Volume 44 May 2014 ISSN 0095-4470

Journal of
Phonetics

Special Issue: Dynamics of Articulation and Prosodic Structure

Guest Editors:
Taehong Cho, Martine Grice and Doris Mücke

Journal of Phonetics 44 (2014) 8–24

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Laryngeal–oral coordination in mixed-voicing clusters

Philip Hoole [*], Lasse Bombien

Institute of Phonetics and Speech Processing, Schellingstraße 3, D-80799 Munich, Germany

ARTICLE INFO

Article history:
Received 14 March 2013
Received in revised form
27 January 2014
Accepted 6 February 2014
Available online 20 March 2014

Keywords:
Laryngeal–oral coordination
Laryngeal transillumination
German
Consonant clusters

ABSTRACT

Laryngeal–oral coordination was studied in German clusters of voiceless fricative or plosive plus /l/ or /r/ by means of videofiberendoscopy and transillumination. In all cases voice onset time (i.e. the time from release of C1 to onset of voicing) was longer in the clusters compared to the single fricative or plosive controls. However, the coordination patterns leading to this consistent acoustic effect were quite varied, ranging from a passive effect of aerodynamic conditions at release of C1, via shortening of C1 with constant glottal gesture, to enhancement of the glottal gesture. Active reorganization was particularly clear in the rhotic clusters. For the single consonants the duration of the glottal gesture was quite constant over place of articulation but occlusion duration varied systematically. Accordingly, for both clusters and singletons peak glottal opening did not keep a constant timing relationship to landmarks in the oral occlusion of C1. The above findings were robustly present over a range of prosodic conditions. Prosodic strengthening itself had a particularly clear influence on the magnitude of the devoicing gesture.

1. Introduction

This paper will be concerned with laryngeal–oral coordination in syllable-initial clusters of German consisting of a voiceless element (plosive or fricative), followed by a sonorant (/l/ or /r/). Our basic contention is that such sound sequences can be potentially very useful for highlighting the gaps in our knowledge about the principles underlying interarticulatory coordination (just as much as the more frequently investigated purely voiceless clusters; e.g. Löfqvist & Yoshioka, 1980; Ridouane, Fuchs, & Hoole, 2006; Yoshioka, Löfqvist, & Hirose, 1981). Let us assume as the most basic hypothesis that in both a singleton onset such as /p/ as well as in a cluster onset such as /pl/ the devoicing gesture is organized with respect to the underlyingly voiceless segment /p/. The addition of /l/ to the onset should not then have any effect on the timing of landmarks in the glottal gesture relative to those in the oral gesture of the /p/. Such a scenario would account perfectly well for the well-known substantial devoicing of, for example, /l/ in /pl/ and /r/ in /pr/ in languages such as English and German as a simple coarticulatory process, given that the laryngeal–oral timing pattern for syllable-initial /p/ involves peak glottal opening located close to the release of the oral occlusion.[1] As we will see below, this basic scenario is actually not well supported by currently available findings. There is even some support for radically different patterns, for example that addition of an underlyingly voiced sonorant to an otherwise voiceless syllable onset can lead to an increase in the magnitude of the glottal gesture. Thus, the fundamental motivation for our own investigation is that we are currently ignorant about how the laryngeal–oral coordination relations should be formulated, and that mixed-voicing syllable onsets have interesting implications for the level of the syllabic hierarchy at which the devoicing gesture is organized.

In the following paragraphs we will review earlier work, aiming to identify potential patterns of laryngeal–oral coordination, summarizing them schematically in Fig. 1 to provide a framework for further discussion.

A convenient point of departure for consideration of previous findings is Docherty's (1992) acoustic investigation of English. Based on his own results, and other relevant results available at that time, two fairly pervasive generalizations can be identified:

1. VOT (i.e. the period of voicelessness following release of the stop or fricative) is longer in /Cl/ and /Cr/ sequences than in simple CV sequences.
2. It is well documented that stops and fricatives generally have a shorter occlusion duration when they occur in clusters (e.g. Haggard, 1973; Hawkins, 1979; Klatt, 1975, 1973), though in fact we will also be encountering some cases where this effect is quite weak.

---

[*] Corresponding author. Tel.: +49 89 2180 3149; fax: +49 89 2180 5790.
E-mail addresses: hoole@phonetik.uni-muenchen.de (P. Hoole), lasse@phonetik.uni-muenchen.de (L. Bombien).

[1] We will be using the phonemic symbol /r/ to refer to the rhotic that in our German speakers was a dorsal articulation in the velar/uvular region. Depending on context, it can range from voiced approximant to voiceless fricative.

0095-4470/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2014.02.004

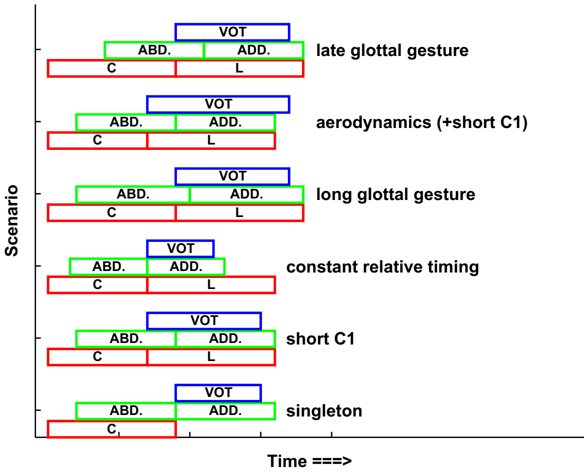

Fig. 1. Schematic illustration of possible laryngeal–oral timing scenarios in clusters (with singleton plosive as reference). For each scenario, the lowest of the three horizontal bars shows oral timing. ‘C’ stands for plosive C1, ‘L’ stands for lateral or rhotic C2. The middle horizontal bar shows glottal activity. ‘ABD’ stands for ‘abduction’ (i.e. opening of the glottis), ‘ADD’ for adduction (closing). Peak glottal opening occurs at the junction between ‘ABD’ and ‘ADD’. The top bar (‘VOT’) extends from the release of C1 to the point at which voicing resumes. To make timing shifts easier to follow across the scenarios, the location of peak glottal opening in the baseline singleton case has been made synchronous with the release of C1. In practice (as seen in the empirical results in Fig. 13), peak glottal opening is typically timed somewhat later (as a further minor simplification the duration of ‘ABD’ has been assumed to be equal to ‘ADD’).

What might explain the increased VOT in sequences such as /pl/ compared to singleton /p/? One possible scenario is that it is a simple link between the above two generalizations: /p/ and /pl/ may have a glottal gesture of similar duration, but a shorter occlusion duration in cluster /pl/ (see scenario “Short C1” in Fig. 1). In Hoole and Bombien (2010) we have suggested that an organizational principle of this kind may be in operation for, in effect, the mirror-image case in Icelandic: pre-aspirated plosives, on the one hand, and voiceless nasal plus plosive sequences, on the other hand, have a glottal gesture of similar duration but shorter plosive occlusions in the latter case. In fact, Docherty's data did not give much support for this scenario, since there were cases where VOT increased more (in the clusters) than would be predicted by the reduction in occlusion duration (i.e. for some clusters the reduction was negligible). In effect, for some clusters, the total duration of voicelessness was longer than in the singletons, indicating that the duration of the glottal devoicing gesture might be longer in the clusters (scenario “long glottal gesture” in Fig. 1).[2]

This is a striking result, because probably no coarticulation theory would predict that adding an underlyingly voiced sound to a syllable onset could result in an increase in devoicing. There may well be one simple passive explanation for the unexpected fact that the stop-continuant cluster has a longer period of devoicing than the simple stop, namely that the aerodynamic conditions in the continuant are not conducive to initiation of phonation, due to the fact that the oral tract is still partially occluded (see Hanson & Stevens, 2002, especially p. 1175 ff). Recent work of our own using articulatory synthesis (Hoole, Pouplier, Beňuš, & Bombien, 2013) confirms the potential relevance of partial constrictions for aerodynamic conditions: In simulations of /br/ the re-initiation of voicing after devoiced /b/ was delayed if strong dorsal constriction for /r/ was already present at release of /b/. A schematic illustration of how the presence of an aerodynamic effect would be reflected in our measurements is given by scenario “aerodynamics (+short C1)” in Fig. 1. This simply uses the oral and glottal (abduction and adduction) timing patterns from the “short C1” scenario and shifts the right edge of the VOT phase (i.e. the location of voice onset) to a later point in time. Comparison of “short C1” and “aerodynamics (+short C1)” also illustrates the important point that depending on the strength of the influence of aerodynamic conditions there can be a mismatch between the acoustically measured duration of voicelessness and the articulatorily measured duration of the laryngeal abduction and adduction movement. Since in the present study we measure the laryngeal movements themselves, it is possible to weigh up the contributions to any increase in duration of voicelessness more directly than would be possible in a purely acoustic study. Specifically, for the baseline singleton case in Fig. 1 (an aspirated plosive) we assume that at the time when glottal adduction is approaching completion no aerodynamically relevant constriction of the vocal tract remains. Accordingly, the precise time point of voice onset is directly determined by the glottal configuration itself: in Fig. 1 we show a situation that is typical in our recordings, namely that the acoustically determined voice onset occurs slightly before glottal adduction is complete (as measured from the transillumination signal). In the cluster scenario “short C1” the position of voice onset relative to the end of glottal adduction is the same as in the singleton case. If such a pattern occurs in the empirical data below we would take this as evidence that any constriction related to C2 has ***no*** aerodynamically relevant effects. In the “aerodynamics (+short C1)” scenario the voice onset shifts to a later time relative to completion of glottal adduction, leading to the opposite conclusion.[3]

Based on the results of previous investigations we assume that it is simply an open issue whether the longer period of voicelessness in clusters compared to singletons reflects a longer glottal abduction–adduction cycle in clusters or, rather, reflects aerodynamically less favorable conditions for voicing in clusters (or some combination of these two mechanisms; further discussion in Docherty (1992, pp.148–149)). But given that lengthening of the glottal gesture itself is a realistic possibility (confirmed by instrumental results we discuss below), it is interesting to consider what could drive speakers to adopt this pattern of movement.

An output oriented style of explanation for lengthening of the glottal gesture might be that it is perceptually important to have a substantial amount of devoicing on the second element in a cluster when the first element is underlyingly voiceless (e.g. to separate “played”, “blade”, “prayed”, “braid”). In other words, languages such as English and German exhibit a large number of contrasting combinations of C1 and C2 in syllable-onset position in which the listener needs to be able to recover the voicing status and place of articulation of C1 as well as the manner of articulation of C2. Since a low onset frequency of voiced F1 is known to support voicing percepts it may be advantageous to delay voice onset in voiceless clusters to reduce the

---

[2] Note that the patterns shown in Fig. 1 are not intended to be exhaustive, but rather illustrative of the basic possible mechanisms. For the “long glottal gesture” pattern only the glottal gestural duration is changed, but the C1 duration is kept the same as in the basic singleton case. But this pattern could, of course, be combined with the “short C1” to give an even stronger increase in VOT.

[3] See Cho and Ladefoged (1999) for a discussion of the greater relevance of aerodynamic conditions for voice onset in unaspirated compared to aspirated plosives.

salience of low F1 due to the constriction of C2. Perhaps also it is easier for listeners to distinguish voiceless (devoiced) sonorants if their duration is long.

A further possibility is that given the aerodynamic conditions in the vocal tract, completion of glottal adduction early in C2 might not lead to reliable re-initiation of voicing anyway, so speakers reorganize the gesture to be completed closer to the time at which aerodynamic conditions are favorable, potentially reinforcing the total duration of voicelessness (we have argued elsewhere (Hoole, Pouplier, & Kühnert, 2012) that speech movements are often organized to exploit the physical conditions in the vocal tract: to 'go with the flow').

As already emphasized, empirical verification is required as to whether lengthening of the glottal gesture actually occurs. But previous results do, however, make it seem unlikely that duration of the glottal gesture is shorter in clusters. This is in itself a significant finding since, given a tendency to shorter occlusion durations, a shorter glottal gesture could well be expected under the plausible assumption that the component gestures of an aspirated plosive become modified in parallel (this is what we referred to as the "basic hypothesis" in the first paragraph of Section 1). For example, working within the framework of the Task-Dynamics model, Saltzman and Munhall (1989) point to evidence from perturbation experiments that the laryngeal gesture is modified when the bilabial closure for /p/ is interfered with experimentally. They cite this as evidence for a level of intergestural cohesion that must exist (otherwise speakers would hardly be able to systematically produce VOTs in the range appropriate for their language). We include the scenario in which oral and glottal gestures shorten in parallel as the pattern labeled "constant relative timing" in Fig. 1. The key feature distinguishing this scenario from the other ones shown in the figure is that peak glottal opening (at the boundary between abduction and adduction) stays at the same location relative to the release of the C1 occlusion as it does for the baseline singleton case. Even though previous results have provided little support for this scenario (note also that in the form shown in Fig. 1 it incorrectly predicts a shortening of VOT) it is conceptually important. Indeed, based on the above discussion of aerodynamic effects, a combined scenario can be envisaged in which the glottal and oral kinematics show constant relative timing but voice onset is delayed beyond the end of glottal adduction if constriction conditions in C2 are very unfavorable for voicing (in fact, a delay in voice onset may be more likely in the "constant relative timing" case than, for example, in the "short C1" case because in the former case glottal adduction is completed at a time when the C2 constriction is still fully formed, whereas in the latter case C2 may already have weakened).

Before turning to discussion of currently available instrumental findings one further important element in setting the scene is given by the gestural approach of Browman and Goldstein (1986). Based on their analysis of voiceless clusters Browman and Goldstein (1986) come to the conclusion that it can be stated as a regularity of English that a word (syllable) can only begin with one devoicing gesture (and we assume initially, because of the similar status of aspiration, that German is likely to pattern in a similar way).

(1) If a fricative gesture is present, coordinate the peak glottal opening with the midpoint of the fricative.
(2) Otherwise, coordinate the peak glottal opening with the release of the stop gesture.

These two simple rules are very attractive because they already account for important regularities in English such as short VOT in a plosive following a fricative, and devoicing of sonorants in clusters of concern here like /pl/ (see also Iverson & Salmons, 1995; Kehrein & Golston, 2004.[4]).

The rules thus form a useful framework in the present context. But the scenarios we have presented above indicate that departures from these coordination patterns may be observed. Thus one aim of the present work is to determine whether these rules need to be supplemented by statements taking the segmental context into account (we come back to the idea of context-sensitive implementation of the devoicing gesture again below, e.g. in discussion of the effect of place of articulation in single consonants).

Turning to investigations that have looked directly at laryngeal behavior in the relevant clusters (i.e. as in our own investigation by means of laryngeal fiberscopy and transillumination), Jessen (1999) considers explicitly the question of whether duration of voicelessness in the sonorant is directly related to shortening of the plosive occlusion. The material came from one speaker of German and consisted of /p, t, k/ combined (to the extent possible in German) with /l, r, n/. The results were somewhat mixed: The clusters with initial /k/ (/kl/, /kn/) were overall reasonably consistent with an account relating VOT duration to the shortening of the plosive occlusion duration, while the p-initial cluster /pl/ showed the expected increase in VOT but with only negligible shortening of the occlusion duration. A rather striking finding in this investigation involved stop-/r/ clusters. All the combinations (/pr, tr, kr/) showed a substantial increase in duration of the glottal gesture. In addition, Jessen (1997) reported on three-element clusters: Those with /r/ as the last element (e.g. /ʃpr/) were the only ones to show a double-peaked abductory movement in the transillumination signal. In the discussion section we will consider in more detail what might contribute to the emergence of the particularly extensive glottal abduction in /r/-final clusters. For the moment we simply note that the net result is that /r/ is realized as a voiceless uvular fricative, and voiceless fricatives are, of course, known to require strong glottal abduction. In any case it is interesting that a contrast such as /pr/ vs. /br/, that underlyingly only appears to involve the first element in the onset, is actually planned in terms of a pattern of activity that involves the whole onset.

Tsuchida and Cohn (2000) examined clusters with /l/ for one speaker of American English. For plosives, they found the expected slight shortening of the oral occlusion (for cluster vs. singleton), but the glottal gesture was actually somewhat longer in the clusters. By contrast, in the comparison of single fricatives and fricative-initial clusters, the clusters showed a reduction in duration of the glottal gesture that roughly paralleled the reduction of the fricative oral occlusion duration. Thus, in terms of the scenarios introduced above we have a first indication that plosive-initial and fricative-initial clusters need not necessarily behave in the same way: fricative-initial clusters following "constant relative timing", plosive-initials showing a combination of "short C1" and "long glottal gesture".

Hoole (2006, some preliminary results of this study also in Hoole, Fuchs, and Dahlmeier, 2003) examined clusters with /l/ for three speakers of German. The only plosive-initial cluster recorded was /pl/. On the one hand there was a consistent finding of the expected longer VOT in the cluster compared to the singleton /p/, and the alignment of the glottal gesture with the oral gesture for /p/ was different in the cluster compared to the singleton. However, it was not possible to identify a consistent change in the coordination pattern over the three subjects leading to the longer VOT. This was tentatively seen as an indication that lengthened VOT in the cluster condition might indeed be the key output constraint guiding the speakers' articulatory behavior; this constraint could be fulfilled by quite a variety of coordination strategies: reducing the occlusion duration or lengthening the glottal gesture are perhaps just the two most obvious ones; there was even an indication that keeping glottal gesture duration constant but delaying its onset might be a further possibility (illustrated as the "late glottal gesture" scenario in Fig. 1). In this study two fricative + /l/ clusters were recorded (/fl/ and /ʃl/). There was once again a consistent finding that the alignment of the glottal gesture with the oral gesture of C1 changed when going from

---

[4] Kehrein and Golston's (2004) conclusion of their wide-ranging typological study fits in well with this general point of view: "laryngeal features are properties not of segments, but of the onsets, nuclei and codas that dominate them" (p. 325).

the singletons to the clusters. Reduction of the duration of C1 occlusion seemed to make a more substantial contribution to the shift in coordination than was the case for the /pl/ cluster.

The small number of subjects in the studies just discussed would itself be sufficient motivation for the present study. But this is all the more so because on the basis of previous results not even the simplest statements about laryngeal–oral coordination in these still relatively straightforward sound sequences can be made with any confidence: in other words, glottal gesture durations in the cluster that are shorter, longer or the same as in the corresponding singleton are all realistic possibilities. For the particular case of /pl/ in our earlier study we advanced an output-oriented account that might rationalize apparently heterogeneous articulatory strategies, but three subjects is clearly insufficient to be confident that this heterogeneity really exists; conceivably, with more subjects a clearly preferred strategy might emerge. A further specific area where more subjects would be highly desirable relates to the striking findings of Jessen for clusters with /r/: this raises the possibility that superficially rather similar syllable onsets such as /pr/, /pl/ might require differing laryngeal specifications. This is particularly intriguing because both would be regarded on paper as ‘good’ (unproblematic) onsets in terms of preferred sonority profiles, while an active realization of /r/ as a voiceless fricative would result in a much less preferred sonority profile for the syllable onset.

1.1. *Laryngeal–oral coordination in single consonants*

Although the main focus of the present investigation is on laryngeal–oral coordination in clusters we will also lead in to the cluster results by looking at coordination in single consonants. The motivation for this is that in particular the relationship between place of articulation of plosives and VOT duration potentially involves similar scenarios to those introduced above for clusters. Specifically, it has been a common finding that VOT is shorter in /p/ than /k/, while occlusion duration is typically shorter in the latter. Thus, comparably to the “short C1” scenario shown for singleton vs. cluster in Fig. 1 it is conceivable that glottal gesture duration remains the same, with the increased VOT being a direct consequence of the shorter occlusion. All the studies quoted above present data relevant to this issue (see in particular Jessen (1999), for a very clear exposition of the possible scenarios; see also Cooper, 1991; Hoole, Pompino-Marschall, & Dames, 1984; Hutters, 1984). Broadly speaking, these studies indicate that shortening of the occlusion duration is certainly a significant element in the lengthening of VOT, but on the other hand the duration of the glottal gesture does not necessarily remain completely constant over all places of articulation. In their wide-ranging cross-language study of VOT, Cho and Ladefoged (1999) suggest that, just as the specification of target values of laryngeal–oral timing for a category such as voiceless aspirated must be part of the language's grammar (because across languages different choices are made from the range of possible values), so, too, it may also be necessary to include context-specific rules referring to place of articulation.

In any case, by incorporating single consonants into our discussion we extend the range of material on which to discuss the extent to which speakers try to get as much mileage as possible out of a rather constant glottal gesture – even if the duration is not fixed in a hard and fast sense – with oral gestures showing a shifting pattern relative to this laryngeal substrate. Here it is worth recalling Shipp's (1982) suggestion that the highly preprogrammed nature of the abductory–adductory cycle may make the larynx ‘one of the basic metronomes of the speech production process’ (p. 111; see also Weismer, 1980).

1.2. *Prosodic effects*

For several reasons it was considered interesting to investigate the organization of the laryngeal devoicing gesture over a range of prosodic conditions in which the target sounds occurred in contexts differing with respect to prosodic strength (achieved, for example, by manipulating phrasal accent position; further details below). Firstly, prosodic manipulation can be regarded as a probe to test the robustness of any differences between singletons and clusters, or between different clusters. For example, if it did indeed turn out that clusters showed enhancement of the glottal gesture compared to singletons then we wanted to be able to test whether the effect only emerged in prosodically strong contexts. Secondly, we were interested in prosodic strengthening as a phenomenon in its own right. There has been considerable interest in recent years in investigating how prosodic structure is signaled not just in contrasting intonation patterns but also in segmental changes. A fair amount of knowledge has now also accumulated on the articulatory substrate of such segmental changes (e.g. using electropalatography or electromagnetic articulography; Bombien, Mooshammer, Hoole, & Kühnert, 2010; Bombien, Mooshammer, & Hoole, 2013; Byrd & Choi, 2010; Keating, Cho, Fougeron, & Hsu, 2003). Of particular interest in the present case is the fact that, at least for English, VOT is often found to be longer in prosodically strong contexts (see e.g. Cho & McQueen, 2005, for discussion). In articulatory terms this suggests that the laryngeal gesture is enhanced in such cases, but detailed information is currently lacking. There has been a fair amount of analysis of word-stress effects using transillumination (e.g. Cooper, 1991; Fuchs, 2005; Hoole et al., 1984; Löfqvist & McGarr, 1987), giving a consistent picture of more glottal abduction in more strongly stressed syllables, but to our knowledge nothing is known about whether such findings extend to further sources of prosodic strengthening. We will focus, in particular, on the influence of contrastive focus, since this was the condition that was easiest to embed in a corpus suitable for transillumination. But we also included a condition designed to provide information on domain-initial strengthening because of the equally well-documented relevance of prosodic boundaries for articulatory strengthening processes (e.g. Cho & McQueen, 2005; Keating et al., 2003). Even if the expectation is indeed that the English findings will extend to German, and that word-stress findings will extend to phrasal stress and to boundary-induced strengthening, there still remains much scope for clarification, i.e. does the glottal gesture simply lengthen with longer VOT, or does it also increase in magnitude? Is it perhaps simply timed later, rather than lengthened? Does gestural enhancement apply equally to fricatives and plosives, even though enhancement of VOT is not relevant in the former case? The third and final reason for including prosodic variation in the experimental design was motivated by the fact that the present work forms part of a larger study investigating laryngeal articulation in languages differing in voicing typology (German, Dutch, French); we expect that the effect of prosodic strengthening should reflect different underlying organization of the voicing contrast in these languages. This will be the subject of a future report.

2. Method

2.1. *Experimental setup*

Five native speakers of German (2 female, age 26 and 28; 3 male, age 35, 35, and 26), one of them the second author, were recorded by means of photoglottographic transillumination (PGG). For this purpose a flexible endoscope (Olympus ENF-P3) connected to a Rehder rpCam-X light source and video camera was inserted through the nose into the pharyngeal cavity by a physician such that the tip of the endoscope was positioned above the glottis. After positioning, the endoscope was mounted on an adjustable positioning arm (Aesculap M-TRAC). By using this device the speakers' discomfort was minimized and the attending physician was relieved from holding the endoscope throughout the course of the entire experiment. Two phototransistors (PGG1 and PGG2) were attached to the skin just above and just below the cricoid cartilage, shielded from ambient light and connected to an amplifier. The light source served not only to illuminate the pharyngeal cavity for video acquisition but also facilitated gauging the glottal width: The light falling through the glottis is transformed by the phototransistors into an electrical potential which correlates with the glottal width. A digital data acquisition system (Sony Ex) was used to record audio (at 32,768 Hz) and PGG data (each at 8192 Hz) along with a synchronization impulse generated by the prompting software. The PGG data were downsampled to 1024 Hz and were available both in raw form and low-pass filtered (cut-off frequency 25 Hz). The latter was used for kinematic analysis since it removed the influence of phonatory movements of the vocal folds on the transillumination signal, Video was captured using a DV recorder (Sony HVR-M15AE) along with the audio and the synchronization impulse, thus facilitating synchronized trial-based video extraction. After transfer to computer the video was de-interlaced, giving a frame rate of 50 Hz. Methodological issues in the use of transillumination are discussed in detail in Hoole (1999). In particular, use of synchronized video filming is crucial in order to ensure that the transillumination signal is not impaired through shadowing of the glottis by the tongue root or epiglottis. The speech material (see Section 2.2) was presented visually on a computer screen to the subjects 6 times in randomized order. Erroneous utterances were repeated at the end of the respective block.

2.2. Speech material

While the complete list of target words includes a large number of legal onsets this study is specifically concerned with words with the initial obstruents /f ʃ p t k/ as singletons and as clusters with a following /r/ or /l/ (see Table 1). Basic considerations in the design of the material were that all carrier phrases should be very similar in length, and contain the target word in a similar utterance-medial position, and that the target consonants should be flanked by high front vowels. Ideally, we would like to have used only unrounded high front vowels, but because of gaps in the German lexicon the rounded counterpart /yː/ had to be used in two cases. As discussed in the introduction, the design also included a set of prosodic conditions. For a direct comparison with respect to prosodic strength, carrier phrases were devised to contrast a deaccented with a (narrowly) focused condition. For these utterance pairs, the sentence structure was basically identical and had the target word in phrase-medial position. However, a contrasting element at the end of the phrase caused the speakers to place contrastive focus on either the target word (focused condition) or the immediately following word (deaccented condition). The third prosodic condition used similar phonetic material but different sentence structure. A prosodic boundary is elicited by placing the target word immediately after addressing a fictitious person (Lissie). We expect this boundary condition to constitute a relatively strong prosodic context, but since, in this corpus, we were not able to explicitly contrast different boundary conditions while simultaneously controlling focus we will base statements about the effect of prosodic strength on the comparison of the deaccented and the narrowly focussed condition. Nevertheless, the inclusion of the third prosodic condition is motivated, as explained in the introduction, by the fact that we are not just interested in prosodic strengthening per se but also in prosody as a “probe” to test the robustness of the segmental and syllable-structure effects. Thus we would have considered an even wider range of prosodic variation if the duration of the experiments had not been constrained by the rather invasive nature of the procedures.

The speakers were familiarized with the different sentence types before the start of the experiment. For the post-boundary conditions, speakers were instructed to avoid pauses at the location of the boundary to prevent additional glottal activity, e.g. for respiration. The general instruction was to speak normally at a comfortable rate. Examples of the sentence types used for each prosodic condition are as follows:

(**xxx**: target; *xxx*: focus; xxx: contrast)
Deaccented/phrase-medial
“Bis sie **piep** *sieht*, nicht hört.”
“Until she *sees* **piep** instead of hearing it.”
Focused/phrase-medial
“Bis sie ***piep*** sieht, nicht Tisch.”
“Until she sees ***piep*** not table”
Post-Boundary
“Lissie, **piep** sieht nicht gut aus.”
“Lissie, **piep** doesn't look good”

2.3. Segmentation and measures

Two acoustic intervals were analyzed in the current study: C1 occlusion (OCC) and voice onset time (VOT). C1 occlusion duration was determined in the waveform and the spectrogram. For stops, a significant drop of energy and the loss of formant structure constituted the start of occlusion while

Table 1
Target words by initial obstruent and onset type. Initial /tl/ is illegal in German.

	Fricatives		Stops		
	f	ʃ	p	t	k
C	fies /fiːs/	schief /ʃiːf/	piep /piːp/	tief /tiːf/	Kies /kiːs/
Cl	Fleece /fliːs/	schlief /ʃliːf/	Plüsch /plyːʃ/	–	klüger /klyːgɐ/
Cr	Fried /friːt/	schrieb /ʃriːp/	pries /priːs/	trief /triːf/	Krieg /kriːk/

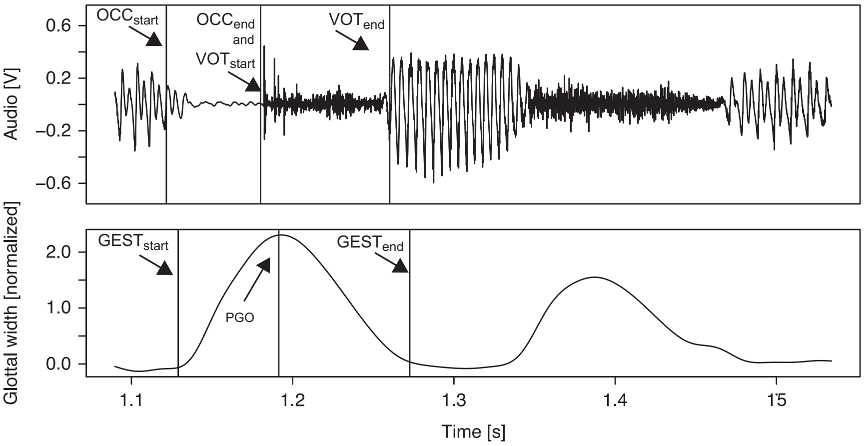

Fig. 2. Labeling of audio (upper panel) and PGG (lower panel) for a token of ‘Kies’ /kiːs/ in focused condition. OCC: oral occlusion of C1; GEST: glottal opening–closing gesture; PGO: timepoint of peak glottal opening; VOT: voice onset time.

voicing might extend into closure (as is the case in Fig. 2). The stops' release marked the end of occlusion. For fricatives, occlusion duration corresponded to the duration of frication. VOT is the interval from the end of the C1 occlusion to the onset of voicing as determined in the waveform and the spectrogram (note that as defined here VOT is equally applicable to items with either stop or fricative as C1). For the kinematic analysis we used the transillumination signal obtained from the sensor placed below the cricoid cartilage. In order to facilitate the comparison of the magnitude of the PGG signal, it was normalized by subtracting the baseline in the neighboring vowels and dividing by the average peak glottal opening in the target items in each block of repetitions. Three kinematic time stamps were analyzed for the purpose of this study: The start and end of the glottal opening–closing gesture (GLOTT) at the 20% threshold of peak abduction/adduction velocity and the timepoint of peak glottal opening (PGO). Fig. 2 provides an example of the segmentation procedure. In a few cases the epiglottis was retracted during the glottal opening cycle causing a distortion of the PGG signal and consequently a failure to detect PGO correctly. In these cases, PGO was determined from the synchronized video (the movements of the arytenoids were always clearly visible, even if the anterior part of the glottis was obscured by the epiglottis).

Based on the intervals and time-points defined above we now list the actual measures that went into the statistical analyses, briefly summarizing the main expectations or points of interest. (Strong expected effects that are not of central concern to the present study will not be mentioned in this list, but will be pointed out in the detailed results below. For example, the timing of peak glottal opening can be expected to differ radically for stops vs. fricatives.)

VOT ($=VOT_{end}–VOT_{start}$)
Defined above to apply to both fricatives and stops, and both singletons and clusters. Main expectation is for longer values in clusters and in prosodically strong conditions

Duration of C1 obstruent occlusion ($=OCC_{end}–OCC_{start}$)
Assuming that C1 occlusion duration shortens as expected in clusters the main point of interest is to what extent this accounts for any increase in VOT. Expected to lengthen in prosodically strong conditions.

Duration of the glottal gesture ($=GEST_{end}–GEST_{start}$)
One of the main motivations for the present study is that it is simply unclear what expectation to have for mixed-voicing clusters compared to singletons: constant glottal gesture over changing supra-glottal articulations, gestural strengthening to enhance voicelessness in specific clusters, or gestural shortening in parallel with C1 shortening. Expected to lengthen under prosodic strengthening (if occlusion duration also lengthens).

PGO amplitude
This is the only measure based on signal amplitude, rather than duration or timing. It complements the measure of gestural duration, particularly with a view to identifying whether gestural strengthening can in fact occur in clusters vs. singletons. Similarly, gestural strengthening related to strong prosodic conditions could lead to higher amplitudes (i.e. stronger abduction).

Timing of PGO ($=(PGO–OCC_{start})/(OCC_{end}–OCC_{start})$)
Previous investigations indicate reasonably consistently that the timing of peak glottal opening relative to the occlusion of C1 in clusters changes relative to the corresponding singleton. The specific expectation in clusters is for later timing (higher values of this measure), i.e. PGO occurs later in the occlusion phase or (quite likely in stops) a longer time after release of the C1 occlusion. Of interest is whether such effects occur for all clusters, or more strongly for specific combinations of C1 and C2.

Proportion of glottal cycle after occlusion offset ($=(GEST_{end}–OCC_{end})/(GEST_{end}–GEST_{start})$)
This is a complementary measure to the preceding one (timing of PGO), i.e. it captures the timing of an oral event (release of the occlusion) relative to a glottally defined interval, rather than the timing of a glottal event relative to an orally defined interval. In practice, the two measures are expected to show very similar patterns (the present measure can also be expected to show greater proportional values for clusters vs. singletons). However, there were some indications in Hoole (2006) that the present measure may give somewhat more stable results than the measure based on timing of PGO, perhaps because it is not dependent on a *single* time-point in the glottal gesture (also, as mentioned above, in a few cases the time instant of PGO had to be measured from the video rather than the PGG signal, perhaps giving slightly more measurement errors). In any case, in order to assess the robustness of any timing shifts from singletons to clusters it seemed important to incorporate two measures that assess similar underlying phenomena from two different points of view.

Occlusion onset to gesture onset ($=GEST_{start}–OCC_{start}$)
This measure provides the final piece of information necessary for a complete picture of laryngeal oral timing. For example, if the timing of PGO is observed to shift to a later location for clusters vs. singletons (and VOT to lengthen), this could be due to a shortening of C1 occlusion duration, to a

Table 2
Predictors for statistical models.

Predictor	Levels	Description
POA	‘p’, ‘t’, ‘k’	Place of articulation in models excluding fricative onsets. Reference level ‘p’ (labial) contrasted with ‘t’ (alveolar) and ‘k’ (velar; dummy coding)
Manner	‘fricative’, ‘stop’	Manner of articulation of the obstruent in single obstruents or obstruent+sonorant clusters. Reference level ‘fricative’ contrasted with ‘stop’ (dummy coding)
Type	‘C’, ‘Cl’, ‘Cr’	Onset type; simple obstruents (‘C’, reference level) are contrasted separately with obstruent+/l/ (‘Cl’) and obstruent+/r/ (‘Cr’) clusters (dummy coding)
Prosody	‘D’, ‘B’, ‘F’	Prosodic condition; deaccented items (‘D’, reference level) are contrasted separately with post-boundary items (‘B’) and with focused items (‘F’; dummy coding)
Repetition	(numeric)	Number of the experimental block

lengthening of the glottal gesture duration, but also to a later onset of the glottal gesture (as well as to the combined effect of all three measures). Although this measure is not expected to be the timing parameter with the most prominent changes, previous investigations indicate that it would be hazardous to neglect it.

Gesture offset to voicing onset ($=VOT_{end}–GEST_{end}$)

We use this measure to provide an assessment of aerodynamic conditions in the vocal tract at the end of the glottal gesture. More positive values of this measure indicate that the re-initiation of voicing may be delayed because the constriction in the vocal tract is still strong enough to delay the equalization of the intraoral pressure increase that will have built up during the C1 occlusion. This would represent a passive aerodynamic component in increased VOT in clusters (as opposed to an active consequence of changed timing patterns). It should be noted that it is by no means *a priori* clear that there has to be an aerodynamic component to increased VOT in clusters; in the present corpus the singleton C1 is followed by a high front vowel, which also exhibits a strong constriction, hence the necessity to assess the situation empirically.

2.4. Statistics

We used linear mixed effects models (LMEMs) to test the effect of place of articulation (POA, Section 3) as well as Manner and (onset) Type (with interaction, Section 4) on the eight measures described in Section 2.3 with subjects as random effect (random intercept) (Baayen, Davidson, & Bates, 2008) and random slopes (Barr, Levy, Scheepers, & Tily, 2013) for the effects of Manner and Type. Since the design of the experiment included three prosodic conditions the factor Prosody was included in the model. In order to account for the variability introduced by speech rate changes across the experiment, our model also included the control variable Repetition. Prosody and Repetition were checked for interactions with the main test variables, i.e. POA in Section 3 and manner and onset type in Section 4. These interactions were only retained in the model if Likelihood-Ratio-Tests indicated a significant improvement of the model fit. We report *p*-values based on Likelihood-Ratio-Tests comparing the model *with* the fixed effect in question against the model *without* the fixed effect in question. To assess the significance of pair-wise comparisons within a 3-level factor, we provide *p*-values based on normal approximation using the *t*-value of the coefficient in question. Differences between levels are given as $x \pm y$ where x represents the estimate and y the standard error as returned by the model. All predictors are explained with their levels in Table 2.

3. Results (1) – single stops

We will start the presentation of the results by looking briefly at timing patterns with respect to place of articulation in stops. It turns out that one of the coordination scenarios we outlined above as potentially applicable to both place of articulation in stops, as well as to the singleton-cluster comparison can indeed be identified in the place of articulation analysis: specifically, differences in VOT linked to differences in occlusion duration while glottal gesture duration varies very little.

This will serve to illustrate the patterns in our measurement parameters for a fairly clear case, and thus serve as a useful background for the subsequent discussion of singletons vs. clusters, where once again fairly clear VOT differences emerge, but where the coordination patterns leading to this result on the acoustic surface are less easy to identify.

Similarly, we will also briefly consider the influence of prosodic condition in single stops, since most previous work on, for example, the influence of prosodic strength on VOT has focussed on such sounds. They thus again provide a useful point of orientation before moving on to the more complex sequences.

3.1. Place of articulation

VOT is, as expected, longer in /k/ and /t/ than /p/. The difference amounts to 15.3 ms±2.9 ($t=5.3$, $p<0.001$) for /p/ vs. /k/ and to 11 ms±2.8 ($t=3.92$, $p<0.001$) for /p/ vs. /t/ (overall effect of POA: $\chi^2(2)=9.48$, $p<0.01$). What elements in the coordination patterns contribute to this expected result? Occlusion duration is, again as expected, shorter in more posterior places of articulation, the difference being 12.8 ms±4 ($t=3.19$, $p<0.01$) for /p/ vs. /t/ and 19.5 ms±5 ($t=3.89$, $p<0.001$) for /p/ vs. /k/ (overall effect of POA: $\chi^2(2)=7.07$, $p<0.05$). Fig. 3 displays the results for VOT and C1 occlusion duration. On the other hand, there is no difference in the glottal gesture duration as a function of POA (mean overall difference is 1 ms±3 for /t/ vs. /p/ and −3 ms±2 for /k/ vs. /p/; see left panel of Fig. 4) nor in the magnitude of peak glottal opening. Thus although increase in VOT does not trade-off one-to-one with reduction in oral occlusion duration it is clear that the latter is a major component in the differences in VOT.

The interval from occlusion onset to glottal gesture onset showed no effects related to place of articulation (this is a first example of a parameter that will, however, warrant more detailed consideration with respect to prosody and to singletons vs. clusters).

We also mention here merely in passing the parameter with which we test for the relevance of aerodynamic conditions at release of the occlusion, namely the interval from end of the glottal gesture to onset of voicing: this showed no consistent differences between /p/ and /k/ or /t/ (overall mean differences in the range of 1–2 ms), but will again turn out to be more relevant for the influence of prosody and for the comparison of singletons and

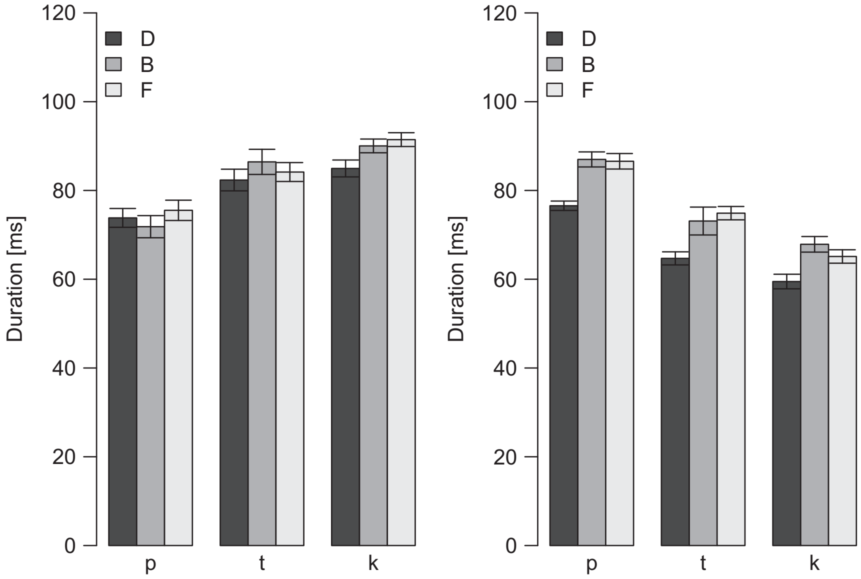

Fig. 3. VOT (left panel) and occlusion duration (right panel) in stops as a function of place of articulation and prosodic condition, averaged across speakers.

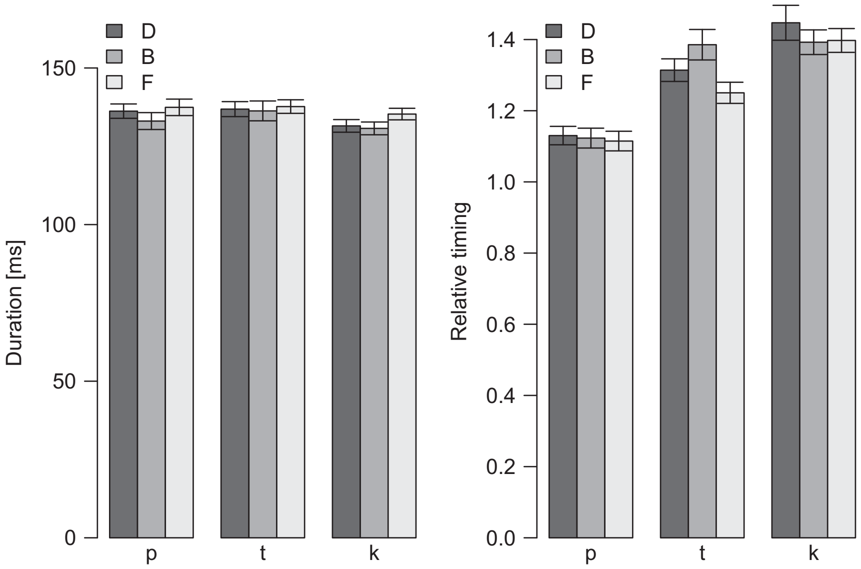

Fig. 4. Gesture duration (left panel) and PGO timing relative to occlusion (right panel) in stops as a function of place of articulation and prosodic condition, averaged across speakers. In the right panel, values>1 indicate that PGO occurs after release of the oral occlusion.

clusters below. (Nevertheless, it confirms the estimate in Cho and Ladefoged (1999, pp. 213–214) that by the time the glottis closes for an aspirated plosive some 50 ms or more after release of the oral occlusion then it is unlikely that place-of-articulation-related differences in the time-course of the release phase will still be influencing aerodynamic conditions in the vocal tract.)

Given the durational patterns listed above it is not surprising that the two parameters that we use to capture relative timing of peak glottal opening and release of the oral occlusion show clear differences. The timing of PGO relative to the occlusion (see right panel of Fig. 4) has a value of about 1.1 for /p/, i.e. PGO is more or less synchronous with the release, but is about 0.2±0.05 higher for /t/ ($t=4.07$, $p<0.001$) and 0.3±0.07 higher ($t=4.42$, $p<0.001$) for /k/, i.e. PGO for /t/ and /k/ is typically not reached until sometime after the release of the oral occlusion (overall effect of POA: $\chi^2(2)=8.38$, $p<0.05$). The complementary measure based on the relative position of the stop release within the glottal cycle shows an identical pattern ($\chi^2(2)=12.25$, $p<0.01$). Approximately 60% of the glottal gesture remain after the release of /p/ but 11%±1 more for /t/ ($t=9.23$, $p<0.001$) and 14%±2 for /k/ ($t=5.67$, $p<0.001$), i.e. the release of the occlusion occurs earlier within the glottal cycle for /t/ and /k/. This indicates that if speakers' representation of laryngeal–oral coordination for stops is, following Browman and Goldstein's (1986) rule above, in terms of PGO and release, then, as indicated in the introduction (Cho & Ladefoged, 1999), the rule is implemented in a context-specific manner. However, there are still grounds for caution: to prove or disprove a specific phasing relation we require access to the dynamical state of both the laryngeal and oral articulators, which we do not have here. At the very least, we would additionally require the position and velocities of the oral articulators in order to be able to view relative timing in a phase-plane representation (e.g. Kelso & Tuller, 1985).

In any case, the results for place of articulation in stops thus serve already to confirm that at least one of the coordination scenarios brought into play in the introduction, namely change in oral occlusion duration relative to a rather constant glottal gesture duration, is indeed one potentially realistic candidate for laryngeal–oral coordination in general.

3.2. Prosodic effects

Turning to the influence of the prosodic conditions, for VOT there was a significant main effect along the lines expected from the literature, specifically with a longer VOT in the focussed than in the deaccented condition ($p<0.01$), but the difference between these two conditions was actually not very large (approx. 3.5 ms±1.1; see left panel of Fig. 3).

For occlusion duration the findings were completely in line with expectations for prosodic strengthening, and rather greater in magnitude than the VOT results, i.e. a difference of approx. 8.5 ms±1.1 for the focussed vs. deaccented comparison ($p<0.001$; see right panel of Fig. 3).

Given this situation in which both oral occlusion and VOT lengthen, there is a clear expectation that glottal gesture duration should also lengthen under prosodic strengthening. This, however, was not the case ($p>0.05$; see left panel of Fig. 4). Why then does VOT lengthen, at least slightly? In the light of the above discussion of place-of-articulation effects one could even ask why it does not shorten, given the longer occlusion duration but same glottal gesture duration. There are two main reasons: (1) The glottal gesture starts somewhat later (about 5.5 ms ±1; $p<0.001$) relative to occlusion onset in the focussed vs. the deaccented condition (this corresponds to the "late glottal gesture" scenario in Fig. 1). (2) The time lag from glottal gesture offset to voice onset is greater by about 4.5 ms±1 for focussed vs. deaccented. This is similar to the effect we illustrated with the scenario labeled "aerodynamics (+short C1)" in Fig. 1, but is here more likely to be an effect of increased vocal fold tension in the focussed condition (i.e. for a given adductory state of the glottis, stiffer vocal folds will delay phonation onset, just as increased intraoral pressure will do). This interpretation is reinforced by the clearest finding of all related to prosodic conditions: Even though the duration of the glottal gesture does not change there is a substantial increase in the magnitude of glottal opening for focussed vs. deaccented (approx. 0.37±0.05, $p<0.001$). Hoole (2006) argues that increased cricothyroid activity (lengthening the vocal folds and raising F0) also increases the efficiency of the abductory movements at the arytenoids (in terms of change in glottal area).

In short, the expected acoustic result of longer occlusion and VOT turns out to be linked to quite a subtle combination of glottal adjustments.

4. Results (2) – full material

In this section we will first look one by one at the detailed results for each of the parameters outlined in Section 2, and then draw the threads together by summarizing the patterns for each of the main cluster categories (fricative+/r/, fricative+/l/, plosive+/r/, plosive+/l/) relative to the corresponding singletons. As pointed out in Section 2.4, the main test variables in this section are Manner and Type and their interaction. The factor Prosody was already introduced in the previous section, and is equally relevant to the present section. Effects of Repetition often turned out to be highly significant. However, the corresponding estimate was often very low especially in comparison to the effects observed for the main test variables. We will therefore only report effects of Repetition when they are clearly relevant for explaining the data.

4.1. Voice onset time

Manner affects VOT ($\chi^2(3)=28.75$, $p<0.001$) such that it is 51.7 ms±4.1 longer for stops than for fricatives. This is a completely expected result given the categorically different patterns of laryngeal–oral coordination for fricatives and aspirated plosives. Of more interest is the fact that Type also has a highly significant influence on VOT ($\chi^2(4)=25.42$, $p<0.001$), with overall longer VOT in clusters compared to singletons (see Fig. 5). However, there is also a significant interaction between Manner and Type ($\chi^2(2)=6.81$, $p<0.05$) since the effect of type is greater for fricatives than stops. For fricative C1, compared to single obstruents VOT is 30.4 ms±4.1 longer in /Cl/ sequences ($t=7.5$, $p<0.001$) and 39.9 ms±5.2 longer in /Cr/ sequences ($t=7.72$, $p<0.001$). For stops, the clearest significant difference is that VOT is 24.7 ms±1.5 longer in /Cr/ than in /C/ ($t=16.41$, $p<0.001$). For clusters with /l/ Fig. 5 shows a small increase compared to the singleton stop over all prosodic conditions. However, the overall effect was not, in fact, significant: while the difference between /p/ and /pl/ amounted to almost 20 ms (significant at $p<0.01$) the difference between /k/ and /kl/ was effectively zero. Regarding Prosody, even though Fig. 5 indicates slightly longer VOT for the focussed compared to the deaccented condition for the stop-initial sequences (there is clearly no trend at all in the fricative-initial case) the main effect of Prosody was not in fact significant (nor was the interaction with manner). This was slightly surprising following the significant effect for singleton stops in Section 3.2 (and in the light of typical findings in the literature), but it will be recalled that in Section 3.2 the difference while significant only amounted to a few milliseconds. So in this study the effect of Prosody on VOT does seem to be quite weak.

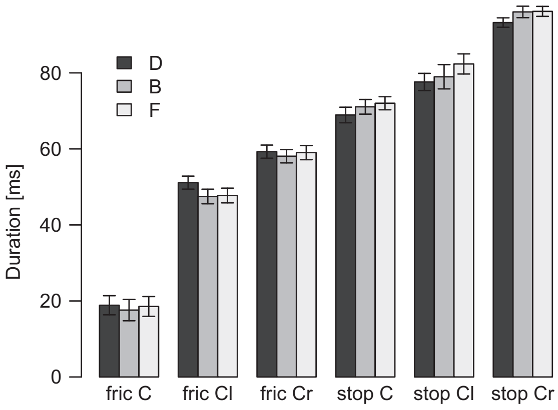

Fig. 5. Voice onset time as a function of manner, onset type and prosodic condition averaged across speakers.

4.2. *Obstruent occlusion duration*

Manner is a significant factor in the model ($\chi^2(3)=20.05$, $p<0.001$) with occlusion duration of stops 50.2 ms±5 (standard errors) shorter than that of fricatives. The main effect for Type ($\chi^2(4)=9.81$, $p<0.05$) accounts for a decrease of occlusion duration in fricative+sonorant cluster compared to single fricatives: /Cl/: 28.8 ms±5.6 (standard errors; $t=-5.13$, $p<0.001$); /Cr/: 23.1 ms±5.5 (standard errors; $t=-4.2$, $p<0.001$). The interaction of Manner and Type ($\chi^2(2)=9.29$, $p<0.01$) and the associated estimates indicate that no such effect is present for stops. Prosodic variation significantly influences the occlusion duration ($\chi^2(2)=100.92$, $p<0.001$). Compared to the deaccented reference level, occlusion is longer by 9.5 ms±1.2 (standard errors; $t=8.11$, $p<0.001$) under focus, i.e. very similar to the difference found when just considering single stops (8.5 ms) The results are visualized in Fig. 6.

4.3. *Duration of the devoicing gesture*

The duration of the devoicing gesture is affected by Manner ($\chi^2(3)=9.24$, $p<0.05$) in that it is 17.9 ms±3.8 longer for fricatives than for stops. The main effect of Type is in fact weakly significant ($\chi^2(4)=11.4$, $p<0.05$) but the more salient feature here is the complicated pattern of interactions between Manner and Type ($\chi^2(2)=5.77$, $p<0.05$). Accordingly we find shorter devoicing gestures for fricative+/l/ compared to the singletons (13.8 ms ±2.7, $t=5.18$, $p<0.001$), but no difference for fricative+/r/ while for the stops, stop+/r/ is *longer* than the singletons (11.5 ms±3.8, $t=3$, $p<0.01$), but stop+/l/ shows no difference. In spite of the significance of the effects observed in this section, the test statistics and visual inspection indicate that they are weaker than the results in Section 4.2. Moreover, there is no significant effect of Prosody on duration of the devoicing gesture. Thus, rather as we indicated at the end of the introduction, even though it would not be justifiable to regard glottal gesture duration as fixed in a hard and fast sense, it is noticeable (taking these results together with those for place of articulation) that occlusion durations vary over a much wider range than do glottal gesture durations (see Fig. 7).

4.4. *Magnitude of peak glottal opening*

The model reveals that for the magnitude of peak glottal opening (PGO) Manner does not seem to play a crucial role. Type has a significant effect on the PGO magnitude ($\chi^2(4)=9.91$, $p<0.05$) as does the interaction of Manner and Type ($\chi^2(2)=9.98$, $p<0.01$). For fricative obstruents, PGO magnitude is 0.25±0.11 smaller in /Cl/ than in /C/ ($t=2.05$, $p<0.01$) while there is no effect for /Cr/ clusters. For stops, on the other hand, the model revealed (after releveling the Manner factor) a significant increase of PGO magnitude by 0.36±0.07 in /Cr/ compared to /C/ ($t=5.09$, $p<0.001$). A very important observation is that Prosody has an immense impact in the data ($\chi^2(2)=90.7$, $p<0.001$). Compared to the reference level (deaccented), PGO magnitude for the focused condition shows an increase of 0.34±0.04 (standard errors, $t=9.46$, $p<0.001$). Once again, this closely follows the finding for single stops. Moreover, since there were no interactions of Prosody with Manner or Type, it would appear, firstly, that prosodic strengthening

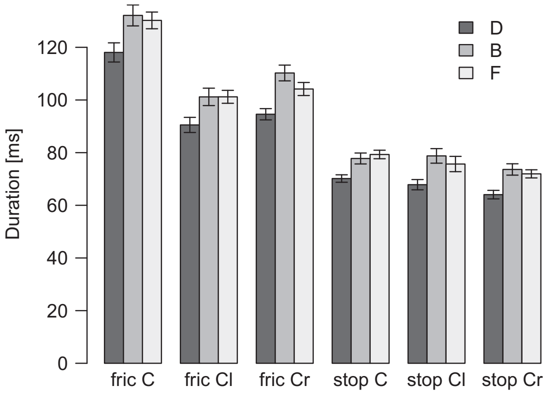

Fig. 6. Obstruent occlusion duration as a function of manner, onset type and prosodic condition averaged across speakers.

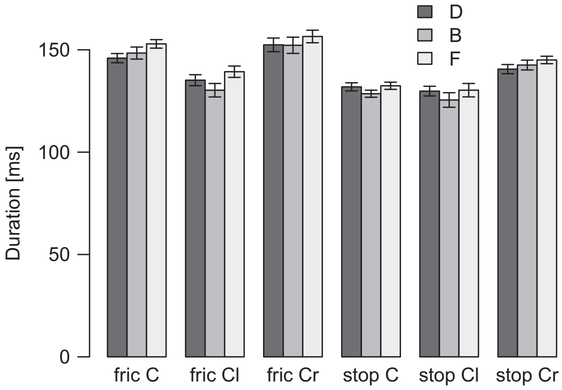

Fig. 7. Duration of the devoicing gesture as a function of manner, onset type and prosodic condition averaged across speakers.

causes a similar increase for fricatives and stops, and secondly, that the increase in glottal opening for /Cr/ vs. /C/ is not confined to strong prosodic locations. These results are graphically presented in Fig. 8.

4.5. *Timing of peak glottal opening*

Trivially, Manner significantly affects the relative timing of PGO ($\chi^2(3)=20.14$, $p<0.001$) in that PGO occurs later relative to the occlusion interval in stops than in fricatives (0.58±0.04). For fricatives, PGO occurs around the center of the occlusion interval while for stops it occurs after the release of the occlusion. Of more interest in the present context is once again the significant effect for Type ($\chi^2(4)=18.8$, $p<0.001$), due to the fact that /Cr/ shows higher values than /C/. But the interaction of Manner and Type ($\chi^2(2)=12.98$, $p<0.01$) is also significant, reflecting the fact that the increase from /C/ to /Cr/ is more substantial in stops than fricatives (though significant in both cases): a delay of PGO for /Cr/ relative to /C/ by 0.35±0.035 ($t=9.87$, $p<0.001$) for stops and by 0.12±0.03 ($t=4.58$, $p<0.001$) for fricatives. Timing of PGO for /Cl/ is not significantly affected. There were no significant effects involving Prosody. Fig. 9 displays the relevant results.

4.6. *Proportion of glottal cycle after occlusion offset*

As described in Section 2.3, this measure describes the proportion of glottal gesture after the offset of the oral occlusion. The results are expected to closely parallel those just presented for timing of peak glottal opening. In this model, fitting interaction with Prosody and Repetition was meaningful. Further inspection revealed that it was sufficient to model two-way interactions of Manner and Type, Manner and Prosody, and Type and Repetition. This entails an increase of the degrees of freedom in Likelihood-Ratio-Tests compared to the models in the previous sections. As in the previous section, the essentially trivial main effect of Manner ($\chi^2(5)=35.37$, $p<0.001$) estimates that 0.43±0.036 more of the glottal cycle remain after the occlusion offset of stops than of fricatives. There is also a main effect for Type ($\chi^2(6)=26.2$, $p<0.001$) which accounts for larger values of this proportion parameter (by 0.17±0.029, $t=5.74$, $p<0.001$) in the case of /Cr/ vs. /C/ onsets. This closely parallels the principal non-trivial finding in the previous section. The main effect of Prosody ($\chi^2(4)=37.01$, $p<0.001$) and the interaction of Manner and Prosody ($\chi^2(2)=15.81$, $p<0.001$) indicate an influence of prosodic contexts on the timing of occlusion offset but only for fricatives: Compared to the deaccented reference level, the proportion of the glottal cycle after occlusion release is 0.048±0.016 smaller in the focused condition. For a display of the results see Fig. 10.

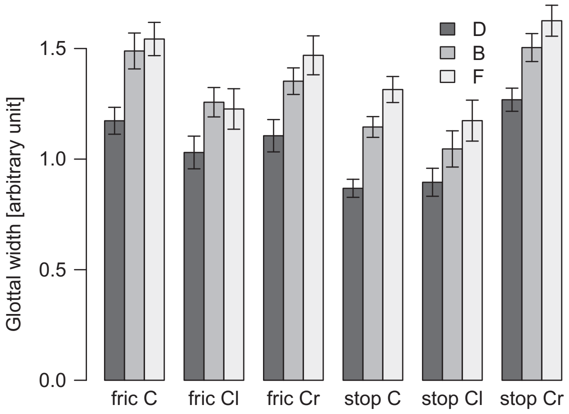

Fig. 8. Magnitude of PGO as a function of manner, onset type and prosodic condition averaged across speakers.

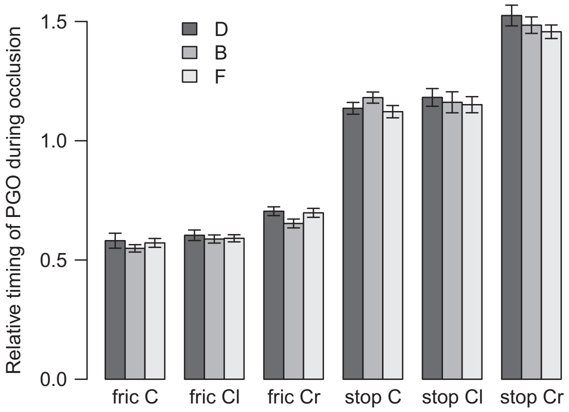

Fig. 9. Relative timing of peak glottal opening as a function of manner, onset type and prosodic condition averaged across speakers. A value of 0.5 indicates PGO timed at the center of the occlusion phase. Values > 1 indicate PGO located after release of the occlusion.

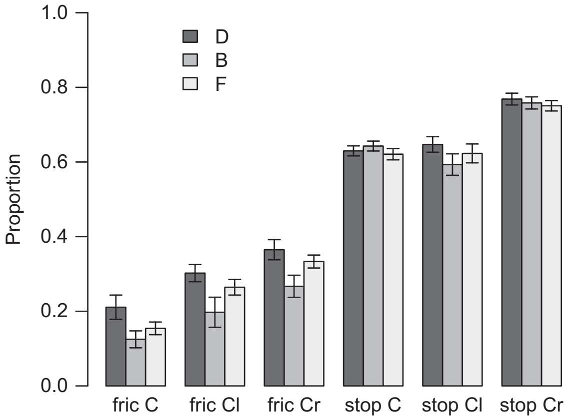

Fig. 10. Proportion of glottal cycle after occlusion offset as a function of manner, onset type and prosodic condition averaged across speakers.

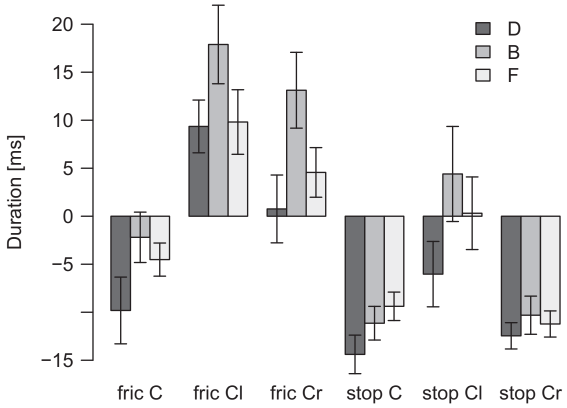

Fig. 11. Gesture offset to voice onset interval as a function of manner, onset type and prosodic condition averaged across speakers.

4.7. Gesture offset to voice onset

This measure aims to capture whether adding a sonorant to a voiceless C onset delays the onset of voicing relative to the devoicing gesture, i.e. whether any supraglottal constriction still remaining at the termination of glottal adduction results in aerodynamic conditions that are less conducive to voicing initiation. Manner: ($\chi^2(3)=11.15$, $p<0.05$) voice onset occurs after gesture offset for fricatives but before for stops (6.26 ms±3). The main effect of Type ($\chi^2(4)=10.6$, $p<0.05$) indicates that compared to simple /C/ onsets voice onset is delayed by 17.7 ms±4 ($t=4.51$, $p<0.001$) in /Cl/ and by 11.6 ms±4.4 ($t=2.61$, $p<0.01$) in /Cr/ clusters. However, as is clearly evident in Fig. 11, the interaction of Manner and Type ($\chi^2(2)=6.12$, $p<0.05$) appears to play an even more crucial role here than for the measures reported so far. A model with the factor Manner releveled to have 'stop' as reference level reveals that there is indeed no Type effect in stop+/r/ clusters, while voice onset is delayed by 10.1 ms±4.3 in stop+/l/ clusters which is a less clear change than in fricative+/l/ clusters. Prosody has a highly significant effect on the gesture offset to voice onset interval ($\chi^2(2)=28.12$, $p<0.001$). Compared to the deaccented condition, voice onset is delayed by approximately 3.9 ms±1.3 ($t=2.95$, $p<0.01$) in focused onsets. This might possibly reflect an influence of increased vocal fold tension in the prosodically strong conditions, as mentioned above in the presentation of the results for single stops (where the effect was very similar). A significant effect of Repetition ($\chi^2(1)=53.17$, $p<0.001$) is due to a trend for this measure to decrease during the course of the experiments.

4.8. Occlusion onset to gesture onset

The interval from occlusion onset to the onset of the devoicing gesture is the final timing parameter that we need to consider in order to have a full understanding of what lies behind the shifting patterns of laryngeal–oral coordination already captured in the analyses of Sections 4.5 and 4.6. The most obvious effect (see Fig. 12) was the main effect of Manner ($\chi^2(5)=35.94$, $p<0.001$): Glottal abduction occurs approximately 22.1 ms±2.8 later relative to occlusion onset in stops than in fricatives. This is entirely expected from previous investigations, and, as discussed further in e.g. Hoole (2006), is probably one of the most stable features of laryngeal–oral coordination ever encountered. This reflects the fact that in fricatives prompt glottal abduction is essential to ensure that the main pressure drop in the vocal tract is at the supraglottal constriction and not at the glottis. A counterpart to this is that the precise time at which glottal abduction starts is much less crucial or constrained in stops. For voiceless stops, voicing will typically die out quite rapidly even without much glottal abduction (the more crucial aerodynamic task for stops involves the release phase). If the precise timing of the start of glottal abduction is indeed less crucial for stops than fricatives then this should predict more variability in the present parameter for stops. Looking at the variability within conditions in Fig. 12 there is no clear evidence of this. However, across conditions it is clear from Fig. 12 that stops vary much more than fricatives, in other words stops are much more affected by onset type and prosodic condition. This is reflected in clearly significant interaction terms in the statistics: Manner × Type: $\chi^2(4)=18.78$, $p<0.001$; Manner × Prosody: $\chi^2(2)=15.36$, $p<0.001$.

The main effects of Type ($\chi^2(4)=12.3$, $p<0.05$) and Prosody ($\chi^2(4)=77.49$, $p<0.001$) are in fact significant but given the strong interactions just noted it is only interesting to consider their effects in detail for the stops (for fricatives the effects are negligible). Regarding timing in the clusters, the

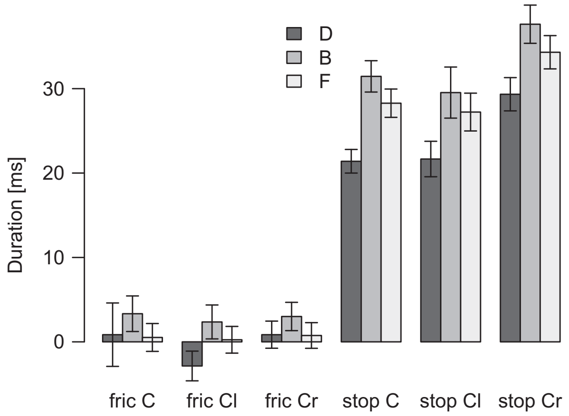

Fig. 12. Occlusion onset to gesture onset interval as a function of manner, onset type and prosodic condition averaged across speakers.

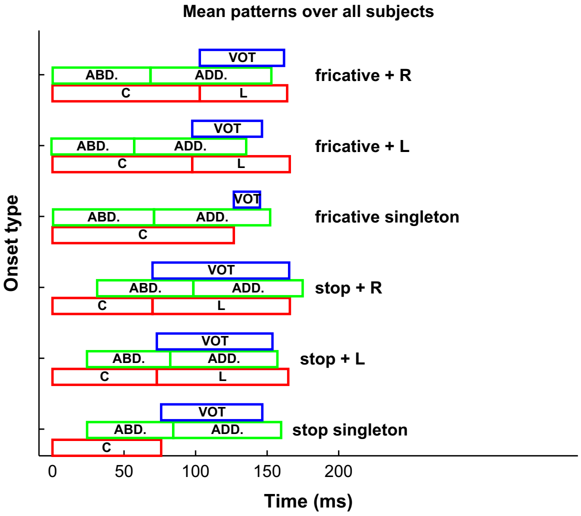

Fig. 13. Schematic illustration of timing patterns for main onset categories averaged over all subjects. As in Fig. 1 each onset type consists of three horizontal bars with lowest for oral timing, middle for glottal activity and top for VOT. In the oral timing bars 'C' stands for fricative or stop C1, 'L' stands for lateral or rhotic C2 (other details as for Fig. 1).

most interesting effect is the small (6.6 ms±2.3, $t=2.82$, $p<0.01$) but significant delay of start of abduction in stop+/r/ compared to the singleton, which thus contributes to the clear shift in relative timing observed in Sections 4.5 and 4.6. The effect of Prosody in stops manifests itself as follows: compared to the deaccented condition, glottal abduction is delayed by 5.6 ms±1 ($t=4.8$, $p<0.001$) in focused onsets. Note that in stops the effect of prosody is very similar for the clusters and singletons.

4.9. Summary

Since all the clusters showed an increase in VOT with respect to the corresponding singleton, a convenient way of summarizing the results is to consider for each cluster category what were the principal temporal adjustments (relative to the singleton) leading to the increased VOT. The relevant temporal patterns are also shown schematically in Fig. 13 for the plosive and fricative singletons together with the four cluster types discussed below. This figure forms a counterpart to the illustration of possible timing scenarios given in Fig. 1, showing with the same style of display the empirically observed patterns.

4.9.1. fricative+/l/

C1 occlusion duration shortens fairly clearly, but glottal gesture duration also shortens somewhat (and the amplitude of peak glottal opening is smaller). So there are only small effects on %gesture after occlusion release, and no effect on position of PGO relative to the occlusion. The results for the interval from end of the glottal gesture to voice onset indicate that aerodynamic conditions are a plausible component in increased VOT for the cluster. The overall scenario seems to be somewhere between occlusion shortening keeping a constant glottal gesture, and occlusion shortening with the glottal gesture shortening in parallel.

4.9.2. fricative+/r/

There is clear shortening of C1 occlusion duration, while the glottal gesture duration is similar to or even slightly longer than in the singleton (though there does not appear to be any increase in the amplitude of glottal opening). Given these findings it is not surprising that clear shifts were found in the relative timing measures, i.e. for the cluster higher %gesture after release of the occlusion, and shift of PGO position towards the release of C1. While there is thus clear evidence for the contribution of the temporal organization to the increased VOT the aerodynamic conditions also appear to be relevant (though perhaps rather less than for fricative+/l/).

4.9.3. plosive+/l

The first point to note is that while VOT does increase, this increase is weaker than for plosive+/r/ (and also weaker than for the fricative-initial clusters). No change is found in occlusion duration, gesture duration (or in glottal opening amplitude), and accordingly no changes in any of the relative timing parameters are found. Thus the sonorant does indeed appear here to be irrelevant to the planning of laryngeal–oral timing. As indicated by the difference between singleton and cluster with respect to the interval from end of glottal gesture to onset of voicing, it seems that the slight increase in VOT can be completely accounted for by aerodynamic conditions. Just one caveat should be made here: these were the only target sounds in a rounded vowel context (refer back to Table 1), so we cannot completely rule out the possibility that some unidentified timing effect related to rounded vowel in the cluster vs. unrounded vowel in the corresponding singletons cancels out timing shifts that would actually be present in the consonants themselves.

4.9.4. plosive+/r/

Only very slight shortening of C1 occlusion duration was observed, but there was some increase in gesture duration, as well as a slight delay in start of the abduction movement (and quite a clear increase in glottal opening amplitude). Taken together the occlusion shortening and gestural lengthening lead to clear changes in the relative timing parameters. Thus there can be assumed to be substantial glottal opening for a fairly extensive period during the articulation of the rhotic. Since the plosive+/r/ and plosive+/l/ clusters thus behave rather differently, this could be taken as evidence for active planning of a voiceless fricative realization of /r/. The reason why increase in glottal opening from singleton to cluster is more marked for the plosive+/r/ cluster than for the fricative+/r/ cluster may be because the amplitude for the singleton plosive is somewhat below the level required for optimal generation of frication noise (according to calculations in Stevens (1998), the opening at the glottis should be roughly double that at the oral constriction to maximize the strength of the supraglottal frication source).

4.9.5. Effects of prosodic context

A number of measures presented in Section 4 are clearly influenced by prosodic context. We concentrated on the deaccented and focus condition, because this provides the most direct comparison of prosodic strength, but in fact boundary strength consistently affected the given measures in the same direction as the focus condition, even if not necessarily to quite the same extent. As expected (e.g. Bombien et al., 2010; Cho & Keating, 2009; Cho & McQueen, 2005), obstruent occlusion duration undergoes prosodic lengthening under narrow focus and in post-boundary position (see Section 4.2). Two other results are directly linked to this finding given that the duration of the devoicing gesture appears to be stable across prosodic context: The interval between occlusion onset and onset of the devoicing gesture (Section 4.8) in stops is longer in the strong contexts while in fricatives it appears to be fixed around 0 ms, i.e. the onsets of occlusion and of the devoicing gesture occur simultaneously. The proportion of the glottal cycle after occlusion release (Section 4.6), on the other hand, is smaller in the prosodically strong contexts for fricatives but remains stable for stops in this respect. A possible interpretation of this pattern is that for fricatives there are strong timing specifications for the onset of the occlusion to the onset of the devoicing gesture. For stops, on the other hand, it is the offset of the occlusion which is timed to a certain point during the glottal cycle such that 60–75% (depending on stop place of articulation and onset type) of the glottal cycle remain.

A further measure that varies as a function of prosodic context is the interval from gesture offset to voice onset (Section 4.7). Under narrow focus but especially in post-boundary position voice onset is delayed with respect to the offset of the devoicing gesture. This is, as mentioned above, just possibly a consequence of increased vocal fold tension.

Finally, the magnitude of the peak glottal opening provides evidence for prosodic strengthening (as opposed to lengthening) in that glottal width is larger under narrow focus and in post-boundary position. However, there is no way of telling whether this effect is in some way relevant for the listener from the data presented here. Therefore, this finding should be revisited in a future study which also addresses acoustic measures of frication degree and burst intensity (this seems particularly relevant because one expected acoustic effect of prosodic strengthening, namely lengthening of VOT, was only very weakly present). Cho and Keating (2009), for example, investigated RMS burst energy and spectral center of gravity (COG) in word-initial English /t/ in various prosodic contexts. Both measures tended to be enhanced under prominence. The authors propose an account which is based on faster tongue kinematics under prominence: Faster stop release has been reported to yield higher burst energy (e.g. Stevens, 1998). Conceivably, the increase of PGO magnitude under prominence reported here might be viewed as an additional factor contributing to the burst energy. Results for positional effects in Cho and Keating (2009), however, are more complicated. While burst intensity was rather consistently *smaller* in initial than in medial position, COG was more prone to speaker variability, undergoing initial strengthening in one group of speakers and weakening in another. It is not clear how these results can be reconciled with larger PGO amplitude in initial vs. medial position as found in the present study.

Since magnitude of peak glottal opening was the parameter most clearly affected by prosodic conditions it is worth recalling here explicitly the absence of interactions with factors Manner and Type. Picking up on the questions raised in the introduction, this gestural enhancement thus applies equally to fricatives and stops, even though one might have expected it to apply more clearly to the latter, since (in German) aspiration is a more salient feature of stops. Since Prosody and Type did not interact, this indicates that the increased gestural magnitude for stop+/r/ clusters is robust, in the sense that it is not restricted to specific prosodic conditions.

In any case, the present results form a useful basis for future work. The finding of a larger magnitude of glottal abduction under prosodic strengthening is not really unexpected for German. The interesting issue now is how German will compare with languages such as Dutch and French in which the status of features such as [spread glottis] is generally assumed to be radically different. This will be the subject of a future report.

5. Discussion

In this discussion section we will first pull together the results related to prosodic conditions (as presented in Section 3.2 for single stops and just summarized in Section 4.9.5 for the full material including clusters and fricatives) and then move on to the overall more complex pattern of results that emerged from the comparison of different syllable onsets.

At a very basic level, the clearly greater glottal opening in the strong prosodic conditions appears to indicate that this study extends to the laryngeal level the substantial evidence for prosodically-driven articulatory strengthening that has previously been extensively documented for supraglottal articulators. As pointed out in the introduction, this result is not entirely surprising given that previous work had provided evidence for the effect of word-stress. But as far as we are aware, this is the first time that phrasal prominence and domain-initial strengthening have been explicitly considered with regard to laryngeal kinematics. At the same time, we believe there is room for debate as to how active the laryngeal adjustment in the strong contexts actually is. Recall that no evidence was found that the laryngeal gesture lengthens in the prosodically strong conditions. As already discussed above, this is surprising given that occlusion duration of C1, as expected, lengthens very clearly, and VOT in the stop contexts also lengthens, albeit very slightly. Although in such a situation gestural lengthening might seem the most natural laryngeal strategy to accompany overall strengthening in the syllable onset, in fact the main timing difference was that in stops laryngeal abduction was initiated later, and we suggested that the weak increases in VOT might be accounted for by the increased vocal-fold tension associated with intonational prominence (making it more difficult for the vocal folds to resume vibration). This in turn raises the possibility that the greater glottal opening area is simply a by-product of lengthening of the vocal folds as a result of increased cricothryoid activity for intonational prominence. In other words, a given amount of separation of the arytenoids (i.e. the actual abductory movement) might lead to a greater change in glottal area (i.e. roughly what transillumination actually measures) when the vocal folds are longer. Testing this will be methodologically somewhat tricky because it really requires analysis of image data of the larynx to determine the relationship between arytenoid separation and total glottal area over a range of intonational patterns. The other main avenue for further work has already been mentioned above, and links in here quite well: it would be interesting to determine whether the increased glottal area under prominence leads to measurable acoustic differences in the frication noise (for fricatives) and the burst (for stops) that could facilitate recovery of the underlying prosodic structure by the listener. If the results of this are negative (i.e. no acoustic differences) then at least we already have a potential explanation for what would otherwise be a puzzling situation: Why should speakers open the glottis wider if this provides no useful modulation of the acoustic signal? This is less unexpected if the wider opening falls out at no extra physiological cost from the laryngeal adjustments required for control of the intonation contour.

Turning now to the results gained from systematic variation of the segmental structure of the syllable onset, the patterns for each cluster category just summarized at the end of the results sections (Sections 4.9.1–4.9.4) make it clear that of the laryngeal–oral coordination scenarios presented in the introduction there is no single one that applies across the board. The situation for clusters vs. singletons thus appears more heterogeneous than that of place of articulation within single stops (Section 3.1). In both cases we have differences in VOT to account for, but while for place of articulation effects one clear pattern emerged with differences in occlusion duration overlaid on a fairly constant glottal gesture, for the various cluster-singleton comparisons the role of occlusion shortening in the cluster ranged from clearly present to negligible, and moreover combined in varying ways with changes in glottal gesture duration.

Given this situation the task now is to consider what more general conclusions can nonetheless be pulled out of the results. Based on our earlier work (Hoole, 2006) we advanced the idea that varying coordination patterns can be regarded as fulfilling an acoustically or auditorily defined goal of actively ensuring increased VOT in the clusters (as outlined in the introduction the common feature was an increase in VOT, but the articulatory patterns leading to this varied over contexts and subjects, e.g. occlusion shortening, longer and/or later glottal gesture). In general terms, this point of view was not confirmed in the more extensive and balanced material available in the present investigation. For example, plosive + /l/ clusters exhibited a null-scenario, which we had not even explicitly envisaged in the introduction (but note the caveat just made in Section 4.9 about the use of rounded vowels in this context), in which the fairly restricted increase in VOT can probably be accounted for as a passive effect of the aerodynamic conditions in the lateral following C1. However, this point of view may still apply to one specific area, namely the different behavior of rhotic and lateral clusters, which is perhaps the most interesting pattern in the results.

The fact that the rhotic clusters unlike the lateral clusters, show clear shifts in the coordination patterns relative to the singletons (as expressed in relative position of PGO and of release of C1) suggests that speakers are actively selecting adjustments (and in fact slightly different ones for fricative + /r/ and plosive + /r/) that will ensure a substantial period of voicelessness following release of C1. Note that they may even be aiming to ensure substantial frication noise during the rhotic: if the glottal amplitude measurements with transillumination can be trusted then it is intriguing that specifically the amplitude increases from plosive to plosive + /r/ (as mentioned above, for fricative + /r/ we assume that sufficient glottal amplitude for good frication may already have been reached in C1). As can be seen in Fig. 13 note also that glottal adduction is completed late relative to the end of the C2 constriction for plosive + /r/; this is a further articulatory adjustment that would ensure a large glottal aperture throughout the dorsal constriction phase of the rhotic. Overall, this account thus supports the observations originally made by Jessen (1997, 1999) regarding larger-magnitude glottal gestures in rhotic clusters.

The results for place of articulation in stops as well as for the rhotic clusters indicate that statements about laryngeal–oral coordination cannot simply relate peak glottal opening to a specific supraglottally-defined time-point (whether release for plosives, or mid-frication for fricatives) as in the rules proposed by Browman and Goldstein (1986). This indicates that the idea of context-sensitive implementation of laryngeal timing put forward by Cho and Ladefoged (1999) with respect to place of articulation effects could well be extended to take into account the nature of the sounds adjacent to the voiceless occlusives. Now, Goldstein (1990) has acknowledged that PGO can be perturbed away from its ‘canonical’ location when segments compete for control of the glottal articulator (as in clusters like /sp/). The interest of the present situation is that it is not obvious that it is desirable to assume such competition here. To do so would appear to imply that /r/ in e.g. /kr/ is underlyingly specified as voiceless (here specifically a voiceless uvular fricative), which would break up obvious symmetries in the German (and many other) sound systems (/kl, gl, kr, gr/), and introduce highly marked syllable onsets such as [ʃχ, fχ, pχ].

Rather than assuming such a categorical change in the representation of a specific segment we believe that it is more fruitful to view the glottal gesture for the voiceless plosive or fricative as being planned in a context-sensitive manner to ensure sufficient contrastiveness over a wide range of syllable onsets. We furthermore believe that our recent work on articulatory coordination in supraglottal articulations provides evidence that this point of view is plausible. In Hoole et al. (2013) we used EMA to study overlap patterns of the principal articulators of C1 and C2 in clusters such as /pr/ vs. /pl/, /br/ vs. /bl/ and /fr/ vs. /fl/. A consistent pattern of lower overlap in the rhotic clusters was found. We argued that the driving force for the low overlap pattern may be the voicing requirements for clusters with *voiced* C1 such as /br/. Using articulatory synthesis we generated /br/ sequences with more overlap between the labial and the (dorsal) rhotic than is typically observed in practice. Strikingly, the main effect of increased overlap was that the duration of voicelessness at the release of /b/ increased, making it sound rather more like /pr/. This is probably due to the fact that the formation of a dorsal constriction very soon after the release of the labial one results in only a very slow decay in the increased intraoral air-pressure in the region above the glottis, making initiation of phonation aerodynamically difficult (phonation usually ceases during German phonologically voiced stops). By timing the dorsal constriction somewhat later, it becomes possible to re-initiate phonation before the disadvantageous effects of this constriction fully come into play. Whether this overlap pattern simply generalizes to rhotic clusters with voiceless C1 (such as /pr/), or whether there are other currently unknown reasons favouring low overlap in rhotic clusters, the interesting point is that the effect in the voiceless clusters is precisely opposite to that in the voiced clusters: If the glottis is wide open at the release of C1 then it would be virtually impossible for voicing to be initiated in the interval between release of C1 and formation of the dorsal constriction for C2. But then it is unlikely that, regardless of the state of the glottis, the speaker will be able to re-initiate voicing until after the end of the dorsal constriction, given the well-known fact that posterior constrictions are unfavorable for voicing, and the equally well-known hysteresis effect that the aerodynamic requirements for initiation of voicing are more stringent than those for maintaining ongoing voicing (e.g. Hirose & Niimi, 1987). In short, given the typical supraglottal coordination patterns for rhotic clusters there is an extensive period during which aerodynamic conditions are extremely unfavorable to voicing if C1 is voiceless. Speakers may then learn that adducting the glottis before the end of the dorsal constriction does not lead to reliable voicing anyway and thus plan the laryngeal movement to somewhat enhance the voicelessness and frication noise that is present in any case (on the general concept that we have termed 'go with the flow', i.e. planning speech movements to exploit the physical forces arising during articulation see further discussion in e.g. Hoole et al., 2012). The basic assumption is, then, that the specific laryngeal adjustments for plosive+/r/ help to enhance the salience of the distinctions among the rather wide range of complex syllable onsets in German. The most direct way to turn this assumption into a hypothesis for future work would be to make further use of articulatory synthesis. For example, /pr/ could be synthesized both with the empirically observed laryngeal pattern, as well as with that found for /pl/. The expectation would be that the empirically observed pattern results in acoustic output for /pr/ that listeners can distinguish more easily from e.g. /br/, /pl/, and /tr/.[5]

Much as we favor an approach to sonorant devoicing that sees it, at root, as a passive coarticulatory effect (following Browman and Goldstein (1986) and Hoole and Bombien (2010)), if the account in the previous paragraph is correct then more active enhancement processes can sometimes also be involved.

It is interesting to note that on the basis of the present data the more active account applies to the consonants (rhotics) that would be located lower on a scale of consonantal strength in theories of syllable structure such as Vennemann (2012).

In conclusion, the explanatory framework proposed here is still basically consistent with the contention of Browman and Goldstein (1986) that words begin with at most one glottal gesture but tries to make more explicit how and why the coordination patterns are affected by contextual conditions in the onset as a whole.

Acknowledgments

This work was supported by the German Research Council (Grant HO 3271/3-1 to Philip Hoole). We express our sincere gratitude to Bodo Winter for valuable advice with the statistic analyses in this paper. The second author thanks Louis Goldstein and Khalil Iskarous for kindly providing a working environment in the Linguistics Department at USC, Los Angeles. To René Bombien for being the physician conducting the endoscopy: Mercy buckets! Thanks to Elizabeth Heller for help with segmentation and kinematic analysis. We much appreciated the constructive comments of Taehong Cho and two anonymous reviewers on an earlier version of the paper.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.
Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.
Bombien, L., Mooshammer, C., Hoole, P., & Kühnert, B. (2010). Prosodic and segmental effects on EPG contact patterns of word-initial German clusters. *Journal of Phonetics*, *38*, 388–403.
Bombien, L., Mooshammer, C., & Hoole, P. (2013). Articulatory coordination in word-initial clusters of German. *Journal of Phonetics*, *41*, 546–561.
Browman, C., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, *3*, 219–252.
Byrd, D., & Choi, S. (2010). At the juncture of prosody, phonology, and phonetics — The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. In: C. Fougeron, B. Kühnert, M. d'Imperio, & N. Vallé (Eds.), *Papers in laboratory phonology 10: Variation, detail and representation* (pp. 31–60). Berlin/NewYork: Mouton de Gruyter.
Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*, 207–229.
Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, *37*, 466–485.
Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, *33*, 121–157.
Cooper, A. M. (1991). Laryngeal and oral gestures in English /p, t, k/. In *Proceedings of the XIIth international congress of phonetic sciences* (pp. 50–53). Aix-en-Provence.
Docherty, G. J. (1992). *The timing of voicing in British english obstruents*. Berlin; New York: Walter de Gruyter.
Fuchs, S. (2005). Articulatory correlates of the voicing contrast in alveolar obstruent production in German. *ZAS Papers in Linguistics*, *41*, 1–238.
Goldstein, L. (1990). On articulatory binding: Comments on Kingston's paper. In: J. Kingston, & M. E. Beckman (Eds.), *Papers in laboratory phonology 1: Between the grammar and physics of speech* (pp. 445–450). Cambridge: Cambridge University Press.
Haggard, M. (1973). Abbreviation of consonants in English pre- and post-vocalic clusters. *Journal of Phonetics*, *1*, 9–24.
Hanson, H. M., & Stevens, K. N. (2002). A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn. *Journal of the Acoustical Society of America*, *112*, 1158–1182.
Hawkins, S. (1979). Temporal coordination in the speech of children: Further data. *Journal of Phonetics*, *7*, 235–267.
Hirose, H., & Niimi, S. (1987). The relationship between glottal opening and the transglottal pressure differences during consonant production. In: T. Baer, C. Sasaki, & K. Harris (Eds.), *Laryngeal function in phonation and respiration* (pp. 381–390). Boston: College Hill.
Hoole, P. (1999). Techniques for investigating laryngeal articulation section A: Investigation of the devoicing gesture. In: W. J. Hardcastle, & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques* (pp. 194–300). Cambridge: Cambridge University Press.
Hoole, P. (2006). *Experimental studies of laryngeal articulation – Part II: Laryngeal–oral coordination in consonant sequences (unpublished habilitation thesis)*. Ludwig-Maximilians-Universität Munich (〈http://www.phonetik.uni-muenchen.de/∼hoole/pdf/habilpgg_chap_all.pdf〉 Accessed 07.03.13).
Hoole, P., & Bombien, L. (2010). Velar and glottal activity in Icelandic. In: S. Fuchs, P. Hoole, M. Zygis, & C. Mooshammer (Eds.), *Between the regular and the particular in speech and language* (pp. 171–204). Peter Lang: Frankfurt am Main.
Hoole, P., Fuchs, S., & Dahlmeier, K. (2003). Interarticulator timing in initial consonant clusters. In S. Palethorpe, & M. Tabain (Eds.), *Proceedings of the 6th international seminar on speech production* (pp. 101–106). Sydney.
Hoole, P., Pompino-Marschall, B., & Dames, M. (1984). Glottal timing in German voiceless occlusives. In M. Van der Broecke, & A. Cohen (Eds.), *Proceedings of the 10th international congress of phonetic sciences* (pp. 309–403). Utrecht, The Netherlands.
Hoole, P., Pouplier, M., Beňuš, Š., & Bombien, L. (2013). Articulatory coordination in obstruent-sonorant clusters and syllabic consonants: Data and modelling. In L. Spreafico, & A. Vietti (Eds.), *Proceedings of ratics3* (pp. 79–94). Bolzano: Bolzano University Press.
Hoole, P., Pouplier, M., & Kühnert, B. (2012). System related variation. In: A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford Handbook of Laboratory Phonology* (pp. 115–130). Oxford University Press, Oxford.
Hutters, B. (1984). Vocal fold adjustments in Danish voiceless obstruent production. *Annual Report of the Institute of Phonetics, University of Copenhagen*, *18*, 293–385.
Iverson, G., & Salmons, J. (1995). Aspiration and laryngeal representation in Germanic. *Phonology*, *12*, 369–396.

---

[5] One initially puzzling result that we left out of the summary of plosive+/r/ clusters in Section 4.9.4 was that the measurement of the interval from completion of glottal adduction to voicing onset did not indicate aerodynamically unfavorable conditions at the end of glottal adduction in the plosive+/r/ clusters, apparently running counter to our assumption of dorsal constrictions being unfavorable for phonation. The explanation may lie precisely in the fact that the speakers have enhanced the duration and the magnitude of the gesture such that it is not completed until the dorsal constriction has been relaxed to such an extent that it no longer impedes phonation. This is exactly what is indicated by the schematic overview of temporal structure in Fig. 13: The end of the glottal adduction movement comes much later relative to the end of C2 for plosive+/r/ compared to plosive+/l/.

Jessen, M. (1997). Word boundary marking at the glottal level in the production of German obstruents. *ZAS Papers in Linguistics*, *11*, 147–166.

Jessen, M. (1999). Redundant aspiration in Geman is primarily controlled by closure duration. In *Proceedings of the XIVth international congress of phonetic sciences* (pp. 993–996). San Francisco.

Keating, P., Cho, T., Fougeron, C., & Hsu, C. (2003). *Domain-initial articulatory strengthening in four languages*. *Laboratory phonology*, Vol. 6. Cambridge: Cambridge University Press143–161.

Kehrein, W., & Golston, C. (2004). A prosodic theory of contrast. *Phonology*, *21*, 1–33.

Kelso, J. A. S., & Tuller, B. (1985). Intrinsic time in speech production: Theory, methodology and preliminary observations. *Haskins Laboratories Status Report on Speech Research*, *81*, 23–39.

Klatt, D. H. (1973). Aspiration and voice onset time in word-initial consonant clusters in English. *Journal of the Acoustical Society of America*, *54*, 319.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, *18*, 686–706.

Löfqvist, A., & McGarr, N. (1987). Laryngeal dynamics in voiceless consonant production. In: T. Baer, C. Sasaki, & K.S Harris (Eds.), *Laryngeal function in phonation and respiration* (pp. 391–402). Boston: College Hill.

Löfqvist, A., & Yoshioka, H. (1980). Laryngeal activity in Swedish obstruent clusters. *Journal of the Acoustical Society of America*, *63*, 792–801.

Ridouane, R., Fuchs, S., & Hoole, P. (2006). Laryngeal adjustments in the production of voiceless obstruent clusters in Berber. In: M. Tabain, & J. Harrington (Eds.), *Speech production: Models, phonetic processes and techniques* (pp. 275–297). New York: Psychology Press.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, *1*, 333–382.

Shipp, T. (1982). Aspects of voice production and motor control. In: S. Grillner, B. Lindblom, J. Lubker, & A. Persson (Eds.), *Speech motor control* (pp. 105–112). Oxford: Pergamon.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: The MIT Press.

Tsuchida, A., & Cohn, A. C. (2000). Sonorant devoicing and the phonetic realization of [spread glottis] in English. *Working papers of the Cornell phonetics laboratory* (pp. 167–181), Vol. 13.

Vennemann, T. (2012). Structural complexity of consonant clusters: A phonologist's view. In: P. Hoole, L. Bombien, M. Pouplier, C. Mooshammer, & B. Kühnert (Eds.), *Consonant clusters and structural complexity* (pp. 11–31). Berlin: de Gruyter.

Weismer, G. (1980). Control of the voicing distinction for intervocalic stops and fricatives: Some data and theoretical considerations. *Journal of Phonetics*, *8*, 427–438.

Yoshioka, H., Löfqvist, A., & Hirose, H. (1981). Laryngeal adjustments in the production of consonant clusters and geminates in American English. *Journal of the Acoustical Society of America*, *70*, 1615–1623.