

Interarticulator cohesion within coronal consonant production^{a)}

Christine Mooshammer^{b)}

Institut für Phonetik und digitale Sprachverarbeitung, Christian-Albrechts Universität, Kiel, 24098 Germany

Philip Hoole

Institut für Phonetik und sprachliche Kommunikation, Ludwig-Maximilians-Universität, Munich, Germany

Anja Geumann

School of Computer Science and Informatics, UCD, Dublin, Ireland

(Received 3 May 2005; revised 29 March 2006; accepted 5 May 2006)

If more than one articulator is involved in the execution of a phonetic task, then the individual articulators have to be temporally coordinated with each other in a lawful manner. The present study aims at analyzing tongue-jaw cohesion in the temporal domain for the German coronal consonants /s, ʃ, t, d, n, l/, i.e., consonants produced with the same set of articulators—the tongue blade and the jaw—but differing in manner of articulation. The stability of obtained interaction patterns is evaluated by varying the degree of vocal effort: comfortable and loud. Tongue and jaw movements of five speakers of German were recorded by means of electromagnetic midsagittal articulography (EMMA) during /aCa/ sequences. The results indicate that (1) tongue-jaw coordination varies with manner of articulation, i.e., a later onset and offset of the jaw target for the stops compared to the fricatives, the nasal and the lateral; (2) the obtained patterns are stable across vocal effort conditions; (3) the sibilants are produced with smaller standard deviations for latencies and target positions; and (4) adjustments to the lower jaw positions during the surrounding vowels in loud speech occur during the closing and opening movement intervals and not the consonantal target phases.

© 2006 Acoustical Society of America. [DOI: 10.1121/1.2208430]

PACS number(s): 43.70.Bk, 43.70.Aj [AL]

Pages: 1028–1039

I. INTRODUCTION

In the execution of a speech task, e.g., complete occlusion of the vocal tract with the lips, the articulators involved constitute a system with multiple degrees of freedom. Therefore, it is generally acknowledged that multiarticulatory tasks are not accomplished by individual control of each composite articulator. Instead, hierarchically ordered coordinative structures are assumed to orchestrate spatially and temporally the individual movements, thereby simplifying the control of multiple muscle activities (see, e.g., Fowler *et al.*, 1980).

Coordinative structures as organizational units have been included in speech production models with radically different assumptions concerning the nature of the motor plan. Thus, for example the task dynamic model (Saltzman and Munhall, 1989) is based around the concept of key vocal-tract constrictions, while the DIVA model (Guenther *et al.*, 1998, 1999) assumes that the goals of speech movements are defined in an auditory or perceptual space. As a further key issue the temporal aspect of speech motor control has been given much attention in recent years because it was found that the contributing articulators do not all start moving at the same time but with a certain order and timed with respect to specific articulatory events of each other (see, e.g.,

Gracco and Abbs, 1986). The focus of most studies was on the strength of the functional linkage between varying component articulators such as lip-jaw vs tongue tip-jaw in Hertrich and Ackermann (2000) or lip-jaw vs velum-jaw in Kollia *et al.* (1995) and across varying suprasegmental conditions (stress and speech tempo in Kelso *et al.*, 1986; speech tempo in DeNil and Abbs, 1991; Nittrouer *et al.*, 1988; Nittrouer, 1991; and Shaiman *et al.*, 1995). Up to now, no articulatory timing data have been available for varying manners of articulation within a single set of component articulators. The general aim of the current study is to investigate the timing of the tongue tip and jaw for the coronal consonants /s, ʃ, t, d, n, l/, i.e., for consonants produced with the same set of articulators but different constriction degrees and additional features such as voicing or velar opening.

A fixed succession of articulatory events for achieving a phonetic goal has been interpreted as strong interarticulator cohesion and evidence for coordinative structures (see Fowler *et al.*, 1980; Saltzman and Munhall, 1989). For example, for the bilabial closure a very consistent advancement of the upper lip velocity peak for the closing movement relative to the lower lip and jaw velocity peaks has been found (see, e.g., Gracco and Abbs, 1986; Gracco, 1988; van Lieshout, 1995; Kollia *et al.*, 1995). The stability of such relatively time-locked interarticulator cohesion has been experimentally tested by varying external parameters such as speech rate and stress placement. The assumption is that the task-specific organizationally invariant timing of composite articulators is achieved by a motor program which function-

^{a)}Portions of this work were presented at the 15th International Congress of Phonetic Sciences, Barcelona, Spain, August 2003.

^{b)}Author to whom correspondence should be addressed. Electronic mail: timo@ipds.uni-kiel.de

ally organizes multiarticulate speech movements for a phonetic gesture, while local parameters of the pattern, such as speech tempo, are left to vary freely and are considered parameters of the program (see Shaiman *et al.*, 1995).

The timing between the executing organs is assumed to be almost invariant and stronger for gestures within a phoneme than between phonemes (see Saltzman *et al.*, 1998). The evidence for different degrees of cohesion, termed “glue” by Saltzman *et al.* (2000), was found by perturbation experiments: the relative timing was shifted to a lesser degree when the perturbation occurred within an actively controlled gesture rather than at the borders.

In former studies the strength of cohesion has been found to vary with several aspects, e.g., closing movements are more tightly coupled than opening movements (e.g., Gracco, 1988; Hertrich and Ackermann, 2000). Furthermore, articulators controlled by different tract variables such as the vocal folds and velum exhibit a smaller degree of interarticulator cohesion with the jaw than articulators controlled by a single tract variable such as upper lip and jaw (see, e.g., Gracco and Löfqvist, 1994; Kollia *et al.*, 1995) as measured by a greater variability in timing parameters. This tendency implies that interarticulator cohesion is stronger than intergestural cohesion. Finally, consistent timing relationships between articulators are more often found for the peak velocity as compared to on- and offsets of movements (Gracco, 1988; van Lieshout 1995).

Evidence for a fixed timing relationship between the upper lip and the jaw across suprasegmental manipulations was found by, e.g., Kelso *et al.* (1986) using a phase plane plot, which shows the velocity of the jaw during a VCV sequence on one axis and its position on the other. The timing relationship to the upper lip was then expressed as an angle in this plane. This phase angle varies with phonetic identity but not with speech rate and stress according to Kelso *et al.* (1986). However, later studies yielded contradictory results: the phase angle was found to decrease if the jaw cycle duration decreased, e.g., at fast speech rate or for unstressed syllables more of the cycle was occupied by the upper lip lowering for bilabial consonants (see Nittrouer *et al.*, 1988; Shaiman *et al.*, 1995) or the tongue tip closing and closure for apical consonants (Nittrouer, 1991). These latter results indicate that rate and stress manipulations do also affect the spatiotemporal relationships among articulators.

Most studies focused on bilabial consonants and the spatiotemporal coordination between the lips and the jaw. Up to now, the *timing* between the involved articulators for different manners of articulation has not been studied. However, a number of studies have examined the *spatial* contribution of the jaw to the production of vowels and consonants with varying manners and places of articulation. Since in the present study articulatory properties of the coronal consonants will be investigated, only earlier results on /s, ʃ, t, d, n, l/ will be reviewed here. Based on jaw positions, previous studies found that the jaw’s contribution varies for these consonants, e.g., a closed and very precisely controlled jaw position is essential for the sibilants /s/ and /ʃ/ in order to provide a second noise source by a small distance between the upper and lower teeth (Geumann *et al.*, 1999; see also, e.g.,

Shadle, 1990; Lee *et al.*, 1994; Howe and McGowan, 2005). For /l/ a low jaw position is advantageous in order to provide space for the more apical articulation—as opposed to a flat and laminal articulation—and to avoid lateral contact between the tongue sides and the posterior parts of the alveolar ridge (see, e.g., Lindblad and Lundqvist, 1999 and Geumann, 2001a). Geumann (2001a) suggested that apicality also seems to play a role for the voiced stop /d/ which is then produced with a lower jaw position than /t/, but as Dart (1991) and Geumann (2001a) pointed out, the choice between an apical or laminal articulation in languages such as French, English, and German depends on the speaker. Highly consistent results have been found for the jaw positions during /t/, which were only slightly lower than the sibilants’ and almost as invariant (see, e.g., Kühnert *et al.*, 1991; Keating *et al.*, 1994; Lee *et al.*, 1994; Geumann *et al.*, 1999).

The present study aims at investigating the temporal coordination between the tongue tip and the jaw for the coronal consonants /s, ʃ, t, d, n, l/ in German. As was found for spatial parameters, the jaw does not contribute uniformly to the production of the consonants under consideration. In the first part of the Results section, jaw and tongue tip positions will be analyzed for the coronal consonants. Whether the differential role of the jaw also has consequences for the timing between the jaw and the tongue tip will be assessed in the second part by analyzing the intervals between specific articulatory events of the tongue tip and the jaw in VCV sequences, such as the closing movement offset and opening movement onset. In addition to these latencies, the latencies of the peak velocities are taken into account because it was found that the peak muscle activation correlates quite well with peak velocities (see Gracco, 1988). One specific hypothesis concerning the timing is that, if the task of the jaw is simply to lift the tongue tip up towards the alveolars, the tongue tip and the jaw should move in relative synchrony, with any time lag attributable to differences in kinematic properties of the articulators involved such as generally slower jaw movements as compared to tongue tip movements (see Tasko and Westbury, 2002). A third aim of this study is to investigate the strength of temporal cohesion for varying manners of articulation. Therefore, the variability of positions and latencies was compared for the six coronal consonants.

In order to check whether the observed differences between manners of articulation are stable across different conditions, the data were recorded at two vocal effort levels: normal and speaking up without shouting. As was found, e.g., by Schulman (1989) and Geumann (2001a), the excursion of the jaw movement towards the vowel is larger in loud speech, whereas the consonants were less affected. Therefore, speaking up can be interpreted as an up scaling for the vowel-directed movement but not the consonant related, and the closing gesture will be mainly affected because of lower jaw positions during the vowel. If tight cohesion obtains, then the temporal and spatial tongue-jaw coordination during the consonants should not be affected by the lower jaw positions of the surrounding vowels. Furthermore, Dromey and Ramig (1998) showed that token-to-token variability of articulatory measures decreased for higher levels of vocal ef-

fort. More subtle differences in timing between manners of articulation are therefore expected to be found more easily in loud speech.

II. METHOD

A. Speakers

Five native speakers of German, one female (AW) and four male (KH, RS, SR, UR), were recorded by means of electromagnetic midsagittal articulatory graphy. The age of the speakers ranged from 23 to 31 and none of the speakers had a known history of speech or hearing problems. The speakers were students or faculty staff of the Institute of Phonetics and Speech Communication at the University of Munich. They were not familiar with the aims of this study.

B. Speech material

This study is based on the same set of data as reported in Geumann *et al.* (1999); Geumann (2001a, b). The six coronal consonants /s, ʃ, t, d, n, l/ were recorded in symmetrical VCV sequences. The vowel context consisted of /i/, /e/, and /a/. Only items with surrounding low vowels /a/ will be considered here because jaw movements for high vowels were too small and noisy for analysis of movement kinematics. The first vowel was always stressed and long and the second one unstressed but unreduced. All VCV sequences were embedded in the carrier phrase “*Hab das Verb ___ mit dem Verb ___ verwechselt*” (I mixed up the verb ___ with the verb ___) and occurred equally often in the first and in the second positions. Therefore, both target sequences received contrastive sentence accent. The sentences were repeated six times in randomized order, which gives 12 repetitions per item and vocal effort condition. Stimuli were presented on a computer screen.

The increase in vocal effort was elicited by instructing the subjects to speak as loud as possible without shouting. They were told to imagine that, with the microphone turned off, they had to be heard in the control room adjacent to the recording room. In the normal condition, the speakers were instructed to speak at a comfortable vocal effort level. Since both conditions were randomly varied, the loud condition was additionally marked on the prompt screen below the test sequence.

By measuring the rms amplitude during the vowels, we assessed whether all speakers increased the intensity significantly (for details see Geumann, 2001a). Speakers varied in the amount of vocal effort increase. Speaker UR almost shouted; he generally spoke with the highest intensity for loud speech and largest difference between the two vocal effort levels (mean sentence intensity for UR for the normal condition was 61 dB and for the loud condition 72 dB). The smallest changes were observed for speakers AW and KH, with a change from normal to loud vocal effort level of about 5 dB.

C. Procedure

Articulatory data were collected by using the electromagnetic midsagittal articulatory graphy AG100 manufactured by

Carstens Medizinelektronik (for details on the measurement principle see Hoole and Nguyen, 1999). Four sensors were glued on the tongue surface by using dental cement (Ketac). For the current study only the tongue tip sensor, placed approximately 1 cm behind the apex, was analyzed. For monitoring jaw movements three sensors were placed on the outer and inner surface of the lower gums and the angle of the chin. Two sensors on the bridge of the nose and the upper incisors were recorded for the correction of head movements.

After the recording session, data were rotated to the occlusal plane and the origin of the new coordinate system was located at the lower edge of the upper incisors. The procedure to orient the data with the horizontal axis parallel to the occlusal plane was as follows: The investigator made a trace of the subject’s hard palate during the experiment using a spare sensor. Then, this trace was aligned with a hard-palate trace taken from a dental impression placed in the EMMA apparatus. A plastic t-bar bearing two sensors was placed on the dental impression (resting on the upper incisors at the front and the second molars at the back) to provide a definition of occlusal plane orientation.

The articulatory data were sampled at a frequency of 500 Hz. For further processing all signals were down-sampled to 250 Hz and low-pass filtered with a finite impulse response (FIR) filter (Kaiser window design, -6 dB at 50 Hz). Horizontal, vertical, and tangential velocities were calculated and smoothed with a further Kaiser window filter (-6 dB at 20 Hz).

The measured tongue tip signal is composed of the active tongue tip and the jaw. Thus, the tongue tip signal has to be decomposed into the active tongue tip movement and the passive consequence of the jaw movements (for an extensive overview see Westbury *et al.*, 2002), which is complicated by the fact that the measured jaw movement consists of a rotational and a translational component. From MRI data (for details of data acquisition see Hoole *et al.*, 2000) for each speaker the exact position of the mandibular condyle was obtained and mapped onto the EMMA coordinates. Distances between condyle and outer jaw and condyle and tongue sensors on the midsagittal plane were calculated at the temporal midpoint of consonant production for each speaker. The tongue-to-condyle distance in percent of the outer-jaw-to-condyle distance was taken as a weighting factor for the jaw. This procedure, which follows that of Edwards (1985), was applied because simple subtraction neglects the fact that jaw rotation affects the tongue tip to a greater degree than the tongue back. The resulting signals are termed intrinsic tongue tip for the remainder of this article.

D. Analysis

Figure 1 shows the labeling criteria. In the upper part the speech signal, the vertical jaw movement, and the derived tangential velocity of the utterance [a:sa] are presented. The last two panels show the vertical intrinsic tongue tip movement and again the derived tangential velocity signal. Movement cycles of the intrinsic tongue tip¹ and jaw were segmented into closing and opening intervals by using a 20% threshold criterion of the peak tangential velocity as shown

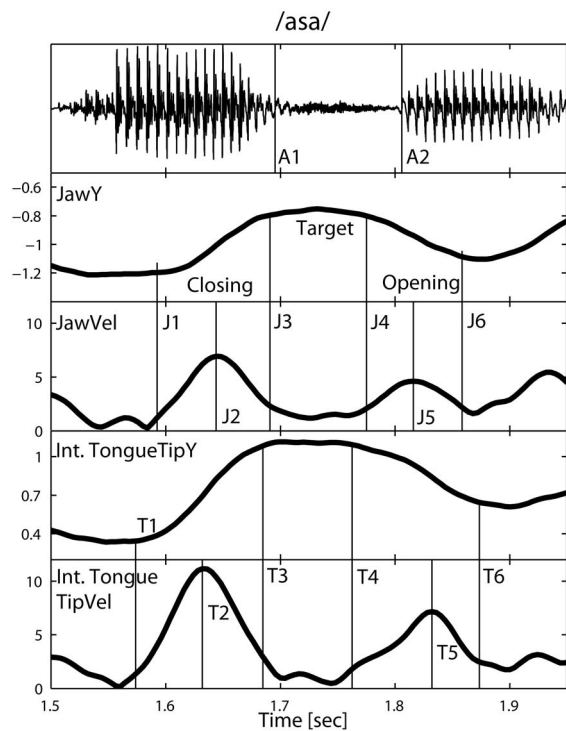


FIG. 1. Labeling criteria: Upper to lower panel: audio signal of [a:sa] by speaker RS, vertical jaw movement in cm, tangential velocity signal of jaw in cm/s, vertical intrinsic tongue tip signal in cm, tangential velocity signal of intrinsic tongue tip in cm/s. Vertical lines: acoustical on- and offset of the consonant (A1 and A2), for the intrinsic tongue tip and the jaw, respectively, onset of closing movements (T1 and J1), peak velocity of the closing movement (T2 and J2), offset of the closing movement (T3 and J3), onset of the opening movement (T4 and J4), peak velocity of the opening movement (T5 and J5), and the offset of the opening movement (T6 and J6).

in this figure. The threshold criterion was applied because the onset of an opening or closing movement cannot be consistently labeled by simply using a zero crossing (or minimum when analyzing the tangential velocity) of the velocity signal, since usually multiple zero crossings occur during and after the target phase. As was shown by Kroos *et al.* (1997), a 20% threshold of the tangential velocity yields the most stable results as assessed by comparing the variability of data for selected absolute and relative threshold values. The hold duration for the consonant was defined operationally as the interval between the offset of the closing movement and the onset of the opening movement (see in Fig. 1 time points T4-T3 for the intrinsic tongue tip hold phase and J4-J3 for the jaw hold phase). Even though we are aware of the fact that this interval, in which relatively little movement occurs, is not the same as the intended target of an abstract gesture, this phase will be termed *target* of the intrinsic tongue tip or jaw for reasons of simplicity.

From these landmarks latencies were computed by subtracting the corresponding time points of the tongue tip from the jaw, with the number landmarks shown in Fig. 1. The following tongue-jaw latencies were calculated: the velocity peaks of the closing movement ($LatV_{cl}=J2-T2$), the target achievement ($Lat_{on}=J3-T3$), the end of the target ($Lat_{off}=J4-T4$), and the velocity peaks of the opening movement ($LatV_{op}=J5-T5$) (the abbreviations given in brackets are used for the tables below).

Latencies were also calculated in relation to the acoustic onset and offset of the consonants. The former was set at the end of high energy in $F2$ for the obstruents or a general energy drop for the nasal or the lateral. The offset was specified at the burst for /t, d/, the beginning of regular voicing for the sibilants, and a rise in energy for the nasal and the lateral. Because the intrinsic tongue tip hold phase onset and offset was well aligned with the acoustically defined events, only the jaw closing movement offset relative to the acoustically defined consonant onset ($AcJaw_{on}=J3-A1$) and the jaw opening movement onset relative to the acoustically defined consonant offset ($AcJaw_{off}=J4-A2$) are discussed in this study. For all latencies negative values indicate a jaw advancement and positive values a tongue tip advancement.

The latencies for consonantal target onsets and offsets were highly correlated with the acoustic duration of the consonants, which varied between the mean values of 130 ms for /s/ and 40 ms for /d/. Therefore, the latencies Lat_{on} , Lat_{off} , $AcJaw_{on}$, and $AcJaw_{off}$ were normalized individually to the acoustic consonant durations, i.e., the latencies were divided by the corresponding acoustic duration ($A2-A1$). Because no reasonable reference duration could be used for the latencies of the velocity peaks, no normalization was applied to $LatV_{cl}$ and $LatV_{op}$.

Intrinsic tongue tip and jaw positions were extracted at the 20% threshold of the closing movement onset during the initial vowel (time points T1 and J1), the target onset and offset (T3, T4, J3, J4), and the opening movement offset during the final vowel (T6 and J6) as shown in Fig. 1. In order to abstract from individual vocal-tract size differences z scores were calculated for all positional data. For computing the z scores, speaker-specific means and standard deviations of the jaw and the intrinsic tongue tip movement signals were calculated for the stretches when the subjects actually spoke. The means pooled for all trials were subtracted from measurement points and then divided by the standard deviation.

Cohesion strength was assessed by calculating statistics for the variability of temporal and spatial parameters. As measure for the variability simple standard deviations were used instead of the coefficients of variation; this is independent of the magnitude of the mean and might therefore be more appropriate. However, since z scores and latencies varied around zero, coefficients of variation could not be calculated.

E. Data exclusion

For the computation of latencies some data had to be excluded because, for the opening gesture towards the second /a/, no jaw downward movement could be detected and/or because the highest jaw position was sometimes not achieved during the consonant but during the following unstressed vowel. At normal vocal effort level exclusion was necessary for 19% of all /d/, 12% of the /n/ realizations, and 27% of the /l/ (11, 7, and 16 items, respectively). The only sound for which some items had to be excluded at loud speech was the voiced stop; the four instances were all produced by speaker SR.

TABLE I. Manner (MN) and vocal effort (VE) effects on the jaw (left) and intrinsic tongue tip (right) positions during the first vowel (V1), the consonant (C) and the second vowel (V2) and on the closing and opening displacements. Statistics are based on repeated measures ANOVAs with Greenhouse-Geisser ϵ corrected degrees of freedom in brackets. Results of pairwise t -tests with Bonferroni adjustments are also given with $<$ indicating lower positions or smaller displacements.

| | | Jaw | | | Intrinsic tongue tip | | | |
|-------|------|---------|--|----------------------|----------------------|---|----------------|-----------------|
| | | MN | VE | MN:VE | MN | VE | MN:VE | |
| Pos. | V1 | $F(df)$ | 1.75 (2.1, 8.6) | 13.05 (1, 4)* N>L | 3.61 (2.9, 11.5)* | 0.22 (1.6, 6.4) | 3.6 (1, 4) | 0.78 (2.3, 9.4) |
| | C | $F(df)$ | 26.9 (1.8, 7.2)*** $s_f > dn_l, t_f > nl$ | 7.74 (1, 4)* | 3.63 (1.3, 5.2) | 16.96 (2.1, 8.4)** $s < f_{tdnl}, t_f > 1$ | 77.7 (1, 4)*** | 3.2 (2.5, 9.9) |
| | V2 | $F(df)$ | 7.19 (2.8, 11.4)** | 12.41 (1, 4)* N>L | 0.87 (1.6, 6.4) | 1.66 (1.9, 7.5) | 1.0 (1, 4) | 1.59 (2.5, 9.8) |
| Disp. | Clos | $F(df)$ | 23.7 (2.4, 9.6)*** $s_f > nl, td > 1$ | 14.4 (1, 4)* N<L | 4.95 (1.5, 5.9) | 13.5 (1.9, 7.7)*** $s < f_{tdn} < 1, f < dn$ | 0.74 (1, 4) | 0.74 (1.5, 6.0) |
| | Op | $F(df)$ | 32.6 (1.6, 6.4)*** $f_{std} > 1, s_f > n$ | 15.1 (1, 4)* | 4.0 (1.8, 7.4) | 25.4 (2.1, 8.4) $s < f_{tdn} < 1, f < dn$ | 6.89 (1, 4) | 3.39 (1.7, 6.7) |

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

These numbers of exclusion and their specific distribution, i.e., only voiced consonants, can be interpreted in terms of an obligatory closed-jaw position for the voiceless obstruents on the one hand and either a transitory lower jaw target (no turning point) or no obligatory jaw target at all for the voiced coronals on the other hand. The latter assumption does not seem to hold if the data for loud vocal effort are taken into account for which exclusion was restricted to four voiced stops, all produced by one speaker. Since the timing for the closing movement might already be affected if no consonant-related jaw target is reached, these items were excluded for all latencies. For the latencies of the peak velocities items were excluded when double-velocity peaks of equal height occurred in either the jaw or the intrinsic tongue tip tangential velocity signal. For the closing gesture a double-velocity peak occurred for 14 cases (1.9%) and for the opening gesture for 26 cases (3.6%).

F. Statistics

Analyses of variance were calculated for individual speakers and pooled over all speakers using the script language R (R Development Core Team, 2005). For the individual speakers all valid data were included. Main effects and interactions were computed. Independent variables were manner of articulation (MN) and vocal effort level (VE).

In order to evaluate speaker-independent strategies, additionally ANOVAs pooled over all speakers were calculated based on the data averaged over up to 12 repetitions so that each speaker contributed only one experimental score per condition (see, e.g., Max and Onghena, 1999). This data reduction is necessary in order to avoid artificially inflating the error terms and degrees of freedom. Whether manner of articulation and vocal effort affected positional and temporal data was evaluated by calculating repeated-measures ANOVAs with the within-subject factors MN and VE. Degrees of freedom were corrected by calculating the Greenhouse-Geisser epsilon in order to avoid violation of the sphericity

assumption. Therefore, fractional degrees of freedom are often given in the tables. Pairwise t -tests with Bonferroni adjustments for multiple comparisons were carried out for individual statistics and for the repeated-measure ANOVAs in order to assess significant differences between the six-level-factor MN.

III. RESULTS

A. Positions and movement amplitudes

First, systematic effects of manner of articulation and vocal effort on spatial parameters were evaluated by calculating repeated-measures ANOVAs with manner and vocal effort as repeated factors. Subject means of jaw and intrinsic tongue tip positions during the first vowel, the consonant, and the second vowel and of the closing and opening amplitudes served as dependent variables (see Table I). Figure 2 shows the displacements of the jaw (left) and the intrinsic tongue tip (right) during the closing (upper panels) and the opening movements (lower panels). All data are z -transformed, i.e., scaled in standard deviations. The height of the white bars indicates the magnitude of the movement amplitudes for the normal condition and gray bars for the loud condition. The zero line for the bars in the upper figures specifies the articulator position in the normal effort condition during the initial /a/ for the closing movement (see T1 and J1 in Fig. 1) and the final /a/ for the opening movements in the lower figures (see also T6 and J6 in Fig. 1). The tops of the bars correspond to the maximal excursions of the articulators during the consonant.

Jaw positions during the consonant were significantly affected by manner but only slightly by vocal effort (see the tops of the bars in panels on the left side). For the jaw positions during the vowels, depicted as the lower edges of the bars, the opposite was the case: for the vowels the jaw position was significantly lower for loud speech but manner of articulation affected the jaw position only slightly, reaching significance only for the second vowel. Therefore, jaw dis-

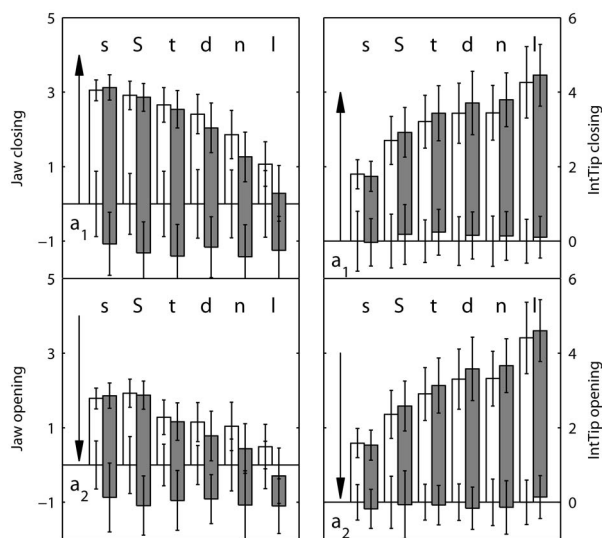


FIG. 2. Positions and displacements of the jaw (left) and the intrinsic tongue tip (right) during the closing (upper panels) and the opening movements (lower panels). All data are z -transformed, i.e., scaled in standard deviations. The height of the white bars indicates the magnitude of the amplitudes for the normal condition and gray bars for the loud condition. The zero line in the upper panels corresponds to the articulator position in the normal effort condition during the initial /a/ for the closing movement and the final /a/ for the opening movements in the lower figures. The tops of the bars indicate the position of the articulators during the consonantal target phase.

placements, shown in Fig. 2 as the height of the bars, increased significantly for loud speech mainly because of the lower jaw positions during the vowels.

Manner effects on jaw positions during the consonant were tested by pairwise t -tests using the Bonferroni adjustments for multiple comparisons; this yielded significantly higher jaw positions for the sibilants as compared to /d, l, n/. The voiceless stop /t/ was produced with a jaw position between the sibilants (no significant difference) and the voiced stop /d/. The lateral was realized with the greatest jaw opening. The effect of vocal effort increased with decreasing consonant-specific position, i.e., more extensive jaw lowering in loud speech was found for consonants with an already open jaw position. However, even in the case of the lateral the t -test did not indicate a significant vocal effort effect. Looking at individuals, four of the speakers had a significantly lower jaw position in the loud condition during the nasal and two speakers during the lateral. Only one speaker produced /t/ and /d/ with a significantly lower position, and no speaker varied the jaw position during the sibilants over the two vocal effort conditions.

As can also be seen in Fig. 2, jaw opening movements towards the second vowel were smaller than the closing movements. This can be attributed to the fact the initial vowel received the main stress and second vowel was unstressed but unreduced, i.e., not a schwa vowel.

The intrinsic tongue tip position during the consonant varied with manner of articulation, with significantly lowest position and smallest amplitudes for the sibilant /s/. /ʃ/ was produced with a significantly higher tongue tip position, which can be attributed to its more retracted place of articulation as compared to /s/. For the production of the lateral the tongue tip had to move more than for the other consonants

(not significant for /l/ vs /d, n/). Thus, the pattern of movement amplitudes for manner of articulation is inversely related for the jaw and the intrinsic tongue tip: the more the jaw moves the smaller is the intrinsic tongue tip movement (see the sibilant /s/) and vice versa for the lateral. Intrinsic tongue tip positions during vowels were affected neither by vocal effort nor by manner of articulation. However, repeated-measures ANOVAs revealed a significant main effect of vocal effort on the intrinsic tongue tip position during the consonant. Since the t -tests pooled for all consonants and split by consonant did not reach significance, and the difference between normal and loud speech was very small (loud speech: mean 1.68, s.d. 1.04; normal speech: mean 1.5, s.d. 0.97) this result will be neglected.

To summarize the results in the spatial domain, manner of articulation affected both the tongue tip and the jaw positions during the consonant, whereas for the vowels only the jaw positions for the final unstressed vowels showed some significant effects for manner variations. Generally, the results from the literature are confirmed: the sibilants and the voiceless stop are produced with a closer jaw position compared to the remaining consonants under consideration here. Vocal effort increase was accompanied by a significantly more open jaw position during the vowels, while for the consonants effects were smaller and less consistent (significant only for the nasal (four speakers) and the lateral (two speakers)). The intrinsic tongue tip positions during the vowels and consonants remained unaffected by vocal effort changes.

B. Temporal coordination

Temporal interarticulator coordination between the tongue tip and the jaw was assessed by analyzing the latencies between the two articulators as well as between the jaw and the acoustically defined landmarks. In the first part, manner and effort effects on the latencies during the consonant target phase are discussed and then in the second part the latencies of the closing and opening velocity peaks. The aim of this section is to determine whether differences in the spatial extent of jaw involvement are accompanied by differences in the temporal coordination close to the constriction phase of the consonants.

As was pointed out in Sec. II D the significantly longer acoustic durations of the fricatives influence the latencies. Therefore, target on- and offsets were normalized to the acoustic consonant durations individually. Results are shown in Fig. 3 with 0 and 1 denoting the acoustically defined begin and end of the consonant respectively (see A1 and A2 in Fig. 1). Unfilled bars show the target duration and relative timing of the intrinsic tongue tip target achievement and release (T3 and T4 in Fig. 1), gray bars the relative hold durations of the jaw (J3 and J4 in Fig. 1). If the lower border of the white bar is close to zero, then the acoustic onset of this consonant (A1) is at the same time as the onset of the intrinsic tongue tip target phase (T3). The height of the lower white bar shows the normalized latency of the target onset (Lat_{on}), i.e., how much later the jaw achieves the target as compared to the tongue tip. For example, the lower white bar for /t/ of

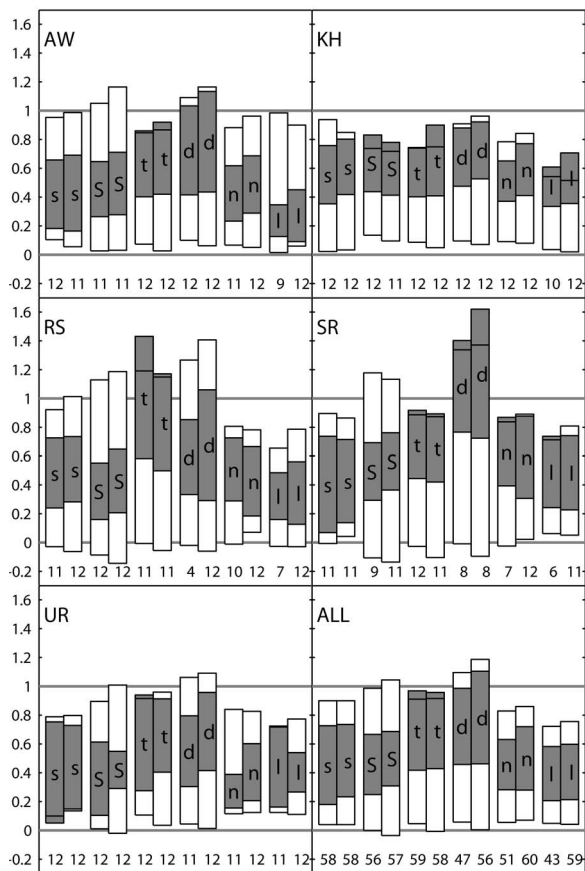


FIG. 3. Normalized durations of intrinsic tongue tip (unfilled boxes) and jaw target (gray boxes) on- and offsets for normal (left) and loud speech (right). Zero denotes the acoustically defined onset of the consonant and 1 the offset with number of measured items.

speaker AW is longer than for /l/, which corresponds to a longer positive onset latency for /t/ than for /l/ and therefore to a later jaw target achievement for /t/ than for /l/. In contrast, in only one case, /s/ of speaker UR, no lower white bar but a gray bar is shown, which indicates a negative onset latency and that the jaw achieves its target earlier than the tongue tip. Accordingly, the height of the upper white bar gives the normalized latency of the onset of the opening movement (Lat_{off}), i.e., the interval during which the jaw has already started the opening movement and the tongue tip still maintains the target position, indicating a negative offset latency (see, e.g., /s/ for all speakers). If no upper white bar is shown and the gray bar is overlapping the upper edge of the white bar, as, e.g., for /t/ of speaker RS, then the tongue tip starts the opening movement before the jaw, corresponding to a positive offset latency.

The results of ANOVAS for individual speakers are given in Table II, with manner and VE (vocal effort) as independent variables and the normalized latency between tongue tip and jaw movement at the target onset (Lat_{on}) and at the target offset (Lat_{off}) as dependent variables. Positive values indicate that the jaw moves later and negative values that it moves earlier than the intrinsic tongue tip. Furthermore, the differences between the acoustic onset and offset of the consonants and the jaw target onset and offset ($AcJaw_{on}$ and $AcJaw_{off}$) were also analyzed in order to check

whether the specified events of jaw movements are coordinated with acoustically defined landmarks such as the on- and offset of nasality or the burst noise.

Generally, on- and offsets of the intrinsic tongue tip tend to vary with the acoustically defined on- and offsets, shown in Fig. 3 as the close proximity of the upper and lower edges of the white bars with the horizontal line at 0 indicating the acoustical onset and the line at 1 indicating the acoustical offset, respectively. There are also some less well-aligned examples: For the postalveolar sibilant the intrinsic tongue tip hold interval was longer than the acoustically defined consonant (shown by the longer white bars overlapping over the two long horizontal lines at 0 and 1 in Fig. 3, especially clear for speakers RS and SR), which can be attributed to the fact that the tongue tip sensor is probably placed in front of the relevant articulator for the postalveolar and therefore does not capture all parts of the relevant movement for the constriction. Only speaker KH, whose articulation of /j/ was much more fronted, had a shorter plateau for the postalveolar fricative compared to the acoustic duration. The intrinsic tongue tip plateau also frequently exceeded the acoustically defined consonant offset for the voiced stop (see especially speakers RS and SR). This can be attributed to the difficulties in labeling the end of the plateau, since during /d/ most speakers showed a high amount of lingual forward and downward movement, probably due to its more apical articulation as reported in the Introduction.

The consonantal target was generally reached first with the tongue tip and then with the jaw, indicated by the lower white boxes in Fig. 3. This short time lag might be attributable to specific kinematic properties of the involved articulators, such as generally slower jaw movements as mentioned above in the Introduction. Only for the alveolar fricative of speaker UR did the jaw reach the target before the tongue tip.

Manner of articulation had significant effects on the timing between the tongue tip and the jaw for several articulatory events. For the stops the onset latency (Lat_{on}) was longer as compared to the other consonants, which implies a longer jaw delay as shown by the longer lower white bars in Fig. 3. This was significant for four speakers and for the tests pooled over all speakers. Furthermore, the onset of the jaw opening movement was consistently somewhat later than or happened at the same time as the onset of the tongue tip opening movement (no upper white bars) for the voiceless stop, whereas the jaw started its opening movement only rarely before the tongue tip for the other consonants (upper white bars). Relative to the acoustic onset of the consonant the jaw reached its target latest for both stops, as shown for the variable $AcJaw_{on}$ (only /d/ for speaker KH) and started its opening movement latest ($AcJaw_{off}$). Therefore, the gray boxes in Fig. 3 are shifted towards the upper end of the white boxes for the stops. For the other consonants the jaw target interval tends to be centered in the middle of the tongue tip target interval with positive onset latencies, i.e., the jaw reaches its maximum later than tongue tip, and negative offset latencies, i.e., the jaw starts its downward movement before the tongue tip.

Comparing the two sibilants /s/ and /ʃ/, the jaw target

TABLE II. Manner (MN) and vocal effort (VE) effects on onset and offset latencies between the tongue tip and the jaw (Lat_{on} and Lat_{off}) as well as between the acoustically defined onset and the jaw target achievement ($AcJaw_{on}$) and the acoustically defined offset and the jaw target offset ($AcJaw_{off}$). Statistics are based on two-way ANOVAs for individual speakers. Results of pairwise t -tests with Bonferroni adjustments are also given with $>$ indicating a later jaw timing.

| Speaker | Effect | Df | Lat_{on} | Lat_{off} | $AcJaw_{on}$ | $AcJaw_{off}$ |
|---------|--------|--------------------------|--|-----------------------|--|---|
| AW | MN | 5 | 32.41 ^{***} | 31.38 ^{***} | 44.48 ^{***} | 54.29 ^{***} |
| | VE | 1 | 2.50 | 0.90 | 0.41 | 7.85 ^{**} |
| | MN:VE | 5 | 1.61 | 1.05 | 0.68 | 0.17 |
| KH | | 125 | $td > fn > sl$ | $td > nsf1$ | $td > fn > sl$ | $d > t > sfn > l$ |
| | MN | 5 | 2.27 | 7.56 ^{***} | 4.33 ^{**} | 10.71 ^{***} |
| | VE | 1 | 4.26 [*] | 5.55 [*] | 1.82 | 11.10 ^{**} |
| | MN:VE | 5 | 0.14 | 1.07 | 0.43 | 2.07 |
| RS | | 129 | $L > N$ | $tf1 > snd$ | $d > snl$ | $d > snl, t > l$ |
| | MN | 5 | 17.83 ^{***} | 22.61 ^{***} | 24.05 ^{***} | 44.19 ^{***} |
| | VE | 1 | 0.15 | 1.73 | 1.17 | 0.03 |
| | MN:VE | 5 | 2.39 [*] | 0.91 | 1.24 | 2.98 [*] |
| SR | | 114 | $N: t > nsf1$ $L: t > nl, d > n$ | $t > nlts > f$ | $t > dsnf1, d > l$ | $N: t > dsnf1, d > l$ $L: td > snf1$ |
| | MN | 5 | 27.00 ^{***} | 27.54 ^{***} | 29.08 ^{***} | 50.43 ^{***} |
| | VE | 1 | 0.20 | 0.65 | 0.01 | 1.15 |
| | MN:VE | 5 | 0.69 | 1.62 | 0.66 | 1.20 |
| UR | | 105 | $d > tf > sln$ | $td > s, tdnl > f$ | $d > tnfls, t > ls$ | $d > tnfls, t > fs$ |
| | MN | 5 | 21.94 ^{***} | 16.08 ^{**} | 14.30 ^{***} | 21.15 ^{**} |
| | VE | 1 | 32.20 ^{***} | 0.87 | 27.88 ^{***} | 0.16 |
| | MN:VE | 5 | 1.59 | 5.69 ^{***} | 0.70 | 3.94 ^{**} |
| | 129 | $td > nls, f > ns L > N$ | $N: tsl > n, tl > f$ $L: ts > fnl, d > f$ | $dt > lfn$ $L > N$ | $N: t > lfn, ds > n$ $L: dt > snf1, s > fl$ | |

* $p < 0.05$.
 ** $p < 0.01$.
 *** $p < 0.001$.

achievement was significantly earlier for /s/ than for /j/ for three speakers. The jaw also tended to start the opening movement earlier for /j/ but this was significant only for two speakers.

Vocal effort affected the tongue-tip jaw coordination only inconsistently. As can be seen in Table II, only for speakers KH and UR did vocal effort increase have a significant main effect with a later jaw target achievement for both speakers and a later jaw target offset for speaker UR (see lower with boxes in Fig. 3). Latencies pooled over all speakers (Table III), however, did not show significantly different means for vocal effort increase. Interactions between the two factors manner and vocal effort were significant in some cases, e.g., for speaker RS the jaw started its downward

movement early for /t/ and later for /j/ when spoken loudly. However, no consistent pattern of vocal effort effects on the timing parameters during the consonant could be found.

Time points further away from the consonant target phases were also affected by manner and vocal effort variations as shown in Table IV. The velocity peak latency of the closing gesture ($LatV_{cl}$) indicated no significant manner-dependent variation when all speakers were pooled. For individual speakers the timing of the closing velocity peaks did vary with manner of articulation, but no consistent pattern could be found. Most consonants were produced with a slight jaw advancement, apart from speakers KH and UR for /l/ and speaker RS for /j/, who showed a later closing velocity peak for the jaw. Vocal effort affected $LatV_{cl}$ of /s/ for

TABLE III. Manner (MN) and vocal effort (VE) effect on onset (left) and offset (right) latencies between the tongue tip and the jaw (Lat) and acoustically defined landmarks and the jaw ($AcJaw$). Statistics are based on repeated measures ANOVAs with Greenhouse-Geisser ϵ corrected degrees of freedom in brackets. Results of pairwise t -tests with Bonferroni adjustments are also given $>$ indicating later jaw events.

| | | Onset | | | Offset | | |
|-------|---------|--|-------------|-----------------|--|---------------------------|-----------------|
| | | MN | VE | MN:VE | MN | VE | MN:VE |
| Lat | $F(df)$ | 6.65 (1.7, 6.9) [*] $df > ls, d > n$ | 2.74 (1, 4) | 2.46 (1.8, 7.1) | 3.90 (3.3, 13.0) [*] $t > f$ | 0.27 (1, 4) | 0.6 (2.3, 9.3) |
| AcJaw | $F(df)$ | 7.21 (1.9, 7.7) [*] $d > nfls, t > ls$ | 0.9 (1, 4) | 1.66 (1.8, 7.2) | 7.83 (2.3, 9.1) ^{**} $td > snf1$ | 10.01 (1, 4) [*] | 1.54 (2.0, 8.1) |

* $p < 0.05$.
 ** $p < 0.01$.
 *** $p < 0.001$.

TABLE IV. Manner (MN) and vocal effort (VE) effects on latencies of the closing (left: $\text{Lat}V_{\text{cl}}$) and opening (right: $\text{Lat}V_{\text{op}}$) velocity peaks. Statistics for individual speakers are based on two-way ANOVAS and the statistics pooled overall speakers are based on repeated measures ANOVAS with Greenhouse-Geisser ϵ corrected degrees of freedom in brackets. Results of pairwise t -tests with Bonferroni adjustments are also given with $>$ indicating later jaw events.

| | | $\text{Lat}V_{\text{cl}}$ | | | $\text{Lat}V_{\text{op}}$ | | |
|-----|----------------|---|-------------------------------------|--|--|--------------------------------------|--|
| | | MN | VE | MN:VE | MN | VE | MN:VE |
| AW | $F(\text{df})$ | 2.24(5, 123) | 0.07 (1, 123) | 1.78 (5, 123) | 44.3 (5, 121) ^{***} d > tnl > sj | 4.56 (1, 121) [*] | 1.62 (5, 121) t: N < L |
| KH | $F(\text{df})$ | 10.6 (5, 127) ^{***} ls > f, l > ndt | 2.72 (1, 127) | 2.74 (5, 127) [*] s: N > L | 11.6 (5, 121) ^{***} l > td, ln > fs | 3.33 (1, 121) | 1.19 (5, 121) |
| RS | $F(\text{df})$ | 4.97 (5, 106) ^{***} f > sld | 6.44 (1, 106) [*] N > L | 1.83 (5, 106) s: N > L | 47.0 (5, 112) ^{***} tndls > f, dnt > s | 5.56 (1, 112) [*] N > L | 5.82 (5, 112) ^{***} t,d: N < L |
| SR | $F(\text{df})$ | 46.42 (5, 105) ^{***} ftndl > s, f > l | 6.60 (1, 105) [*] | 0.63 (5, 105) n: N > L | 29.7 (5, 102) ^{***} lnd > st > f | 0.87 (1, 102) | 2.38 (5, 102) [*] |
| UR | $F(\text{df})$ | 7.34 (5, 127) ^{***} tsdfn > l | 1.42 (1, 127) | 1.09 (5, 127) s:N > L | 14.7 (5, 120) ^{***} sldnt > f | 8.71 (1, 120) ^{**} N > L | 1.89 (5, 120) t: N < L |
| All | $F(\text{df})$ | 0.85 (1.5, 5.8) | 10.12 (1, 4) [*] | 0.85 (2.2, 8.6) | 8.2 (2.6, 10.4) ^{**} nldst > f | 19.51 (1, 4) [*] | 2.2 (1.6, 6.3) t: N < L |

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

three speakers with an earlier jaw velocity peak in loud speech. The latency of the opening velocity peaks was negative and longer for /ʃ/ (for speakers AW and KH also for /s/) than for the other consonants, i.e., the jaw velocity peak occurred earlier than the intrinsic tongue tip velocity peak for the sibilants. Another tendency was that in loud speech the jaw velocity peak occurred later after the voiceless stop, which was significant for three speakers and when pooled over all speakers.

Summarizing the results for manner effects, it was found that two differential patterns emerge for the six analyzed coronal consonants: The first pattern looks more symmetrical, with a jaw target achievement occurring shortly after the tongue tip target achievement and the jaw target offset before the offset of the tongue tip. Therefore, the jaw hold interval is always shorter than the tongue interval and lies approximately in the middle of the tongue tip target interval. As can be seen in Fig. 3, for most speakers the sibilants and the sonorants are produced with this timing pattern. The second pattern is asymmetrical in the sense that the jaw target is achieved in the second half of the tongue tip target phase and the jaw also starts its opening movement later than the tongue tip. This pattern emerges for the two stops (see also Fig. 3).

For vocal effort, due to a high amount of speaker-dependent variability, only some tendencies could be observed: The consonants with the least spatial variability as well as the highest jaw position, i.e., /s, ʃ, t/, showed very inconsistent differences in the onset and offset of the jaw target. Significant timing differences, reported in Table IV, were always in the direction to accommodate both requirements: a lower jaw for loudly produced vowels and a high jaw target for segmental needs. For example, for the alveolar fricative /s/ the latency of the velocity peaks of the closing movement slightly decreased (as shown in Table IV, column MN:VE for $\text{Lat}V_{\text{cl}}$ significant for speakers KH, RS, and UR), i.e., the jaw reached its peak velocity somewhat earlier than

the tongue tip in normal condition (mean = -10 ms) and with a more pronounced advancement in loud speech (mean = -18 ms). This seems to be a possible strategy in order to anticipate an early onset of the jaw target, whereas for /ʃ/ and /t/ the jaw target onset did not matter as much. For the voiceless stop a high jaw position has to be maintained until the burst. Because of the required lower jaw position for the following vowel in loud speech the necessary jaw displacement increases, which in the case of /t/ also leads to a later jaw opening velocity peak (mean for all speakers for the normal condition: -9 ms vs loud: 6 ms). As shown in Table IV, this is significant for speakers AW, RS, and UR as well as the pooled speakers.

C. Spatial and temporal variability

In order to assess the interarticulator cohesion, the standard deviations of temporal parameters were compared for the analyzed consonants assuming that temporally more precisely articulated consonants exhibit a stronger interarticulator cohesion. Standard deviations of intrinsic tongue tip and jaw positions during the consonant were also analyzed in order to compare the relevance of the two articulators for the coronal consonants. Results of repeated-measures ANOVAS and pairwise t -tests are given in Table V. Concerning the spatial variability, the jaw positions clearly varied less during the voiceless obstruents /s, ʃ, t/. A generally higher variability for loud speech was observed for the jaw positions, which is contrary to the results of Dromey and Ramig (1998). The accuracy of intrinsic tongue tip position during the consonant varied neither with manner of articulation nor with vocal effort.

For the offset latencies Lat_{off} and $\text{AcJAW}_{\text{off}}$, significantly smaller variances were obtained for the sibilants as compared to the lateral and the voiced stop. The velocity peak latency of the closing movement was significantly more variable for /ʃ/ as compared to the other consonants.

TABLE V. Manner (MN) and vocal effort (VE) effects on standard deviations of positions (upper part) and onset offset latencies between the tongue tip and the jaw (Lat), the acoustically defined landmarks and the jaw (AcJaw) and the peak velocities (LatV). Statistics are based on reported measure ANOVAs with Greenhouse-Geisser ϵ corrected degrees of freedom in brackets. Results of pairwise t -tests with Bonferroni adjustments are also given with $>$ indicating more variability.

| | | Jaw | | | Intrinsic-tongue-tip | | |
|-------|--------|------------------------------|-----------------------|-----------------|---|-------------|-----------------|
| | | MN | VE | MN:VE | MN | VE | MN:VE |
| Pos | F (df) | 7.16 (2.2, 8.9)* ln > tsf | 8.93 (1, 4)* L > N | 2.59 (2.1, 8.3) | 1.11 (2.4, 9.5) | 1.1 (1, 4) | 0.88 (1.8, 7.2) |
| | | Onset | | | Offset | | |
| | | MN | VE | MN:VE | MN | VE | MN:VE |
| Lat | F (df) | 1.42 (1.5, 6.1) | 0.01 (1, 4) | 0.96 (1.9, 7.7) | 4.8 (2.1, 8.5)* ld > fs | 0.05 (1, 4) | 0.37 (1.8, 7.4) |
| AcJaw | F (df) | 1.56 (1.6, 6.2) | 0.44 (1, 4) | 1.26 (2.0, 8.1) | 12.0 (2.4, 9.6)** dlt > sf, n > f | 0.18 (1, 4) | 0.23 (1.6, 6.6) |
| LatV | F (df) | 10.3 (1.6, 6.5)* f > sntd | 1.61 (1, 4) | 1.04 (1.9, 7.5) | 1.0 (2.9, 11.5) | 1.29 (1, 4) | 1.03 (2.2, 8.8) |

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

In conclusion, during the consonants significantly less spatial and temporal variability was found for the sibilants, whereas for the voiceless stop only the spatial jaw positioning was more precise than for /n, l/ but not the timing.

IV. SUMMARY AND DISCUSSION

This study investigated the timing between the tongue tip and the jaw of the German coronal consonants /s, ʃ, t, d, n, l/ in a low vowel context. We hypothesized that the contribution of the jaw differs for various manners of articulation, and that this is reflected in the strength of cohesion between the tongue tip and the jaw. The strength of the cohesion was investigated by analyzing the variability of temporal and spatial parameters. Vocal effort changes were introduced as a control condition. The assumption was that consonants with a tight temporal and spatial coupling between the tongue tip and the jaw should not change due to vocal effort variation. We obtained the following results:

- (1) With respect to tongue-jaw coordination, two different patterns emerged: a symmetrical one with later jaw target onsets and earlier offsets within the tongue target phase, and an asymmetrical pattern with a late jaw target onset and an offset which occurs approximately simultaneously with the tongue tip target offset. The sibilants /s, ʃ/ were produced predominantly with a symmetrical pattern and the voiceless plosive with an asymmetrical one. For the remaining coronal consonants /d, l, n/ a preference for a symmetrical pattern could be observed but with a high amount of speaker-dependent variability.
- (2) Increasing vocal effort had very little consistent effects on the analyzed timing parameters during the consonants. However, the timing of velocity peaks was affected for two consonants: the jaw-closing velocity peak

was advanced for /s/ in loud speech and the jaw-opening velocity peak was delayed for /t/ in loud speech.

- (3) Sibilants were produced with the least amount of spatial and temporal variability. The voiceless stop showed reduced variability only in the spatial domain. The remaining consonants /d, n, l/ were generally more available and the sonorants also more frequently affected by the vocal effort condition.

These results will be discussed in terms of lingual-mandibular coordination for producing different manners of articulation, and implications for speech motor control.

A. The role of the coordination between the tongue tip and the jaw

The most striking result of this study is the very consistent difference in lingual-mandibular timing between the voiceless stop on the one hand and the sibilants on the other hand. For the remaining consonants, which were also produced with a lower and more variable jaw position, speakers varied in their preferred timing pattern. This confirms the results from previous studies that the jaw does not contribute uniformly to the achievement of consonantal coronal constrictions. In this section various explanations for the emergence of these two timing patterns will be discussed.

As was suggested in the Introduction, a late jaw target for the voiceless stop is produced in order to achieve a salient burst. In Mooshammer *et al.* (2003) we argued that the explosion noise might be enhanced by an obstacle noise source, namely the lower teeth. Among stops it is only the alveolar in which the lower teeth are immediately downstream of the place of articulation, and indeed current evidence suggests that bilabial or velar stops are generally produced with lower and more variable jaw positions (see, e.g., Lee, 1996; Hoole and Kühnert, 1996). This assumption is

supported by the high and almost invariant jaw position which was found in most studies for /t/. Furthermore, a late jaw target, starting the opening movement somewhat later than the tongue tip, was not obligatory for the voiced stops, which were fully voiced for all speakers in the current study. As can be seen in Fig. 3, only speaker SR exhibits an asymmetrical timing pattern with a late jaw target in this case. Finally, in Geumann (2001a) it was found that for three speakers the jaw was significantly lower during the voiced stop compared to the voiceless, which can be attributed to an accommodation to the jaw targets of the neighboring vowels (see also the higher contextual variability for /d/ as found in Geumann, 2001a). This option does not exist for the voiceless stop since a prominent burst is required. It is an interesting hypothesis that the well-known fact that the voiced stop is produced with a weaker and less audible burst (Ladefoged and Maddieson, 1996) might be partly attributable to the lower jaw position and the earlier jaw opening movement onset.

An alternative or additional factor might be that the asymmetrical pattern of the tongue-jaw coordination is a consequence of the target planning: Fuchs *et al.* (2001, in press) and Löfqvist and Gracco (2002) hypothesized that for stops the articulator aims at reaching a target planned above the constriction location (palate or upper lip) for ensuring a rapid pressure buildup. Furthermore, a target above the palate also has the advantage that no precise positioning of the tongue or lower lip is required for the stop as opposed to fricatives, for example. Hence, when the tongue tip crashes into the palate before reaching its target, the jaw might still continue to move upwards to achieve its planned goal. Evidence for these considerations can be found by the late jaw target achievement compared to the fricatives and the sonorants, but palate impact alone cannot explain why the jaw opening movement is timed with the burst and why a high and less variable jaw position seems to be an obligatory characteristic of the voiceless alveolar stop.

B. Implications for speech motor control

Generally, our results confirm that coordinative structures orchestrate individual articulators in a task-specific and flexible manner in order to reduce the degrees of freedom. On the one hand the same executing organs, tongue tip and jaw, can act together and—by combining a variety of spatial and temporal patterns—create distinctive sounds. The tongue tip and jaw are temporally highly fine-tuned but not necessarily moving in synchrony, as was shown here for the voiceless stop. Therefore, the analyzed articulators can move quite independently of each other and recombine in a flexible manner. On the other hand the temporal patterns during the consonantal target regions were quite stable across two vocal effort conditions, which speaks for a high degree of cohesion for producing coronal consonants.

It was proposed by Gracco (1988) and Hertrich and Ackermann (2000) that opening movements are produced with a lesser degree of cohesion than closing movements (see the Introduction). In contrast, we found that the onset of the opening movement and the relative timing of the contribut-

ing articulators at this time point might even be crucial for distinguishing different sounds. Even though the analyzed consonants are not solely distinguished on the basis of the two observed movement patterns, symmetrical and asymmetrical, the latter pattern seems to aim at the production of a prominent burst and therefore provides important cues for the contrast between the voiced and the voiceless stops (by additionally raising the jaw less for the former) and the contrast between different places of articulation. Besides implications for the kind of control underlying the observed kinematics, this result also supports the assumption that intergestural cohesion is stronger within segments than in between (see, e.g., Saltzman *et al.*, 1998), and that—as suggested by the perturbation experiment of Gomi *et al.* (2002)—the cohesion is stronger during the achievement of the goal than further away. In the current study consistent temporal adjustments due to changes in the condition, namely vocal effort, were mainly found at the peak velocities, i.e., during the transition between successive targets (see also delayed jaw peak velocity after /t/ in loud speech). The timing of the consonantal target onsets and offsets, however, was not consistently affected by vocal effort changes.

As was already discussed, the jaw's task varies significantly for the analyzed consonants. Even though many studies emphasize that the jaw is more sluggish due its heaviness than other articulators, the speaker seems nevertheless capable of controlling jaw positions and its movement course in a very exact way. As was found by Lindblom and Lubker (1985), subjects can judge the amount of jaw movement more accurately than their tongue movements. The authors argue that—besides the perceptual distinctiveness—the higher awareness of the speaker for jaw positioning might play an important role for the tendency to favor contrasts along the dimension of opening for the composition of vowel systems. Our study suggests that the speaker's higher awareness of jaw positions is probably exploited for the distinction of consonants to a greater degree than assumed in earlier studies.

ACKNOWLEDGMENTS

This work was partially supported by the German Research Council (DFG) Ti69/31 and GWZ 4/8-1, P.1. We also thank Jonathan Harrington, Barbara Kühnert, Christian Geng, and Susanne Fuchs for very valuable comments on earlier drafts of this paper. Peter Dalgaard and Ernst Dombrowski gave invaluable advice regarding statistics and solutions in R. Furthermore, this work was inspired and initiated by very fruitful discussions of the first author with Peter Alfonso and Pascal van Lieshout during her postdoc term in Knoxville, 2001, supported by a DAAD grant. We also want to express our gratitude to the two anonymous reviewers and the editor, Anders Löfqvist, for their insightful comments and suggestions.

¹For reasons of consistency the intrinsic tongue tip signal was used for labeling for all consonants, even though there is some variation in place of articulation: the sibilant /ʃ/ is usually more retracted than the alveolar consonants.

- Dart, S. (1991). "Articulatory and acoustic properties of apical and laminal articulations." *UCLA Working Papers in Phonetics* **79**, 1–155.
- De Nil, L., and Abbs, J. (1991). "Influence of speaking rate on the upper lip, lower lip, and jaw peak velocity sequencing during bilabial closing movements." *J. Acoust. Soc. Am.* **89**, 845–849.
- Dromey, C., and Ramig, L. O. (1998). "Intentional changes in sound pressure level and rate: Their impact on measures of respiration, phonation, and articulation." *J. Speech Lang. Hear. Res.* **41**, 1003–1018.
- Edwards, J. (1985). "Contextual effects on lingual-mandibular coordination." *J. Acoust. Soc. Am.* **78**, 944–948.
- Fowler, C., Rubin, P., Remez, R., and Turvey, M. T. (1980). "Implications for speech production of a general theory of action," in *Language Production, Volume 1: Speech and Talk*, edited by B. Butterworth (Academic, London), pp. 373–420.
- Fuchs, S., Perrier, P., and Mooshammer, C. (2001). "The role of the palate in tongue kinematics: An experimental assessment in VC sequences from EPG and EMMA data." *Proceedings Eurospeech, Aalborg, Denmark* **3**, pp. 1487–1490.
- Fuchs, S., Perrier, P., Geng, C., and Mooshammer, C. (2006). "What role does the palate play in speech motor control? Insights from tongue kinematics for German alveolar obstruents," in *Towards a Better Understanding of Speech Production Processes*, edited by J. Harrington and M. Tabain (Psychology Press, New York), pp. 149–164.
- Geumann, A. (2001a). "Invariance and variability in articulation and acoustics of natural perturbed speech," *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München*, **38**, pp. 265–393.
- Geumann, A. (2001b). "Vocal intensity: Acoustic and articulatory correlates," in *Proceedings of the 4th International Speech Motor Conference*, edited by B. Maassen, W. Hulstijn, R. Kent, H. Peters, and P. van Lieshout (Uitgeverij Vantilt, Nijmegen), pp. 70–73.
- Geumann, A., Kroos, C., and Tillmann, H. G. (1999). "Are there compensatory effects in natural speech?" *Proceedings of the 14th International Congress Phonetic Sciences, San Francisco*, **1**, pp. 399–402.
- Gomi, H., Honda, M., Ito, T., and Murano, E. (2002). "Compensatory articulation during bilabial fricative production by regulating muscle stiffness." *J. Phonetics* **30**, 261–279.
- Gracco, V. (1988). "Timing factors in the coordination of speech movements." *J. Neurosci.* **8**, 4628–4639.
- Gracco, V., and Abbs, J. (1986). "Variant and invariant characteristics of speech movements." *Exp. Brain Res.* **65**, 156–166.
- Gracco, V., and Löfqvist, A. (1994). "Speech motor coordination and control: Evidence from lip, jaw, and laryngeal movements." *J. Neurosci.* **14**, 6585–6587.
- Guenther, F., Hampson, M., and Johnson, D. (1998). "A theoretical investigation of reference frames for the planning of speech movements." *Psychol. Rev.* **105**, 611–633.
- Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., and Perkell, J. S. (1999). "Articulatory tradeoffs reduce acoustic variability during American English /t/ production." *J. Acoust. Soc. Am.* **107**, 2854–2865.
- Hertrich, I., and Ackermann, H. (2000). "Lip-jaw and tongue-jaw coordination during rate-controlled syllable repetitions." *J. Acoust. Soc. Am.* **107**, 2236–2247.
- Hoole, P., and Kühnert, B. (1996). "Tongue-jaw coordination in German vowel production," in *Proceedings of the 4th Speech Production Seminar, Autrans, France*, pp. 97–100.
- Hoole, P., and Nguyen, N. (1999). "Electromagnetic articulography in coarticulation research," in *Coarticulation: Theory, Data and Techniques*, edited by W. H. Hardcastle and N. Hewlett (University Press, Cambridge), pp. 260–269.
- Hoole, P., Wismüller, A., Leinsinger, G., Kroos, C., Geumann, A., and Inoue, M. (2000). "Analysis of tongue configuration in multi-speaker, multi-volume MRI data," in *Proceedings of the 5th Seminar on Speech Production: Models and Data*, Seon, Germany, pp. 157–160.
- Howe, M., and McGowan, R. (2005). "Aeroacoustics of [s]." *Proc. R. Soc. London, Ser. A* **461**, 1005–1028.
- Keating, P., Lindblom, B., Lubker, J., and Kreiman, J. (1994). "Variability in jaw height for segments in English and Swedish VCVs." *J. Phonetics* **22**, 407–422.
- Kelso, J., Saltzman, E., and Tuller, B. (1986). "The dynamical perspective on speech production: Data and theory." *J. Phonetics* **14**, 29–59.
- Kollia, B., Gracco, V., and Harris, K. (1995). "Articulatory organization of mandibular, labial, and velar movements during speech." *J. Acoust. Soc. Am.* **98**, 1313–1324.
- 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* **35**, 17–25.
- Kühnert, B., Ledl, C., Hoole, P., and Tillmann, H. (1991). "Tongue-jaw interactions in lingual consonants." *PERILUS* **14**, 21–25.
- Ladefoged, P., and Maddieson, I. (1996). *The Sounds of the World's Languages* (Blackwell, Oxford).
- Lee, S. (1996). "Orals, gutturals, and the jaw," in *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, edited by B. Cornell and A. Arvaniti (Cambridge University Press, Cambridge), pp. 235–241.
- Lee, S., Beckman, M., and Jackson, M. (1994). "Jaw targets for strident fricatives," in *Proceedings of the International Conference on Spoken Language Processing, Yokohama*, pp. 37–40.
- Lindblad, P., and Lundqvist, S. (1999). "How and why do the tongue gestures of [t], [d], [l], [n], [s], and [r] differ?" *Proceedings of the 14th International Congress of Phonetic Sciences*, pp. 417–420.
- Lindblom, B., and Lubker, J. (1985). "The speech Humunculus and a problem of phonetic linguistics," in *Phonetic Linguistics, Essays in Honor of Peter Ladefoged*, edited by V. Fromkin (Academic, Orlando), pp. 169–192.
- Löfqvist, A., and Gracco, V. (2002). "Control of oral closure in lingual stop consonant production." *J. Acoust. Soc. Am.* **111**, 2811–2827.
- Max, L., and Onghena, P. (1999). "Some issues in the statistical analysis of completely randomized and repeated measures designs for speech language and hearing research." *J. Speech Lang. Hear. Res.* **42**, 261–270.
- Mooshammer, C., Geumann, A., Hoole, P., Alfonso, P., van Lieshout, P., and Fuchs, S. (2003). "Coordination of lingual and mandibular gestures for different manners of articulation," in *Proceedings of the 15th International Congress Phonetic Sciences*, pp. 81–84.
- Nittrouer, S. (1991). "Phase relations of jaw and tongue tip movements in the production of VCV utterances." *J. Acoust. Soc. Am.* **90**, 1806–1815.
- Nittrouer, S., Munhall, K., Kelso, J., and Tuller, B. (1988). "Patterns of interarticulator phasing and their relation to linguistic structure." *J. Acoust. Soc. Am.* **84**, 1653–1661.
- R Development Core Team (2005). "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria; URL: <http://www.R-project.org/>
- Saltzman, E., and Munhall, K. (1989). "A dynamical approach to gestural patterning in speech production." *Ecological Psychol.* **1**, 333–382.
- Saltzman, E., Löfqvist, A., and Mitra, S. (2000). "'Glue' and 'clocks': Intergestural cohesion and global timing," in *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, edited by M. B. Broe and J. B. Pierrehumbert (Cambridge University Press, Cambridge), pp. 88–101.
- Saltzman, E., Löfqvist, A., Kay, B., Kinsella-Shaw, J., and Rubin, P. (1998). "Dynamics of intergestural timing: A perturbation study of lip-larynx coordination." *Exp. Brain Res.* **123**, 412–424.
- Schulman, R. (1989). "Articulatory dynamics of loud and normal speech." *J. Acoust. Soc. Am.* **85**, 295–312.
- Shadle, C. (1990). "Articulatory-acoustic relationships in fricative consonants," in *Speech Production and Speech Modelling*, edited by W. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 187–209.
- Shaiman, S., Adams, S., and Kimelman, M. (1995). "Timing relationships of the upper lip and jaw across changes in speaking rate." *J. Phonetics* **23**, 119–128.
- Tasko, S., and Westbury, J. (2002). "Defining and measuring speech movement events." *J. Speech Lang. Hear. Res.* **45**, 127–142.
- van Lieshout, P. (1995). *Motor Planning and Articulation in Fluent Speech of Stutterers and Nonstutterers* (NICI, Nijmegen).
- Westbury, J., Lindstrom, M., and McClean, M. (2002). "Tongue and lips without jaws: A comparison of methods for decoupling speech movements." *J. Speech Lang. Hear. Res.* **45**, 651–662.