Articulatory gestures and focus marking in German

Anne Hermes, Johannes Becker, Doris Mücke, Stefan Baumann & Martine Grice

I*f*L Phonetik, University of Cologne, Germany
{anne.hermes; becker.johannes; doris.muecke; stefan.baumann; martine.grice}@uni-koeln.de

Abstract

This study reports on a production experiment investigating tonal and articulatory means of encoding different focus structures in German. Using an electromagnetic articulograph, we examined the movements of the upper and lower lips (related to sonority expansion) during the production of target words occurring in four different focus conditions. We found systematic differences not only between unaccented vs. accented target words (background vs. contrastive focus), but also *within* the category 'accented': the differences in articulatory expression for broad vs. contrastive focus were expressed by greater displacements and lower stiffness of lip aperture (opening and closing movements). Our results suggest that German speakers express discrete linguistic differences, namely differences in focus structure, by gradually but systematically varying sonority expansion in focus exponents across consonants and adjacent vowels, thus enhancing the syntagmatic contrast.

1. Introduction

In most studies dealing with information structure, *focus* and *background* are regarded as two distinct categories. Consequently, it is often assumed that the *prosodic marking* of these categories should be categorically distinct as well, thus reducing the prosodic analysis to the question of whether a constituent is accented or unaccented. More recent studies (e.g. [2]) have shown, however, that different focus structures are encoded by different accent types and/or by varying continuous parameters such as duration or pitch excursion on the focus exponents, thus creating different *degrees of prominence* on the respective items.

In the present study we are primarily concerned with the role of articulatory gestures in focus marking. The few previous investigations in this field are restricted to words in maximally diverging focus structures (contrastive focus vs. background) and thus to the accented-unaccented dichotomy (e.g. [5] for English and [1] for Italian). It is unclear from these studies, however, whether the articulatory differences found (e.g. greater jaw lowering or lip aperture in contrastive focus) are simply due to accentuation or whether the articulatory expression of different focus structures can be regarded as a continuum of prominence or emphasis (as reported in [6] for French).

In order to shed light on this question, we explore the variation in articulatory parameters which are related to lip kinematics (greater displacement, longer duration, higher peak velocity and lower stiffness of lip opening to enhance prominence) in the marking of target words occurring in different types of focus (contrastive, non-contrastive) and different sizes of focus domain (broad, narrow; see e.g. [9]), or in the background. In particular, we investigate differences *within* the category 'accent' (broad vs. narrow focus, broad vs. contrastive focus) as well as *between* accented and unaccented words (contrastive focus vs. background). Results on differences between background and broad focus as well as narrow and contrastive focus are not presented here.

1.1. Reading material

The speech material included question-answer sets eliciting four different focus structures: the NP under investigation occurred either as part of the previously mentioned background or in broad, narrow or contrastive focus. The target words, i.e. the fictitious names after *Dr.* [dɔktɐ], were always disyllabic, with the stressed syllable containing one of the four long target vowels /iː/, /aː/, /oː/ or /uː/. An example of a question-answer set is given below:

Questions:

1. Will Norbert Dr. Bahber treffen? *Does Norbert want to meet Dr. Bahber?*
2. Was gibt´s Neues? *What´s new?*
3. Wen will Melanie treffen? *Whom does Melanie want to meet?*
4. Will Melanie Dr. Werner treffen? *Does Melanie want to meet Dr. Werner?*

	Answers: Melanie will <u>Dr. Bahber</u> treffen.	**test word in:**
1.	[——]$_{focus}$	background
2.	[————————————]$_{focus}$	broad focus
3.	[——]$_{focus}$	narrow focus
4.	[——]$_{focus}$	contrastive focus

(lit.: *Melanie wants Dr. Bahber to-meet)*

1.2. Speakers and recordings

Three native speakers of Standard German (aged 26, 27 and 37) were recorded with a 2D Electromagnetic Midsagittal Articulograph (EMMA) and a time-synchronized DAT-recorder. The kinematic data were recorded at 500Hz, downsampled to 200Hz and smoothed with a 40Hz low-pass filter. The acoustic data were digitized at 44.1kHz.

The subjects listened to the questions (which were presented both visually and auditorily) and were instructed to answer these questions in a contextually appropriate manner and at a normal speech rate. After a test block of five question-answer-pairs each subject read out the target sentences (four focus structures, four target words, seven repetitions) in pseudo-randomised order, leading to 112 tokens per speaker in total.

Lip movements were monitored by EMMA (Carstens AG100), with sensors placed on the vermillion border of the upper and lower lip within the midsagittal plane. Two additional sensors on the nose and the upper gums served as a reference in order to correct helmet movements during the recordings.

1.3. Labelling procedure

Acoustic and articulatory data were labelled by hand using the EMU speech database system. A screen shot including all tiers and labels described below is given in Fig.1.

Segment boundaries of consonants and vowels of the accented and post-accented syllables (c1, v1, c2, v2) were annotated in the acoustic waveform.

In the tonal analysis we identified three different GToBI accent types on the target word (as proposed in [2]): !H* (downstep), ^H* (upstep) or H* (neither downstep nor upstep). Note that up- and downsteps are always related to a preceding prenuclear LH accent on the subject argument. Deaccentuation of the target word was marked with 'Ø'.

For the kinematic data, the lip aperture (LA) index was calculated in terms of the Euclidean distance between the two sensors of upper and lower lip, including movements both in the horizontal and vertical dimension [4]. Minima and maxima of opening and closing gestures (min1, max1, min2, max2) were located at zero-crossings in the respective velocity trace. Additionally, we labelled peak velocities at zero-crossings in the respective acceleration trace (p1, p2, p3). Twelve utterances (all from speaker WP) were removed from the analysis because no clear turning points for the lip kinematics could be identified.

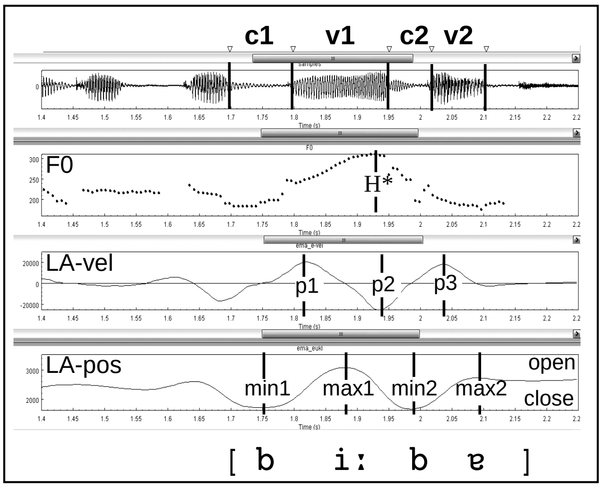

Figure 1: *Labelling scheme; from top to bottom: oscillogram, F0 curve, velocity and position curve of lip aperture (LA); target word B/i:/ber.*

2. Results and Discussion

We analysed all measures with one-way-ANOVAs for each speaker separately and with a Tukey post hoc test. The dependent variables included accent type and word duration for the acoustic measures, and displacement, peak velocity, duration and stiffness for the articulatory measures. The independent variable FOCUS STRUCTURE included broad focus, narrow focus, contrastive focus and background.

2.1. Accent types

Table 1 shows the accent types preferably used by the three speakers in the different focus structures. As expected, all speakers deaccented the target words when they occurred in the background. In broad focus, speakers DM and AH almost exclusively used downsteps (DM 85.2%; AH 100%), whereas speaker WP typically produced upsteps (84%; only 4% downsteps). In the narrow focus condition, speakers DM and WP both produced upsteps (DM 82.6%; WP 100%), while speaker AH used all three accent types nearly to the same extent (36% upsteps; 32% downsteps; 32% unmodified H*). In contrastive focus, all three speakers always used upsteps.

speaker	**back-ground**	**broad**	**narrow**	**contras-tive**
DM	Ø	!H*	^H*	^H*
AH	Ø	!H*	!H* H* ^H*	^H*
WP	Ø	^H*	^H*	^H*

Table 1: *Most frequently produced accent types per speaker and focus condition.*

2.2. Acoustic durations

We examined the duration of the target words for all speakers. Since our target words are disyllabic, the domain 'word' is identical with the domain 'foot'. Fig.2 shows mean durations of the target word B/i:/ber for the different focus conditions. For all three speakers, we found a significant increase in word duration from background to contrastive focus (e.g. AH: 33ms longer, $p<0.001$) as well as from broad to contrastive focus (e.g. DM: 45ms longer, $p<0.001$; WP: 34ms longer, $p<0.001$). However, none of the three speakers showed a significant increase in word duration from broad to narrow focus.

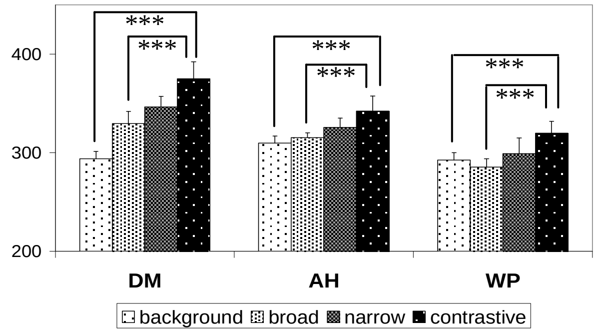

Figure 2: *Duration (ms) of target word B/i:/ber; speaker DM, AH and WP.*

2.3. Kinematic results

Kinematic results are presented for two speakers for the vowel /i:/. Fig.3 shows averaged trajectories for the distance between upper and lower lip (Lip Aperture) during the production of the target word B/i:/ber, for each focus condition separately.

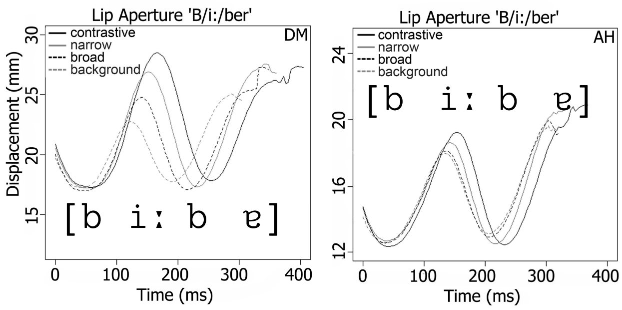

Figure 3: *Averaged contours for Lip Aperture, speaker DM and AH, target word B/i:/ber.*

Low displacements indicate that the lips are closed for the production of the stop consonants, while high values indicate open lips during the vowels. Going from background through broad and narrow to contrastive focus we can see an increase in duration and displacement (corresponding to lip aperture).

Table 2 provides mean durations, displacements, peak velocities (=maximum speed) and stiffness values for the opening and closing gestures during the production of the CV:.CV sequence in B/i:/ber. The opening gesture is defined as the movement of the primary constrictors (upper and lower lip) from the initial consonant (maximum of consonantal closure) to the following vowel (maximum of transvocalic opening) in the accented syllable. The closing gesture is the movement from the vowel to the intervocalic consonant (maximal closure); the intervocalic consonant is the onset of the post-accented syllable.

OPENING GEST.		back-ground	broad	narrow	contrast
DM	dur (ms)	79 (5)	92 (7)	100 (6)	107 (12)
	dis (mm)	5.6 (0.6)	7.9 (1)	9.9 (0.9)	11.4 (1.2)
	p-vel (mm/s)	121 (11)	153 (17)	183 (17)	199 (10)
	stiffness	21.6 (0.9)	19.3 (1)	18.5 (1.1)	17.5 (1.6)
AH	dur (ms)	99 (6)	100 (4)	104 (5)	112 (5)
	dis (mm)	5.5 (0.6)	5.6 (1)	6.1 (0.5)	7.1 (0.8)
	p-vel (mm/s)	92 (10)	95 (21)	95 (11)	103 (10)
	stiffness	16.8 (0.9)	16.9 (0.7)	15.7 (1)	14.5 (0.8)
CLOSING GEST.		back-ground	broad	narrow	contrast
DM	dur (ms)	68 (2)	75 (5)	78 (3)	86 (10)
	dis (mm)	5.2 (0.6)	8.1 (1.1)	10.1 (1.1)	11.9 (1.4)
	p-vel (mm/s)	128 (11)	177 (15)	214 (19)	238 (11)
	stiffness	24.4 (0.9)	22.1 (1.2)	21.3 (0.8)	20.1 (1.4)
AH	dur (ms)	70 (3)	71 (3)	72 (2)	75 (4)
	dis (mm)	5 (0.7)	5.2 (1)	6.3 (0.4)	7.2 (1.1)
	p-vel (mm/s)	114 (16)	121 (22)	141 (7)	157 (19)
	stiffness	22.9 (0.7)	23.1 (1.3)	22.4 (0.8)	21.8 (0.9)

Table 2: *Durations, displacements, peak velocities and stiffness of opening and closing gestures in the CV:.C sequence; standard deviations in brackets.*

All mean values in Table 2 *increase* from background through broad and narrow focus to contrastive focus (see Table 2). We found a large effect of the variable FOCUS STRUCTURE on all parameters for speaker DM in the opening gesture (duration: [F(3, 28) = 17.055, p<0.001], displacement: [F(3, 28) = 48.768, p<0.001], peak velocity: [F(3, 28) = 41.076, p<0.001], stiffness: [F(3, 28) = 14.802, p<0.001]) and in the closing gesture (duration: [F(3, 28) = 10.512, p<0.001], displacement: [F(3, 28) = 50.308, p<0.001], peak velocity: [F(3, 28) = 79.166, p<0.001], stiffness: [F(3, 28) = 18.500, p<0.001]). For speaker AH, the factor FOCUS STRUCTURE also reached significance in the opening gesture for the parameters duration [F(3,28) = 8.141, p<0.001], displacement [F(3, 28) = 6.391, p<0.01] and stiffness [F(3, 28) = 12.285, p<0.001], and in the closing gesture for the parameters displacement [F(3, 28 = 10.556, p<0.001] and peak velocity [F(3, 28 =9.404, p<0.001].

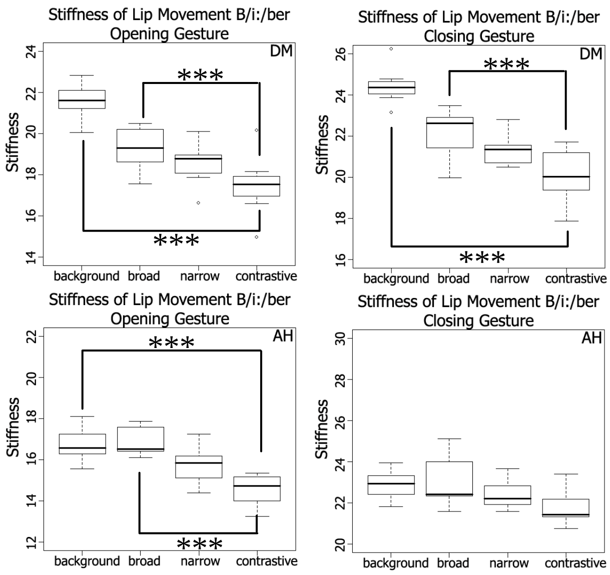

Figure 4: *Stiffness (calculated as peak velocity in relation to max. displacement), speaker DM and AH.*

In addition, Fig.4 provides medians and quartiles for stiffness in the opening gesture (left) and in the closing gesture (right), for speaker DM (above) and speaker AH (below) separately (for the vowel /i:/). Stiffness is related to the relative speed of the articulatory movement, and affects the duration of a movement. A gesture with a lower stiffness reaches the target later than a gesture with a high stiffness. Lower stiffness has been reported on in the literature for accented syllables in contrast to unaccented ones (e.g. [3] for English). We calculated stiffness as specified in [7] and [8]:

$$\text{Stiffness} = \frac{\text{peak velocity (cm/sec)}}{\text{maximum displacement (cm)}}$$

There were systematic differences observable in the distribution of the boxes: for both speakers stiffness continuously decreased from contrastive through narrow and broad focus to background.

2.3.1. *Background vs. contrastive focus*

As reported in the ANOVA above, all parameters were affected in measurements for background vs. contrastive focus (except for speaker AH, for whom peak velocity in the opening gesture, and duration and stiffness in the closing gesture did not reach significance).

During the production of the *opening gesture*, speaker DM considerably increased the displacement (on average 5.8mm larger from background to contrastive focus, p<0.001), the peak velocity (78mm/s faster, p<0.001) and the duration (28ms longer, p<0.001). Speaker AH increased the displacement (1.6mm larger, p<0.01) and the duration (13ms longer, p=0.001). Contrary to speaker DM, she did not modify the peak velocity (p=0.436ns). However, both speakers considerably lowered the stiffness parameter (DM: -4.1 lower stiffness, p<0.001; AH: -2.3 lower stiffness p<0.001).

In the *closing gesture* (as in the opening gesture), speaker DM increased the displacement from background to contrastive focus (6.7mm larger, p<0.001), the peak velocity

(110mm/s faster, $p<0.001$), and the duration (18ms longer, $p<0.001$). Speaker AH increased displacement (2.2mm larger, $p<0.001$) and peak velocity (43mm/s faster, $p<0.001$), while the duration remained the same (p=0.068ns). While speaker DM lowered the stiffness (-4.3 lower stiffness, $p<0.001$), speaker AH did not.

To sum up, we found higher targets for contrastive focus than for background. Both speakers reached larger displacements in *both* the opening *and* closing gesture to enhance prominence. However, they differ in the way they reach the targets:

Speaker DM adjusted all parameters in the opening and closing gesture. She increased the maximum displacement, the peak velocity and the duration of the movement. Furthermore, she lowered the stiffness. In a mass-spring model, these adjustments are predictable [4,5] by a non-proportional change of the control parameters TARGET and STIFFNESS. There is a stronger modification of the TARGET control parameter leading to higher velocities in contrastive focus.

Speaker AH did not adjust all parameters. In the opening gesture, she increased the displacement and the duration of the movement, but not the peak velocity. Furthermore, she lowered the stiffness. In the closing gesture, she increased the displacement and the peak velocity, but not the duration. Stiffness values remained the same. In a mass-spring model, those adjustments are predictable [4,5] by a proportional change of the control parameters TARGET and STIFFNESS in the opening gesture, and by a pure TARGET change in the closing gesture.

2.3.2. Broad focus vs. contrastive focus

All parameters reported above (in the ANOVA) reached significance in comparison across broad and contrastive focus. Both speakers adjusted their articulation of the opening and the closing gesture to enhance prominence from broad to contrastive focus. The strategies are comparable to those reported in 2.3.1 for contrastive focus vs. background.

In the *opening gesture*, speaker DM considerably increased the displacement (3.5mm larger from broad focus to contrastive focus, $p<0.001$), the peak velocity (46mm/s faster, $p<0.001$) and produced longer durations (15ms longer, $p<0.01$). Speaker AH increased displacement (1.5mm/s larger, $p<0.01$) and duration (12ms longer, $p<0.01$), but not the peak velocity (p=0.697ns). Both speakers lowered the stiffness in the opening gesture (speaker DM: -1.8 lower stiffness, speaker AH: -2.4 lower stiffness, $p<0.001$).

In the *closing gesture*, speaker DM strongly increased displacement (3.8mm larger, $p<0.001$), peak velocity (61mm/s, $p<0.001$) and duration (11ms longer, $p<0.05$). Speaker AH increased displacement (2mm larger, $p<0.001$) and peak velocity (36mm/s faster, $p<0.01$), but she did not modify the duration (p=0.107ns). While DM lowered the stiffness (-2 lower stiffness, $p<0.001$), speaker AH did not.

To sum up, prominence is enhanced from broad focus to contrastive focus in essentially the same way as from background to contrastive focus.

2.3.3. Broad focus vs. narrow focus

All values increase from broad to narrow focus. However, only a few parameters reached significance: speaker DM increased displacements in the opening (2mm larger, $p<0.01$) and closing gesture (2mm larger, $p<0.01$). Furthermore, she increased the peak velocity in the opening (30mm/s faster, $p<0.01$) and closing gesture (37mm/s faster, $p<0.001$). All other parameters remained the same even though they gradually increased from broad to narrow focus.

3. Conclusions

We found systematic differences in a number of articulatory parameters corresponding to sonority expansion between the marking of constituents as background on the one hand, and as contrastive focus on the other. Since the former is unaccented and the latter always bears an accent, this might suggest that sonority expansion is a concomitant of accentuation. However, we also found differences in articulatory adjustments *within* the category ‘accent’ (recall that both broad and contrastive focus both involve an accent on the target word): Speakers distinguished between broad and contrastive focus by showing larger displacements and lower stiffness.

Thus we show that speakers of German differentiate between discrete linguistic categories relating to focus structure not only with accentuation but also by varying aspects of their supralaryngeal articulation patterns resulting in the enhancement of the syntagmatic contrast between consonants and vowels on and around the target syllable.

4. Acknowledgements

The research reported here was funded by German Research Council (DFG) grant GR1610/3-1 (TAMIS).

5. References

[1] Avesani, C., Vayra, M., Zmarich, C. 2007. On the articulatory bases of prominence in Italian. *Proceedings 16th ICPhS 2007*, Saarbrücken, 981-984.

[2] Baumann, S., Becker, J., Grice, M., Mücke, D. 2007. Tonal and articulatory marking of focus in German. *Proceedings 16th ICPhS 2007*, Saarbrücken, 1029-1032.

[3] Beckman, M. E., Edwards, J., & Fletcher, J. 1992. Prosodic structure and tempo in a sonority model of articulatory dynamics. In: *Papers in Lab.Phon.II.* Cambridge University Press. S. 68-86.

[4] Byrd, D. 2000. Articulatory vowel lengthening and coordination at phrasal junctures. In: *Phonetica* 57, 3-16.

[5] Cho, T. 2005. Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ, i/ in English. In: *JASA* 117(6), 3867-3878.

[6] Dohen, M., Lœvenbruck, H., Hill, H. 2006. Visual correlates of prosodic contrastive focus in French: Description and inter-speaker variabilities. *Proceedings Speech Prosody 2006*, Dresden, 221-224.

[7] Munhall, K.G., Ostry, D.J., Parush, A. 1985. Characteristics of velocity profiles of speech movements. *Journal of Experimental Psychology: Human Perception and Performance* 11, 457-474.

[8] Roon, K.D., Gafos, A.I., Hoole, Ph., Zeroul, Ch. 2007. Influence of articulator an manner of stiffness. *Proceedings 16th ICPhS 2007*, Saarbrücken, 409-412.

[9] Uhmann, S. 1991. Fokusphonologie. Eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie. Tübingen: Niemeyer.