Coordination of lingual and mandibular gestures for different manners of articulation

Christine Mooshammer[†], **Anja Geumann**[‡], **Phil Hoole**[‡], **Peter Alfonso***, **Pascal van Lieshout**° **and Susanne Fuchs**[1]

† IPDS, Christian-Albrechts Universität, Kiel, Germany; ‡ Institut für Phonetik und Sprachliche Kommunikation, LMU, München, Germany; * University of North Dakota, Grand Forks, USA; ° Graduate Department of Speech-Language Pathology, University of Toronto, Canada; [1] ZAS, Berlin, Germany

E-mail: timo@ipds.uni-kiel.de

ABSTRACT

In Articulatory Phonology the jaw is not controlled individually but serves as an additional articulator to achieve the primary constriction. In this study the timing of jaw and tongue tip gestures for the coronal consonants /s, ʃ, t, d, n, l/ is analysed by means of EMMA. The findings suggest that the tasks of the jaw for the fricatives are to provide a second noise source and to stabilise the tongue position (more pronounced for /s/). For the voiceless stop, the speakers seem to aim at a high jaw position for producing a prominent burst. For /l/ a low jaw position is essential for avoiding lateral contact and for the apical articulation of this sound.

1. INTRODUCTION

In the traditional description of sounds as well as in Articulatory Phonology [2, 11] the jaw does not count as a primary articulator with respect to linguistic information but rather as an additional contribution to the principal articulator and is, therefore, missing in the IPA. In Task Dynamics, the jaw is subordinate to the tract-variables of all lingual and labial sounds and therefore a member of the coordinative structures constituting the primary task. At least since Shadle [13], the role of the lower incisors and therefore jaw height position as a second noise source for the production of sibilants has been generally acknowledged. Subsequent studies of jaw movements showed that jaw positions vary little with vowel context if the consonant is one of the two sibilants /s,ʃ/, or the voiceless stop /t/ (e.g. [4,7,9]), compared to e.g. /l/. Furthermore, the position of the jaw is higher, i.e. more closed, for the consonants /s,ʃ,t/ in comparison with /d,n,l/ (see Fig. 1). If the jaw had simply a supporting function for lifting the tongue tip then the order of the consonants should be the same for jaw and tongue tip height, i.e. consonants with a high jaw position should also have a high tongue position. As can be seen in Fig. 1 the opposite is the case: /s/ is produced with a high jaw position but the tongue tip is low. The lateral on the other hand is produced with a low jaw position but a very high tongue position. Therefore the contribution of the jaw for forming the constriction differs according to the manner of articulation. In Task Dynamics, a larger influence of the jaw is modelled by attributing larger weights to this composite articulator. Because of the high and invariant position of the jaw during stridents, Lee *et al.* [9] suggest the introduction of a jaw gesture with its own fixed task.

The general aim of this study is to further investigate the differential role of the jaw for coronal consonants especially with respect to timing relations between tongue tip and jaw.

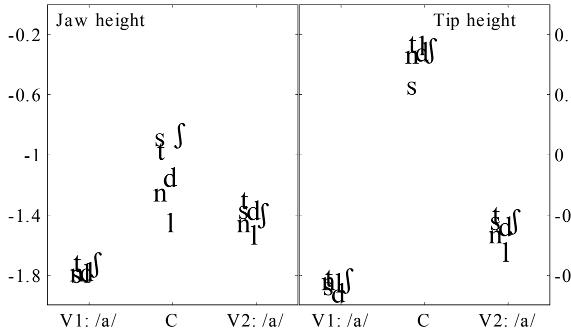

Figure 1: Vertical jaw (left) and tongue tip positions (right) during /ˈaːCa/-sequences averaged over 5 speakers with 12 repetitions each (adopted from Geumann [4], same set of data is used for this study)

2. METHOD

The speech material consisted of /ˈVCV/ embedded in the carrier phrase “Hab das Verb ___ mit dem Verb___ verwechselt” with the target consonants consisting of the coronal phonemes of German differing in manner of articulation /s,ʃ,t,d,l,n/ and the symmetrical long vowel context /iː,eː,aː/ with the first vowel stressed and the second one unstressed. All 18 sequences were produced 12 times in randomised order at normal and loud volume by five speakers. Articulatory data were obtained by means of Electromagnetic Midsagittal Articulography (Carstens Medizinelektronik AG100) at a sampling rate of 250 Hz. Four sensors were placed on the tongue and three on the jaw (inner and outer surface of the gums, angle of the chin).

The intrinsic tongue was estimated by the method proposed by Edwards [3]. From MRI data 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 during the mid part 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.

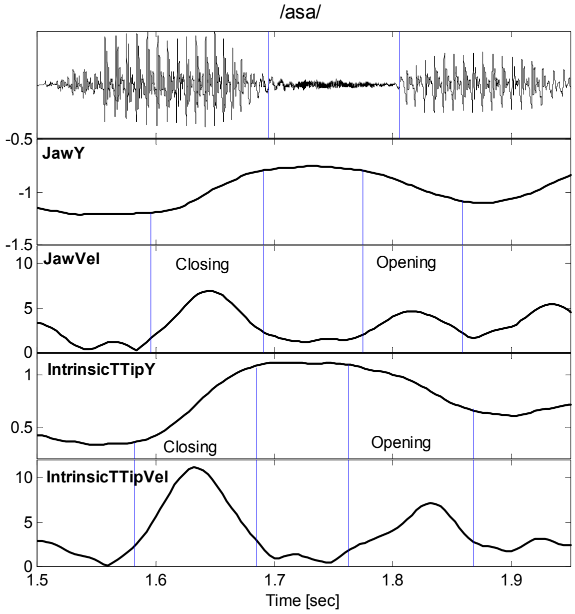

Figure 2: Labelling criteria: Upper panel to lower panel: audio signal of /ˈaːsaː/ by speaker RS, vertical jaw movement in cm, tangential velocity signal of jaw in cm/s, vertical intrinsic tongue tip signal in cm, tangential velocity signal of intrinsic tongue tip in cm/s. Vertical lines: acoustical on- and offset of the consonant, on- and offset for closing and opening movements for jaw and tongue tip.

For this study only items in /a/ context and at normal volume were analysed. On- and offsets of jaw and intrinsic tongue tip movements were measured at 20% of the tangential velocity peak. The interval between the offset of the closing movement and the onset of the opening movement is termed the target interval here. For some speakers and sounds (especially /d/ for speaker AW, /s/ for speaker SR) the minimum in the tangential velocity signal of intrinsic tongue tip during the consonant was very late or very early because of the high amount of forward and downward movement during the consonant. Therefore, these sounds showed very short tongue tip target durations.

Besides on- and offsets, the time points of peak velocities were also measured. Displacements during opening and closing phases were computed as the integral of the tangential velocity signal. Statistics were carried out by using the GLM procedure of SPSS with consonant as independent factor.

3. RESULTS

3.1. DISPLACEMENTS

The amplitude of the jaw closing gesture is highest for the voiceless obstruents (/s/ 3 speakers, /ʃ/ one speaker, /t/ one speaker) and always lowest for /l/. As can be seen in Fig. 1, jaw positions in the unstressed postconsonantal vowel is influenced by the identity of the preceding consonant: after /l/ jaw is lowest and after /t/ highest. Since the consonantal jaw position is already low for /d,n,l/ some cases are found where either the opening movement could not be analysed or even an increase in jaw position occurs.

3.2. LATENCIES

Table 1 shows the latencies between intrinsic tongue tip and jaw, with negative values for jaw advancement and positive values for intrinsic tongue tip advancement. Latencies of the closing movement onset (CLon) are highly variable and speaker dependent, probably due to the fact that the preceding consonant is a bilabial which leaves the tongue tip unconstrained to initiate the movement towards the alveolar constriction.

C	CLon	CLvel	TARon	OPon	OPvel	Dur
s	-10.9 (27.5)	-11.2 (29.5)	4.8 (24.0)	-11.2 (24.5)	-14.3 (17.4)	126.5 (17.2)
ʃ	-3.0 (28.9)	-3.1 (19.1)	26.9 (20.2)	-28.3 (32.4)	-32.3 (20.9)	124.2 (14.9)
t	1.0 (22.6)	-2.0 (11.8)	20.4 (18.5)	5.2 (15.0)	-8.5 (15.2)	71.1 (17.7)
d	-10.6 (25.2)	-6.1 (10.4)	19.9 (13.2)	6.3 (23.1)	2.4 (18.0)	56.2 (13.4)
n	1.0 (31.5)	-3.4 (10.3)	15.5 (17.3)	-4.2 (20.4)	5.7 (15.3)	68.7 (13.0)
l	3.2 (29.1)	-3.4 (18.2)	12.6 (12.7)	-3.2 (17.8)	0.4 (23.3)	64.8 (13.0)

Table 1: Means and standard deviations of latencies between intrinsic tongue tip and jaw movements in ms at onset of closing movement (CLon), peak velocity of the closing movement (CLvel), onset of target (TARon), onset of opening movement (OPon) and peak velocity of the opening movement (OPvel). Negative: jaw first, positive: tip first. Acoustical consonant duration (Dur).

The velocity peak latency of the closing gesture (CLvel) shows a significantly greater jaw advancement for /s/ compared to the other sounds. Gracco [5] interpreted a differential timing of velocity peaks as a “feedforward” articulator information for adjustments of positions or timing of other subcomponents of a coordinative structure. Overall speaker results are considerably influenced by one speaker (SR) whose intrinsic tongue movements for /s/ could not be separated in closing and movement during /s/ with the currently used labelling criteria. Generally, the standard deviations for the velocity peak latency of the closing movement is smaller than the other latencies which is in agreement with Gracco’s [5] and van Lieshout’s [15] results. Tongue tip and jaw reach their targets (TARon) for /s/ almost simultaneously whereas for all other consonants the jaw achieves its target later than the tongue tip. Again this result is highly speaker-dependent: one speaker had a mean latency of –25 ms and another showed the highest latency for /s/ (33 ms), which means that the jaw reached its target with a long delay. /ʃ/ on the other hand showed the longest delay of jaw target achievement.

Tongue tip and jaw started the opening gesture (OPon) more or less synchronously for the consonants /t,d,n,l/ but not for /s/ and /ʃ/. For the fricatives, the jaw release occurred earlier than the tongue tip opening gesture ($p<0.05$ for /ʃ/ all speakers, for /s/ 2 speakers). This differential timing for the fricatives becomes even more pronounced for the latencies between velocity peaks of the opening gesture (OPvel), where the jaw velocity peak precedes the tongue tip peak by as much as 32 ms for /ʃ/ and 14 ms for /s/. The peak velocity latency is significantly smallest for /ʃ/ for all speakers and differs from /s/ for three speakers. These differences in tongue-tip-jaw latencies could be attributed to the fact that the tongue tip sensor is placed in front of the relevant articulator for the post-alveolar.

For the stops there is a tendency that the jaw starts its opening movement later than the tongue tip release but this is significant only for one speaker.

3.3. NORMALIZED DURATIONS

Since it is not clear whether the manner-dependent differences in latencies could be attributed to the longer durations of the fricatives, target on- and offsets were normalised to the acoustical consonant durations individually (see Table 1). Results are shown in Fig. 3 with 0 and 1 denoting the acoustically defined begin and end of the consonant respectively. Unfilled bars show the target duration and relative timing of the intrinsic tongue tip target achievement and release, grey bars of the jaw.

For all consonants the intrinsic tongue tip preceded the jaw for target achievement but for /s/ the difference was very small (n. s. for two subjects) and the jaw reached its target quite early with respect to the acoustically defined consonant. The stops usually showed the latest achievement for the jaw target (/t/ > /n,s,l,ʃ/, /d/ > /ʃ,s,l/, n.s. for speaker KH, /t/ later than /d/ for speaker RS).

For the onset of the opening movement the jaw started its opening movement for the following vowel latest for the stops, which might be due to the fact that the burst, which has a very close relation to supralaryngeal articulation, was used as criterion for the end of the acoustical consonant. Since the latencies for onset of opening movement are also highest for /t,d/ compared to all other consonants (sig. higher than /s,ʃ/), i.e. jaw starts the opening movement somewhat later than tongue tip, we are tempted to conclude that the speakers aim at a high jaw position for the burst.

Since the durations of intrinsic tongue tip plateaux were highly speaker-dependent and very variable they will not be discussed here. Jaw target durations in percent of acoustical durations were longer for /t,d,s/ compared to /n,l,ʃ/. For absolute target durations, the alveolar sibilant showed a significantly longer hold phase of the jaw than the post-alveolar ($p<0.05$ 3 subjects).

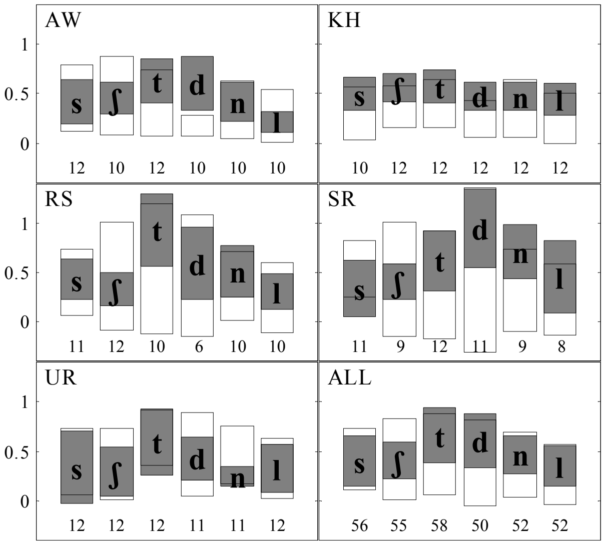

Figure 3: Normalised durations of intrinsic tongue tip (unfilled boxes) and jaw target (grey boxes) on- and offsets. 0 denotes the acoustically defined begin of the consonant and 1 the end with number of measured items.

4. CONCLUSIONS

It has often been observed that there is a high amount of variability for interarticulatory coordination both inter-individually (e.g. [6]) and intra-individually over sessions ([1]). In the present work, especially latencies for the onset and to a lesser degree for the target achievement varied a lot whereas the peak velocities seemed to show a more strict timing for different articulatory structures. This is not only in agreement with the findings of Gracco, who also found a high correlation between peak amplitude of muscle activity and time to peak velocity, but also with the view that gestural activation intervals in Articulatory Phonology should not be modelled as step-rectangular force functions but with a more smooth signal e.g. a continously increasing onset and decreasing offset phase ([8,12]). The full force of the gestural activation would then be reached approximately at the moment of the velocity peak.

The high amount of variability for the onset latencies could also be due to consonant context: since in the present study the preceding consonant is a bilabial the tongue tip is free to vary. In another set of data (unpublished manuscript) where the vowel-preceding consonant was /s/ or /t/ highly significant differences in closing onset and target latencies between /s/ and /t/ were found with synchronous tongue tip and jaw latencies for /s/ and a delay of the jaw for /t/.

In terms of spatial parameters, exact jaw positioning plays an important role for the two sibilants. This has been attributed to the well-known fact that sibilants are produced with a second noise source, the lower incisors. Therefore, one could assume that the temporal parameters for these two sibilants are also similar. As was found in the present study, the alveolar fricative was produced with a longer jaw target duration than the post-alveolar. Therefore, we assume that the task of the jaw differs for the two sibilants. One explanation could be that for both sibilants the task of the jaw is twofold: building an obstacle and supporting the tongue to form a critical constriction. This second task might be reduced for the post-alveolar where the relevant articulator is further back. As can be seen in Fig. 3, the portion of jaw target interval for /ʃ/ compared to the lingual target duration is much smaller than for /s/, which shows that the jaw is less involved in stabilising the tongue position for the former fricative. Further evidence for a more stable lingual constriction for /ʃ/ is given by Tabain [14], who found a higher coarticulatory resistance due to a more constrained tongue shape compared to /s/.

For the voiceless stop an exact and high jaw position seems also crucial. In the present data the target position for the jaw is reached relatively late during both alveolar stops, i.e. close to the burst. Furthermore, the jaw tended to stay at its position even after the tongue tip started its opening movement. The difference between /t/ and /d/ is that the jaw shows a higher contextual variability and also a lower jaw position for /d/. The task of the jaw for the voiceless stop could be to provide a close constriction for a salient burst.

For the lateral a low jaw position is needed for providing space for the more apical articulation of this sound and for avoiding lateral contact of the tongue sides (e.g. [10]). For the nasals the jaw plays only a subordinate role, as suggested in Task Dynamics, in a supporting function for the tongue tip.

In conclusion contrary to the suggestion of Lee et al. [9] we assume that the high and invariant jaw position for sibilants could probably be modelled by very large weights for this articulator and that no additional tract-variable is necessary. Whether the differential timing of tongue tip and jaw for the voiceless apical stop can be controlled by one single gesture is yet not clear to us.

ACKNOWLEDGMENTS

This work was funded by the German Research Council (DFG, Ti69/31) and a DAAD grant for the first author.

REFERENCES

[1] Alfonso, P., P. van Lieshout, “Spatial and temporal variability in obstruent gestural specification by stutterers and controls: comparisons across sessions,” in Hulstijn, Peters, Van Lieshout, eds. *Speech Production: Motor Control, Brain Research and Fluency Disorders*, pp. 151-160. Amsterdam: Elsevier, 1997.

[2] Browman, C. and L. Goldstein, “Tiers in Articulatory Phonology, with some implications for casual speech,” in Kingston, Beckman, eds. *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, pp. 341-376, 1990.

[3] Edwards, J., “Contextual effects on lingual-mandibular coordination,” *JASA*, vol. 78, pp. 1944-1948, 1985.

[4] Geumann, A., “Invariance and variability in articulation and acoustics of natural perturbed speech,” *FIPKM*, vol. 39, 2001.

[5] Gracco, V., “Timing factors in the coordination of speech movements,” *Journal of Neuroscience*, vol. 8, 4628-4639, 1988.

[6] Johnson, K.; Ladefoged, P.; Lindau, M., “Individual differences in vowel production,” *JASA*, vol. 94, pp. 701-714, 1993.

[7] Keating, P.; B. Lindblom, J. Lubker and J. Kreiman, “Variability in jaw height for segments in English and Swedish VCVs,” *Journal of Phonetics*, vol. 22, pp. 407-422, 1994.

[8] Kröger, B.J., G. Schröder and C. Opgen-Rhein, “A gestural-based dynamic model describing articulatory movement data,” *JASA*, vol. 98, pp. 1878-1889, 1995.

[9] Lee, S., M. Beckman and M. Jackson, “Jaw targets for strident fricatives,” *Proceedings of ICSLP*, Yokohama, pp. 37-40, 1994.

[10] Lindblad, P. and S. Lundqvist, “How and why do the tongue gestures of [t], [d], [l], [n], [s], and [r] differ?,” *Proceedings of the 14th ICPhS*, pp. 417-420, 1999.

[11] Saltzman, E. and K. Munhall, “A dynamical approach to gestural patterning in speech production,” *Ecological Psychology*, vol. 1, pp. 333-382, 1989.

[12] Saltzman, E. and D. Byrd, “Task-dynamics of gestural timing: Phase windows and multifrequency rhythms,” *Human Movement Science*, vol. 19, pp. 499-526, 2000.

[13] Shadle, C., “Articulatory-acoustic relationships in fricative consonants,” in *Speech Production and Speech Modelling*, Hardcastle, Marchal Eds., pp. 187-209. Dordrecht: Kluwer, 1990.

[14] Tabain, M., “Variability in fricative production and spectra: Implications for the Hyper- and Hypo- and Quantal Theories of speech production,” *Language and Speech*, vol. 44, pp. 57-94, 2001.

[15] Van Lieshout, P., Motor planning and articulation in fluent speech of stutterers and nonstutterers, Nijmegen: *NICI*, 1995.