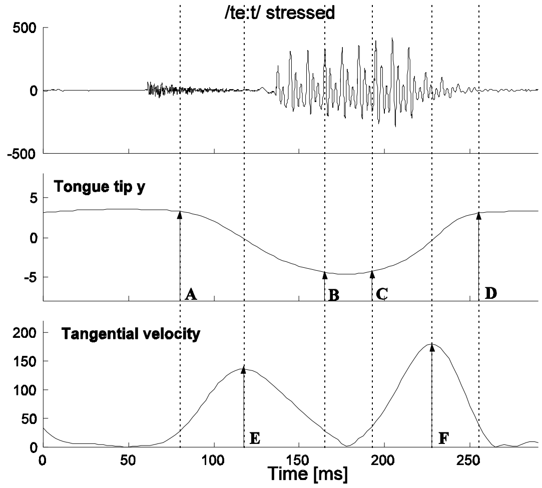

Figure 1: Target syllable /ˈteːt/. Upper panel: speech signal; middle panel: rotated tongue tip movement; lower panel: tangential velocity of tongue tip. A: begin of opening movement, B: end of opening movement, C: begin of closing movement, D: end of closing movement, E: velocity peak of opening gesture, F: velocity peak of closing gesture.

3. SIMULATIONS

Three sets of simulated data were generated by manipulation of tongue tip signals of stressed items: rescaled, truncated and combined (see Fig.2).

Rescaled curves were generated by reducing the maximal tongue tip lowering in linear proportion to the changes in duration in steps of 5 ms, multiplied by a reinforcement factor 1.5 which was determined empirically. For example, if a curve of 100 samples is shortened by two samples then the movement path is multiplied by 0.98 and therefore shrank. Truncation was made by cutting the opening movement at the point of maximal vertical opening for the vowel and overlapping it with the closing movement. Combined truncation and rescaling was achieved by overlapping the opening and closing gestures and shrinking them at the same time.

The amount of compression was determined by the interval between velocity peaks, i.e. the difference between the vowel specific mean of unstressed items and the actual item. As will be shown below the P2P ratio changed too much for simple truncation. Harrington et al. (1995) reported that the relationship between P2P ratio and stepwise overlap is not linear, i.e. the greater the overlap the more the P2P ratio decreases. Therefore the amount of truncation for the combined simulations was computed by using a curve which was approximated by averaging different curves of the P2P ratio change. Further shortening of the remaining portion of compression was achieved by rescaling. The following table shows the amount of compression by speaker and vowel. The first column per speaker, labeled O, gives the mean overall compression for all three simulation types in samples; the second column indicates the portion of combined compression (C).

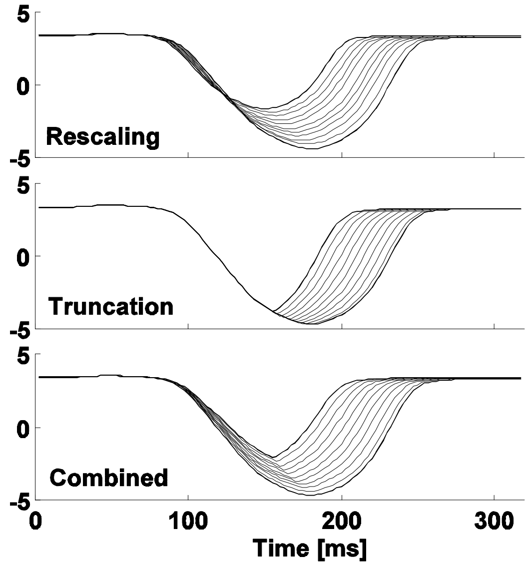

Fig. 2: Compression of /ˈteːt/ in 10 steps of 5 ms. Upper panel: rescaling, middle panel: truncation, lower panel: combined rescaling and truncation.

Table 1: Mean compression in samples by speaker and vowel. O= overall compression, C= portion of combined compression.

Speaker	RW		CG		DF		PJ		JD	
Compr.	O	C	O	C	O	C	O	C	O	C
i	10	7	15	7	10	2	15	9	14	8
y	9	9	18	11	10	4	21	17	13	5
e	14	10	22	13	14	9	22	17	23	18
ø	17	10	20	12	17	13	24	14	23	20
ɛː	21	13	31	18	25	19	26	17	37	27
aː	23	13	40	31	39	24	26	8	40	30
o	20	12	28	19	27	15	24	12	31	29
u	10	5	24	20	14	7	22	21	16	11
ɪ	2	2	5	2	1	1	6	6	8	6
ʏ	4	3	3	2	1	1	6	5	9	5
ɛ	2	2	3	1	3	1	6	3	8	3
œ	2	1	5	2	3	0	5	2	10	7
a	5	3	5	1	3	1	4	2	9	5
ɔ	2	1	5	3	6	4	4	1	9	5
ʊ	2	0	3	3	2	0	4	3	5	2

4. RESULTS

As can be seen in Table 1 the computed amount of compression differs considerably depending on speaker and vowel type. Generally, for lax vowels the compression is much less than for tense vowels. Furthermore for tense vowels the amount of compression decreases with vowel height, i.e. low vowels need more compression than high vowels.

Figure 3 shows the opening and closing durations of stressed, unstressed, truncated, rescaled and combined items averaged over speakers in the upper two panels. Comparing stressed to unstressed tense items deaccentuation reduces the duration of opening gestures to a greater amount than the duration of closing gestures. This effect was also shown for jaw movements by Summers (1987). Furthermore, durations of the opening gestures are much more variable than the closing durations which confirms the results of Gracco (1994).

Simulations of closing durations fit quite well the durations of the unstressed items for all three types, which is not true for simulated opening durations. The simulated durations do not exhibit the asymmetrical influence of stress on opening and closing durations, i.e. for all three simulation types the durations of the opening movements are longer than the opening durations of unstressed syllables with tense nuclei.

The P2P ratio, which is shown in the lower panel of Figure 3, is a measure for the amount of truncation. It decreases considerably due to deaccentuation for tense items but only very slightly for lax items. As was found by Kroos at al. (1997) the temporal parameters of lax stressed vowels exhibit a truncated pattern. Therefore, it is probably the case that lax vowels cannot be truncated further due to deaccentuation (see also Mooshammer et al. 1999).

For given compressions simple truncation yielded P2P ratios much smaller than the measured P2P ratios of unstressed sequences, whereas the influence of rescaling on the P2P ratios was too small, i.e. the P2P ratios of rescaled items were only slightly smaller than the measured ratios of stressed items. The best results were achieved for the combined simulations.

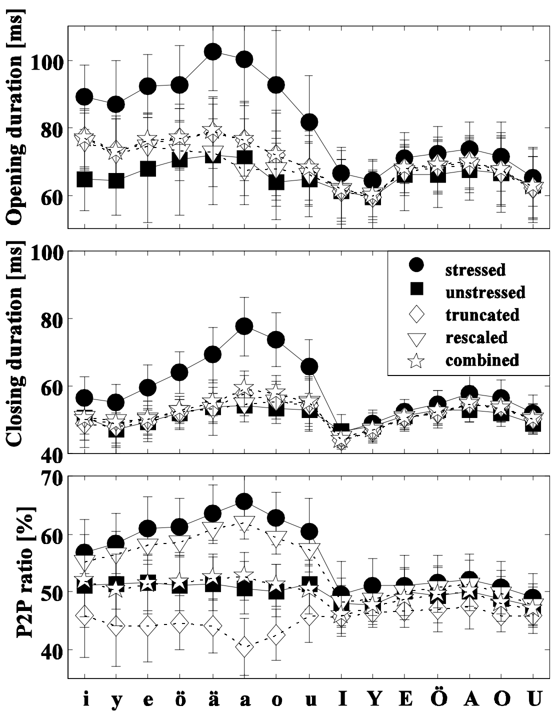

Figure 3: Upper panel: opening durations of stressed, unstressed, truncated, rescaled and combined items in ms; middle panel: closing durations; lower panel: P2P ratios. Lower case letters indicate tense vowels, upper case lax vowels.

Movement amplitudes of the tongue tip sensor during the syllable are shown in Figure 4, upper panel. Measured distances are reduced strongly going from stressed tense to unstressed tense items. Again, as for durations, deaccentuation influences lax items to a lesser degree. Low tense and back tense vowels are affected to a greater amount than high front vowels.

For truncated items the reduction of distances is not sufficient, which was also found by Harrington et al. (1995). The adjustment of distances to unstressed items is much better for simple rescaling and combined simulations, but not sufficient for lax sequences. Since the degree of shrinking depends on the compression from stressed to unstressed syllables and lax vowels are only slightly compressed there seems to be a differential behaviour between tense and lax vowels, i.e. for lax vowels the reduction of amplitudes is not proportional to the temporal shortening. Better results could probably be obtained with a higher reinforcement factor for lax vowels.

Velocity peaks of both, the opening and closing movements, show high correlation with movement amplitude. The reduction of velocity peaks due to deaccentuation is more prominent for closing gestures than for opening gestures. Truncation yielded an increase in peak velocity for both, opening and closing movement, i.e. the influence is in the wrong direction. Rescaling and combined simulation fit quite well to the data of unstressed tense items but not to the lax items.

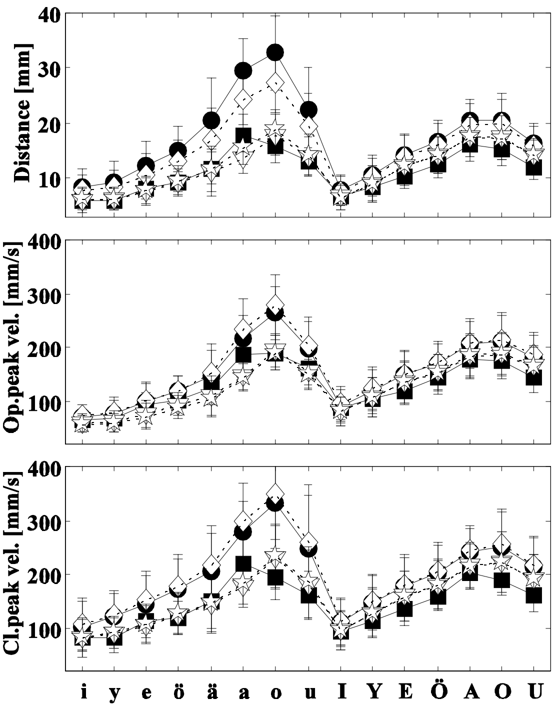

Figure 4: Upper panel: distances traveled during stressed, unstressed, truncated, rescaled and combined items in mm; middle panel: peak velocities of opening movements in mm/s; lower panel: peak velocities of closing movements in mm/s. Lower case letters indicate tense vowels, upper case lax vowels. For markers see legend in Fig. 3.

The last parameter considered here is the number of acceleration peaks between velocity peaks. As was shown by Harrington et al. (1995) acceleration curves of stressed items normally show two acceleration peaks, whereas for unstressed items only one peak was observed. Rescaling did not change the number of acceleration peaks but truncation did. The same tendency was found for the tense/lax distinction in German by Kroos et al. (1997) and confirmed by Mooshammer et al. (1999): Syllables with lax vowels tend to be produced with only one acceleration peak in the signal of the consonant articulator, whereas in sequences with tense vowels two or more peaks occur. For unstressed as well as for lax vowels the peak of the deceleration phase of the opening gesture merges in the acceleration phase of the closing gesture.

Table 2: Number of items with one, two or three acceleration peaks between velocity peaks. S=stressed, R=rescaled, T= truncated, C=combined rescaling and truncation, U=unstressed.

Nr.	Ten.					Lax				
Peak	S	R	T	C	U	S	R	T	C	U
1	37	69	308	259	265	257	286	292	290	270
2	241	247	12	59	49	35	6	0	2	5
3	42	4	0	2	0	0	0	0	0	0

As can be seen in Table 2 the number of acceleration peaks for the majority of tense items is reduced from two to one due to deaccentuation. This change is not achieved for rescaled simulations. For truncated tense items we found too many items with one peak. The best results were achieved for the combined simulations. As expected the reduction of acceleration peaks for lax vowels was not as dramatic as for tense vowels. Here the three types of simulations yield quite similiar results.

5. DISCUSSION

Three different types of manipulated movement paths of stressed CVC syllables were compared with measured tongue tip movements of unstressed syllables: truncation, rescaling and combined truncation and rescaling. The following parameters were used for evaluating the generated movements: durations and velocity peaks of opening and closing gestures, P2P ratio, distance traveled during the syllable and number of acceleration peaks between velocity peaks.

The first aim of this study was to extend Harrington et al.'s finding on the whole inventory of German vowels. It was shown here that vowels differ in the amount of necessary compression: low and back tense vowels are compressed to a greater degree than high front tense vowels. The P2P-ratio of stressed tense vowels varies with vowel height whereas for unstressed tense and lax vowels this parameter remains constant over different vowel types.

The second aim of this study was to compare the results of truncation, rescaling and combined truncation and rescaling with measured data of unstressed sequences. The best results were obtained for the combined model. The parameters duration of closing gesture, number of acceleration peaks between velocity peaks and P2P ratio fitted closely the parameters of the unstressed sequences. Distance and peak velocities resembled only the measured data of tense unstressed vowels. For lax vowels deaccentuation affected the reduction of the movement amplitudes to a greater degree as predicted by the slight temporal shortening. For rescaling and truncation it is assumed that changes in kinematic parameters are a consequence of temporal compression, but this prediction does not seem to hold for lax vowels. An incompressibility limit prbably prevents further shortening of the already short lax vowels. Therefore it can be concluded that none of the tested models accounts for all parameter changes due to deaccentuation. At least for lax vowels different motor control strategies seem to underlie the production of deaccentuation.

The third aim concerns the analysed articulator. Most studies investigated stress effects by mandibular movements in bilabial consonantal context. Tongue tip movements are composed of actively controlled tongue tip gestures and passive consequences of jaw movments. Nevertheless results for tongue tip gestures resembled quite closely the results for jaw gestures with the exception of the opening gesture durations. Therefore, the next step of this investigation will be analyzing the interaction of jaw and tongue movments in the production of the stress distinction.

Acknowledgements

Special thanks to Daniel Pape and Dirk Fischer for assistance with data analysis and to Phil Hoole, Tracy Hall and Bernd Pompino-Marschall for helpful comments. Work supported by German Research Council (DFG) grant GWZ 4/5-1.

References

Beckman, M., Edwards, J. and Fletcher, J. 1992. Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. J. Docherty and D.R. Ladd (eds.) *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, 68-86.

Edwards, J., Beckman, M.E. and Fletcher, J. 1991. The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America*, 89, 369-382.

Gracco, V. 1994. Some organizational characteristics of speech movement control. Journal of Speech and Hearing Research, 37, 4-37.

Harrington, J., Fletcher, J. and Roberts, C. 1995. Coarticulation and the accented/unaccented distinction: evidence from jaw movement data. *Journal of Phonetics*, 23, 305-322.

Jong, de KJ. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, 97, 491-504.

Kelso, J.A.S., Vatikiotis-Bateson, E., Saltzman, E., and Kay. B. 1985. A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. *Journal of the Acoustical Society of America*, 77, 266-280.

Kent, R.D. and Netsell, R. 1971. Effects of stress contrasts on certain articulatory parameters. *Phonetica*, 24, 23-44.

Kroos, C., Hoole, P., Kühnert, B. and Tillmann, H. 1997. Phonetic evidence for the phonological status of the tense-lax distinction in German. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)*, 35, 17-25.

Macchi, M. 1985. Segmental and supersegmental features of lip and jaw articulators. Ph.D. dissertation, New York University.

Mooshammer, C., Fuchs, S. and Fischer, D. 1999. Effects of stress and tenseness on the production of CVC syllables in German. *Proceedings of the 14th International Congress of Phonetic Sciences*, 409-412.

Summers, W.V. 1987. Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *Journal of the Acoustical Society of America*, 82, 847-863.