JSLHR

Research Article

A Cross-Language Study of Laryngeal-Oral Coordination Across Varying Prosodic and Syllable-Structure Conditions

Philip Hoole[a] and Lasse Bombien[a]

Purpose: The purpose of this study is to use prosodic and syllable-structure variation to probe the underlying representation of laryngeal kinematics in languages traditionally considered to differ in voicing typology (German vs. Dutch and French).

Method: Transillumination and videofiberendoscopic filming were used to investigate the devoicing gesture in German, Dutch, and French for material that compared, first, a strong versus weak prosodic condition and, second, singletons versus clusters (stop + /r/ and /l/).

Results: The results showed strengthening of the devoicing gesture in the strong prosodic condition and in the segmental context stop + /r/ for German and French but not for Dutch. In terms of timing (duration of oral occlusion, voice onset time, timing of peak glottal opening relative to stop release), French was intermediate between German and Dutch.

Conclusions: (a) The representation of French voiceless plosives requires an active specification for glottal spreading just as in German. (b) Static features are not well suited to capturing cross-language differences in voicing typology and changes in voicing specification over time.

This study investigates laryngeal-oral coordination for voiceless stops in languages that are traditionally considered to differ in terms of their voicing typology, namely, German and French, and thus extends to a cross-language context some of the issues that we have recently addressed for German in Hoole and Bombien (2014). In traditional terms, German would be regarded as a representative of languages (also including English) that contrast clearly aspirated voiceless stops with substantially devoiced voiced stops, whereas Dutch and French contrast unaspirated voiceless stops with full voicing in the voiced stops. Here, we will be concerned only with the phonologically voiceless stops. In terms of laryngeal coordination, the above pattern translates in German to the location of the peak glottal opening of the devoicing gesture roughly at the time of release of the oral occlusion (Browman & Goldstein, 1986; Hoole & Bombien, 2014; Jessen, 1999), with glottal adduction for the following vowel being completed substantially after the oral release, whereas in French and Dutch, peak glottal opening occurs roughly in the center of the oral occlusion phase, with adduction being completed roughly at the same time as the oral release (Benguerel, Hirose, Sawashima, & Ushijima, 1978; Yoshioka, Löfqvist, & Collier, 1982; also Löfqvist, 1980).

The basic rationale for the present investigation was that prosodic variation, on one hand, and segmental context, on the other hand, can be used as a two-pronged probe to make clearer the nature of the articulatory representations underlying the voicing distinction in general and also underlying traditional phonetic terms such as *aspirated* and *unaspirated,* in particular.

Turning first to prosodic variation, in English, the voice onset time (VOT) of voiceless plosives is often found to be longer in prosodically strong locations (see, e.g., Cho & McQueen, 2005, for discussion), suggesting that the laryngeal abduction gesture is also longer and/or larger. More intriguingly, Cho and McQueen (2005) observed for Dutch that prosodic strengthening can lead to shortening of VOT. This was interpreted (following Keating's [1984] distinction between phonological and phonetic features) as indicating that the phonologically voiceless plosives are implemented phonetically as {+ spread glottis} in English and {– spread glottis} in Dutch (these phonetic features then being reinforced as part of prosodic strengthening). However, neither for English nor for Dutch is anything known about what changes in laryngeal movement actually

[a]Institute of Phonetics and Speech Processing, Munich University, Germany

Correspondence to Philip Hoole: hoole@phonetik.uni-muenchen.de

Editor: Julie Liss
Associate Editor: Leonardo Lancia

Received January 30, 2015
Revision received March 10, 2015
Accepted August 7, 2016
https://doi.org/10.1044/2016_JSLHR-S-15-0034

Disclosure: The authors have declared that no competing interests existed at the time of publication.

take place under prosodic strengthening (but see Hoole & Bombien, 2014, for discussion of effects of word stress). This work thus continues in the tradition of work aiming to understand how prosodic structure is realized in articulation (e.g., Keating, Cho, Fougeron, & Hsu, 2003).

Further phonological accounts of voicing across languages would also appear to predict differential effects of prosodic strengthening that depend on laryngeal specification. For example, Iverson and Salmons (2007) analyzed voiceless plosives as laryngeally specified for [spread glottis] in German and English but as laryngeally unspecified for, for example, French and Dutch (vice versa for the voiced plosives: laryngeally unspecified in German and English, but actively specified for [voice] in Dutch and French). If this reflects a difference in representation between typologically different languages, then prosodic strengthening would not be expected to lead to reinforcement of the laryngeal devoicing gesture in French and Dutch. This line of argument is given very explicit expression in Beckman, Helgason, McMurray, and Ringen (2011). They followed a similar privative phonological framework in which, in typical laryngeal contrasts, only one member of a pair is actively specified. They discuss evidence that in speech rate changes (which they see as closely related to other kinds of prosodic strengthening), it is precisely the actively specified member of a contrast where the laryngeal feature is strengthened/lengthened (i.e., aspiration in [spread glottis] languages and prevoicing in [voice] languages). Their investigation specifically deals with Swedish, which is interesting following this line of reasoning because it indicates that in some languages, instead of contrasting an actively specified stop with an unspecified one (as in German, English, etc.), two active specifications may be required: Swedish speakers lengthened both voicing lead in voiced stops and voicing lag in voiceless stops at slower tempo. This kind of observation will turn out to be very relevant for the discussion of our own results.

Following this privative account, the basic expectation here is that any strengthening of the devoicing gesture in prosodically strong conditions should be confined to German. Following the rather different account of Cho and McQueen (2005), if French and Dutch do show prosody-related changes, they should be qualitatively different from those in German and not just quantitatively weaker; that is, rather than just showing less lengthening of VOT in strong conditions than German, they should show some combination of laryngeal and/or oral durational changes that result in shorter VOT in the strong condition.

Turning now to the second prong in the overall approach of probing the behavior of the devoicing gesture in different contextual conditions, we systematically varied the syllable onsets of the target items, specifically by comparing singletons (e.g., /p/) with clusters (e.g., /pr/, /pl/). On one hand, some timing changes can be expected to be qualitatively similar across languages (even though direct laryngeal information for languages such as French and Dutch is lacking, to our knowledge); partly because of a tendency toward shortening of the occlusion of the initial plosive, peak glottal opening will be timed later with respect to the release of the plosive. This in turn will tend to lead to an increase in VOT in the clusters (see especially Jessen, 1999, for German; also Tsuchida & Cohn, 2000, for English). On the other hand, and more crucially, some evidence has been found that the overall duration of voicelessness is longer in the clusters (e.g., Docherty, 1992, for English; Hoole & Bombien, 2014, for German), even though the second element of these clusters would not normally be regarded as underlyingly voiceless. To the extent that this is an active mechanism involving a strengthening of the glottal devoicing gesture itself, the question arises as to whether the amount of gestural reorganization in the clusters will be sensitive to the underlying voicing typology of the language involved. The basic prediction links up with the prosodic condition outlined above: If it is only in German that plosives are assumed to have an active specification for glottal spreading, then it is only in this language that the gesture will be available for strengthening in specific contextual conditions. Because Hoole and Bombien (2014) found most evidence for strengthening in the context plosive + /r/, the more specific prediction is that the glottal gesture will not be strengthened in this context in Dutch or French.

A general point to be made about the representational frameworks considered above is that they are essentially nontemporal. Given that voicing control is probably the quintessential case in which coordination relations have linguistic relevance, a particular focus in the discussion will be on whether the results are more easily captured in a gesturally oriented framework (Browman & Goldstein, 1986). One potential advantage of a gestural approach is that it may be better suited to capturing a behavioral continuum, including possible shifts in a language's position on such a continuum over time. Cho and Ladefoged's (1999) wide-ranging review of VOT values across languages for voiceless stops shows that, even if there is a fairly strong tendency for languages to be localized in three main VOT ranges, there is still a nonnegligible proportion of intermediate values. Simply dismissing this timing variation as details of phonetic implementation serves only to perpetuate a hard dichotomy between continuous phonetics and discrete phonology. There are, in fact, already some indications that a framework allowing intermediate representations could be crucial for the languages with which we are concerned here. Specifically, there are indications that French is not as firmly at the unaspirated end of the continuum as Dutch is: Torreira and Ernestus (2011) found evidence for a more vigorous devoicing gesture in French than in Spanish (the latter being another typical representative of a [voice] rather than a [spread glottis] language); Kirby and Ladd (2015) showed that cases of quite strong aspiration are not difficult to find in French. This leads to a final very basic motivation for the present work: Direct measurements of laryngeal behavior in French (and also Dutch) are very limited and, such as they are, already quite old (the investigation of Benguerel et al., 1978, is now almost 40 years old). It is not entirely straightforward to derive precise timing information from Benguerel et al. (1978) because of the somewhat limited temporal resolution of the fiber-optic measurements they were using, but the basic pattern seems consistent with the traditional idea of an unaspirated plosive (i.e., glottal adduction completed very shortly after release of the oral occlusion and peak glottal opening probably timed roughly in the middle of the oral occlusion phase). A key point of interest in our results will be whether there is indeed evidence that glottal timing in French may have shifted toward the aspirated end of the continuum. Such shifts in interarticulatory coordination may well be captured more parsimoniously by a gestural approach, because this obviates the necessity to categorically assign the language to the [voice] or [spread glottis] pattern.

Experimental Procedures

Speech Material

The target words all had a voiceless consonant (plosive or fricative) in initial position. Two prosodic conditions were compared: a condition in which the target word was in the focused position versus an unaccented condition. The structure of the syllable onset was systematically varied by comparing singleton plosives and fricatives with all combinations with /l/ and /r/ available in the language. (The material analyzed here was embedded in a larger corpus containing three-element onset clusters, fricative-plosive clusters, and voiced control items, as well as a third prosodic condition that was less comparable across languages.)

The material is summarized for each language below, using the following schematization of prosodic context: **xxx**: target; *xxx*: focus; xxx: contrast (note that bold, italics, and underlining are used here for illustrative purposes; these typographic devices were not used in the orthographic version presented to the subjects).

1. **German.** Plosive: p, t, k, pl, kl, pr, tr, kr   Fricative: f, ʃ, fl, ʃl, fr, ʃr
   Focused: Bis sie ***piep*** sieht, nicht Tisch. [“Until she sees ***peep***, not table”]
   Deaccented: Bis sie **piep** *sieht,* nicht hört. [“Until she *sees* **peep** instead of *hearing* it”]
2. **Dutch.** Plosive: p, t, k, pl, kl, pr, tr, kr   Fricative: f, s, fl, fr
   Focused: Als ’t-ie ***piep*** ziet, niet last. [“If he sees ***peep***, not burden”]
   Deaccented: Als ’t-ie **piep** *leest,* niet weet. [“If he *reads* **peep**, instead of knowing it”]

Note: In the orthographic form of the phrase “Als ’t-ie,” the apostrophe-t was used to indicate to the subjects that “Als ie” should be uttered with a linking /t/, a very common connected speech process in Dutch that does not, however, have a standard orthographic representation. Audio recordings of the intended realization were prepared in advance by a Dutch native speaker and played back as examples to the subject to make it clear that apostrophe-t did not have its more usual orthographic function of coding forms of the definite article or object pronoun (which would not be grammatically possible in this carrier phrase).

3. **French.** Plosive: p, t, k, pl, kl, pr, tr, kr   Fricative: f, ʃ, s, fl, sl, fr
   Focused: C’était ***‘pipe’*** qu’il citait (pas ‘quiche’) [“It was ‘pipe’ that he quoted (not ‘quiche’)”]

Note: The material in parentheses provided the context but was not spoken.

Unaccented: Voici des **pipes** très étroites [“Here are some very narrow pipes”]

Note: The target “pipes” is in nonfinal position in the noun phrase, and together with the use of “très” (“very”), was an attempt to make it more likely that the adjectival part of the noun phrase is emphasized.

Further Notes on the Material

Note that the plosive-initial material is identical for all three languages (all three lack /tl/). In all target words, the vowel following the target sounds was a high front vowel to ensure that light transmission from the endoscope through the glottis was not obstructed by tongue retraction. For the same reason, high front vowels were used as far as possible throughout the carrier phrase; thus, for example, in Dutch, the more colloquial variant “ie” was preferred to “hij” (= Engl. “he”). A point that will be relevant in the discussion is that the rhotic is dorsal in our German and French speakers (ranging from voiced approximant to voiceless fricative, depending on context) and apical in the Dutch speakers (for convenience, the phonemic symbol /r/ is used throughout).

In the results below, the focused material will be referred to as the strong prosodic condition and the deaccented/unaccented as the neutral condition.

Speakers

Five speakers of German and four each of French and Dutch were analyzed. In most cases, five randomized repetitions of each target item were available for analysis (for one French speaker, only two repetitions were completed). Note that the randomization was carried out over the complete corpus, so the different prosodic conditions were not recorded in separate blocks. In each block of repetitions, each item in the corpus occurred once.

Details of the speakers are as follows:

- German: Two women, three men, aged 26–35 years. All spoke standard German, with only minor regional coloring.
- Dutch: Three women, one man; three speakers were in their 20s or 30s, and one female speaker was in her 60s. All spoke a form of standard Dutch typical of the central western conurbation (“Randstad”).
- French: Three women, one man; two speakers were in their 20s and two in their 30s. All spoke a variety of standard French characteristic of the northern half of France, with no major regional coloring.

All recordings were carried out in the phonetics lab at the University of Munich. The German speakers resided in Munich at the time of the recordings. The Dutch speakers all traveled to Munich specially for the recordings. One French speaker came to Munich for the recordings, two had been resident in Munich for a few months as part of a student exchange, and one was a university employee who had been resident in Germany for 2 years (this was the speaker for whom the least data were acquired). The latter three French speakers were all closely integrated in French-speaking communities.

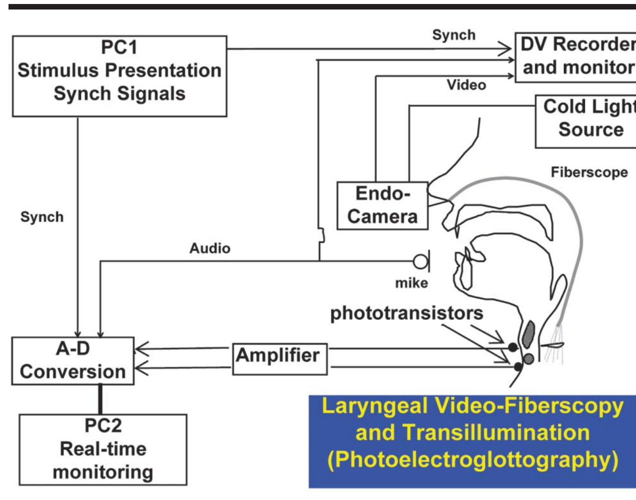
During the recordings, both authors monitored the utterances of the speakers for pronunciation errors (particularly prosodic errors such as focused realization of the target word in the deaccented condition). If possible, such items were repeated. A few errors were overlooked during the recording sessions but were eliminated when the material was checked again for errors prior to starting the actual analysis. Between them, the authors have native or very good knowledge of German and French. For Dutch, the corpus was prepared with the advice of a native Dutch linguist, who, as mentioned above, also provided us with a complete audio recording of the corpus. This allowed the authors to familiarize themselves with a typical realization prior to the experiment and made it possible to illustrate to the subjects the prosodic conditions involved. We are thus confident that the realizations retained for analysis correspond adequately to the intended prosodic conditions.

Methods

The abductory and adductory movements of the glottis associated with the voiceless consonants were monitored by means of transillumination. Briefly, the amount of light passing through the glottis from a light source located in the pharynx is modulated by the laryngeal movements and registered by phototransistors attached externally to the neck below the level of the glottis. The transillumination signals together with the audio signal (Sennheiser MKH40 microphone) and synchronization signal were recorded on a Sony-Ex multichannel instrumentation recorder, using a sample rate of 32,768 Hz for the audio signal and 8,192 Hz for the other signals. An overview of the setup including synchronized videolaryngofiberscopy is given in Figure 1. All procedures were approved by the human subjects committee of the medical faculty of Munich University. See Hoole and Bombien (2014) and Hoole (1999) for further technical details and background. For the measurements of laryngeal kinematics, the transillumination signal was smoothed using a Kaiser-design finite impulse response filter with a cutoff frequency of 25 Hz, a transition band of 25 Hz, and minimum damping in the stop band of 70 dB. These are typical smoothing parameters for kinematic studies of articulation; smoothing is particularly relevant for the transillumination signal because the signal may contain phonatory modulation at the fundamental frequency and above, which is not of interest when analyzing devoicing kinematics.

For the present article, we will concentrate on the following measures (see Figure 2): (a) duration of the oral occlusion of C1 and (b) VOT (from release of C1 to onset

Figure 1. Schematic overview of the experimental setup for combined audio, transillumination, and videofiberscopic recording.

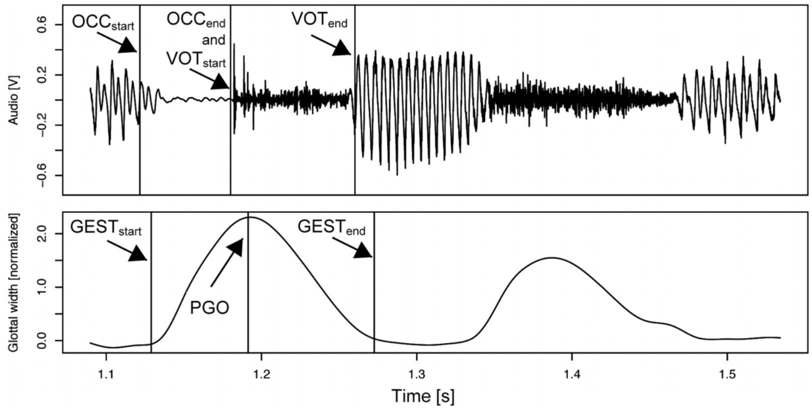


of voicing), both on the basis of the acoustic signal, using waveform and spectrogram (on the basis of the time points in Figure 2, these two measures correspond to calculating OCCend-OCCstart and VOTend-VOTstart, respectively), (c) relative timing of peak glottal opening, (d) magnitude of peak glottal opening, and (e) glottal gesture duration, determined from the transillumination signal.

Relative timing of peak glottal opening was calculated as the time from the onset of oral occlusion to the time of peak glottal opening, divided by the duration of oral occlusion of C1 (i.e., the position of time point PGO relative to the interval delimited by OCCstart and OCCend in Figure 2). For an aspirated plosive (as shown in Figure 2), where peak glottal opening is roughly synchronous with the release of the oral occlusion, the relative timing measure gives a value of about 1. For an unaspirated plosive (and also for fricatives), where peak glottal opening occurs at about the midpoint of the oral occlusion, a value of about 0.5 would be expected. On the basis of previous work (e.g., Hoole & Bombien, 2014; Jessen, 1999; Tsuchida & Cohn, 2000), later relative timing of peak glottal opening is expected for clusters compared with singletons (e.g., values > 1 if peak glottal opening occurs after the end of C1).

Regarding the magnitude of peak glottal opening, because there is no simple way to calibrate the transillumination signal, and because the signal level can vary quite substantially over the course of the experiment, a normalization factor was calculated separately for each block of repetitions. Specifically, this was based on the average glottal opening over all items with fricative onsets, calculated separately for each speaker and block of repetitions. The motivation for this was that we are particularly interested in laryngeal differences for the plosives across languages, whereas there is no particular reason to expect major language-specific differences in glottal opening magnitude for the fricatives. The aerodynamic constraints on voiceless fricative production should be very similar across languages:

Figure 2. Illustration of measurements made on the audio signal (top panel) and low-pass filtered transillumination signal (bottom panel) for a token of German "Kies" /ki·s/ in strong prosodic condition. The target sound is the aspirated initial plosive. OCC = oral occlusion of C1 (i.e., /k/); VOT = voice onset time; GEST = glottal opening–closing gesture; PGO = time point of peak glottal opening.



For example, on the basis of Stevens (1998, e.g., see Figure 2.37), generation of frication noise should be maximized when the area at the glottis is roughly twice that of the supraglottal constriction.[1] Nevertheless, cross-language and even gender-specific differences in constriction formation for fricatives could lead to differences in the supraglottal constriction area that limit in turn the generality of the assumption of comparable glottal opening (see, e.g., Fuchs & Toda, 2010; we are grateful to the associate editor for drawing this to our attention). As noted above, the gender distribution of our subjects is not completely uniform across languages (men in the majority for German, women for French and Dutch).

In any case, in the absence of an absolute measure, our procedure allows us to express glottal opening for plosives as a proportion of glottal opening for fricatives. This is a functionally relevant measure because (even if it is not absolutely precise for cross-language comparisons) it helps us to assess whether, for example, the rhotic in the plosive-rhotic clusters is actively being planned as a fricative (as well as increasing the precision of other within-subject comparisons, such as the influence of prosodic strengthening).

For interpretation of the results, the above two basic glottal parameters on the timing and magnitude of peak glottal opening often require supplemental information on the total duration of the glottal gesture (i.e., the time from the start of glottal abduction to the completion of glottal adduction). This is defined here as the interval from GESTstart to GESTend in Figure 2, determined using a 20% threshold of maximum velocity in the abduction and adduction phase, respectively.

[1]Representative frames of laryngeal images taken from the videoendoscopic films are included in the online supplemental materials: for one speaker of each language, one token of a fricative (in the neutral prosodic condition) and of a plosive in the two prosodic conditions (see Supplemental Material S11).

Analysis and Results

For the five parameters outlined above, a general overview of the results will be given in a set of five figures (see Figures 3, 4, 5, 6, and 7), each organized in the following way.

Each figure consists of a left and right subblock, dividing the data with respect to prosodic condition (neutral on the left and strong on the right). Within each subblock, the arrangement on the x-axis is given by the syllable-structure condition, with P standing for singleton plosive, PL for plosive + /l/, and PR for plosive + /r/. As will be seen from the details of the material given above, the corpus incorporates

Figure 3. Results for occlusion duration of C1, broken down by language (color and symbols), prosodic condition (left and right subblocks for neutral and strong condition, respectively), and structure of syllable onset (position on x-axis, labeled *P* for singletons, *PL* for plosive-lateral clusters, and *PR* for stop-rhotic clusters). Error bars represent 2.5 times the standard error of mean. See text for further details. Color online.

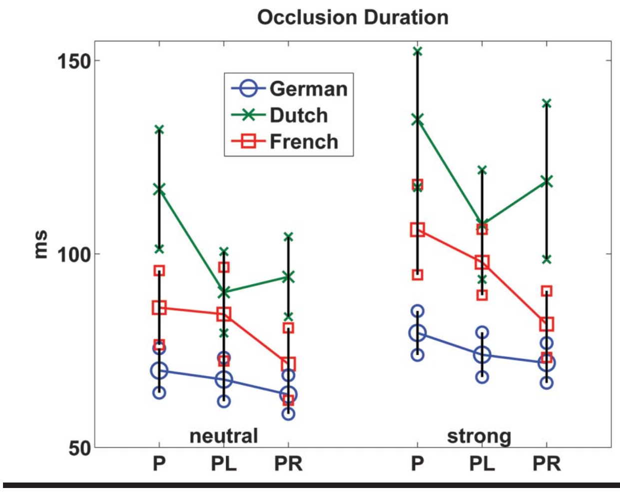
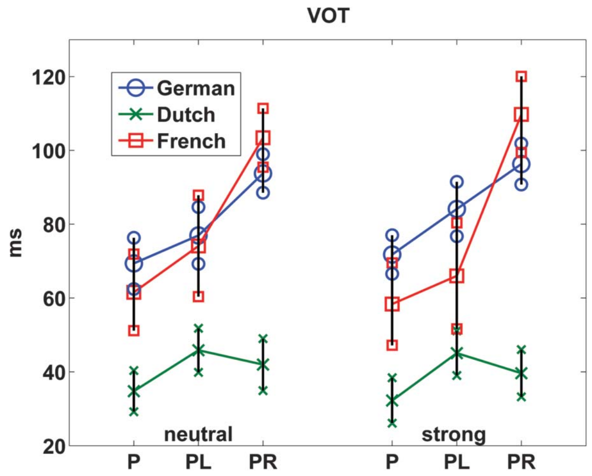


Figure 4. Results for voice onset time. Same arrangement as for Figure 3.



a (reasonably) systematic variation of place of articulation of C1. However, because this is not of particular interest in the present article (see Hoole & Bombien, 2014, for detailed discussion of the German data), all of the results shown in the figures have been averaged over place of articulation (three places for P and PR, two places for PL; note that *P* in the axis labels thus stands for *plosive* and not for bilabial *p*). More specifically, this applies not only to the mean values but also to the error bars. These correspond to 2.5 times the standard error of mean, whereby the standard error of mean was first calculated for each combination of language, prosodic condition, syllable structure, and place of articulation and then averaged over place of articulation. The error bars should thus give a reasonable approximation to the 99% confidence interval of the corresponding mean value that reflects speaker-to-speaker and token-to-token variation (but is not unduly inflated by systematic effects of place of articulation, which are undoubtedly often present) and thus give a good indication of the robustness of any between-language differences, in particular.[2]

In addition, Figure 8 assembles a schematic overview of the temporal structure of each utterance type with respect to all the acoustic and glottal measures for all three languages.

For statistical testing, the basic procedure involved *N*-way analysis of variance (ANOVA) as implemented in the MATLAB function anovan, with speaker as a random factor nested in language and with prosodic condition, syllable structure, and place of articulation as within-subject factors. To avoid inflation of the degrees of freedom, the results for each cell were reduced to their means before being used in anovan (i.e., typically averaged over five repetitions; as mentioned above, for one French speaker, only two repetitions were available). After this averaging process, the data set used as input to the ANOVA consisted of 13 speakers (over the three languages) $\times$ eight combinations of syllable structure and place of articulation $\times$ two prosodic conditions = 208 values per measured parameter. It turned out that a subdivision of the data to compare French separately with each of the other two languages was often convenient; for these reduced data sets, the data amounted to 144 and 128 values, respectively, for the comparison of French with German and French with Dutch.

Figure 5. Results for relative position within C1 occlusion (labeled "re. occ1" on the ordinate) of peak glottal opening. Same arrangement as for Figure 3.

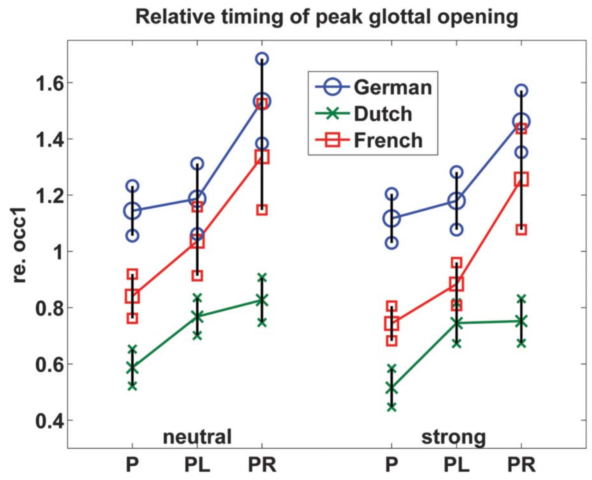
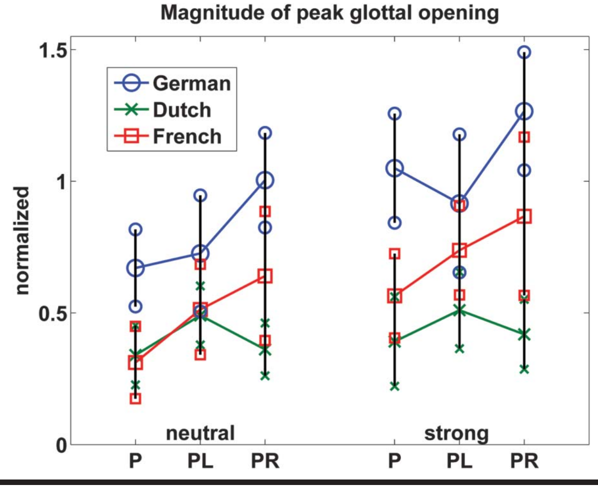
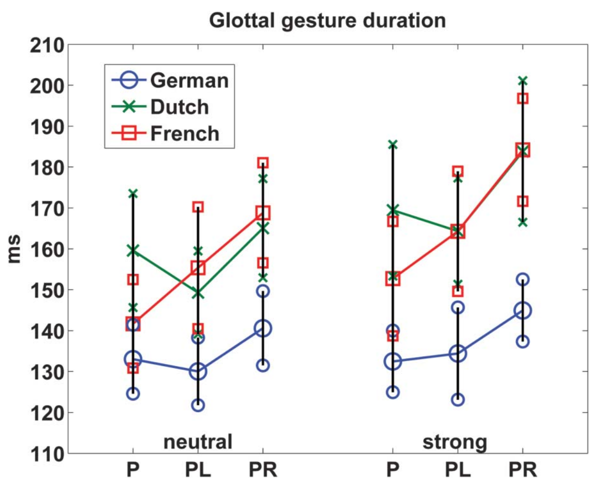


Figure 6. Results for magnitude of peak glottal opening. Same arrangement as for Figure 3.



[2]Because there is relatively little previous transillumination data on any of the languages studied here, we have included an additional set of figures with a complete breakdown of place of articulation in the online supplemental materials: see Supplemental Material S1 and S2 for Occlusion Duration, S3 and S4 for VOT, S5 and S6 for Relative Timing of Peak Glottal Opening, S7 and S8 for Peak Glottal Opening Magnitude, and S9 and S10 for Glottal Gesture Duration (for each parameter, separate figures for normal and strong prosodic context).

Figure 7. Results for glottal gesture duration. Same arrangement as for Figure 3.



Because of the unbalanced place of articulation condition (absence of the /tl/ combination), further details of the most convenient arrangement of the data for input to anovan also varied somewhat depending on the specific comparisons of most interest. This is outlined as required below. Particularly because of the subdivision of various analyses alluded to here, the issue of correction of p values for multiple comparisons arises. On the basis of considerations in Abdi (2007), it was estimated that to achieve an experiment-wide alpha error rate of .05, alpha should be corrected to .0127 for individual comparisons (on the basis of the Šidák correction $1 - (1 - \text{alpha})^{1/n}$, with $n = 4$). This indicates the need for caution for uncorrected p values between $p < .05$ and $p < .01$. In the end, we decided to report uncorrected p values, because whenever key results fell in this critical region (which we will refer to as *marginally significant*), additional ad hoc statistical procedures carried out for the purpose of cross-checking generally indicated that the N-way ANOVA was already quite conservative. Again, details of this are given as required below.

Occlusion Duration

We start the presentation of the results with occlusion duration, because this gives a straightforward indication as to whether the attempt to contrast different degrees of prosodic strength on the target word has been successful (given that it is difficult to assume the prosodic conditions are strictly comparable across languages, especially French vs. German and Dutch).

Comparing the left and right blocks of Figure 3 clearly indicates that, as expected from many previous investigations, occlusion durations are longer under prosodic strengthening. Of the three languages, the differences are smallest for German (about 10 ms), but this still represents a statistically significant difference between the prosodic conditions. Figure 3 also shows the effect of language and syllable type on occlusion duration. For language, values increase from German via French to Dutch.

To perform a basic statistical test of the prosody and language effect, we merged syllable structure and place of articulation into a single syllable onset factor (with a total of eight levels), because, as mentioned above, we are not specifically interested here in place of articulation and thus avoid the problem of an unbalanced design. In this three-factor design, both language ($p < .01$) and prosody ($p < .001$) were significant. The Language × Prosody interaction was not significant: Language, $F(2, 10) = 8.43$; Prosody, $F(1, 10) = 31.47$; Language × Prosody, $F(2, 10) = 1.93$.

With regard to syllable structure, there is a not always very large but nonetheless consistent trend for C1 durations to be shorter in the cluster onsets (PL and PR) compared with the singleton onset (P). The latter effect is certainly not unexpected but needs to be borne in mind when we come to the interpretation of the results for laryngeal-oral coordination.

To test this statistically, we ran two separate ANOVAs in which now both syllable structure and place of articulation were factors (in addition to prosody and language), first comparing P with PR (i.e., using the onsets /p, t, k, pr, tr, kr/) and then P with PL (using the items /p, k, pl, kl/).

For the singleton and rhotic cluster comparison, syllable structure was significant at $p < .001$, $F(1, 10) = 67.02$. The interaction of language and syllable structure was marginally significant, $p < .05$, $F(2, 10) = 5.13$, reflecting somewhat different amounts of shortening of C1 in the clusters across languages.

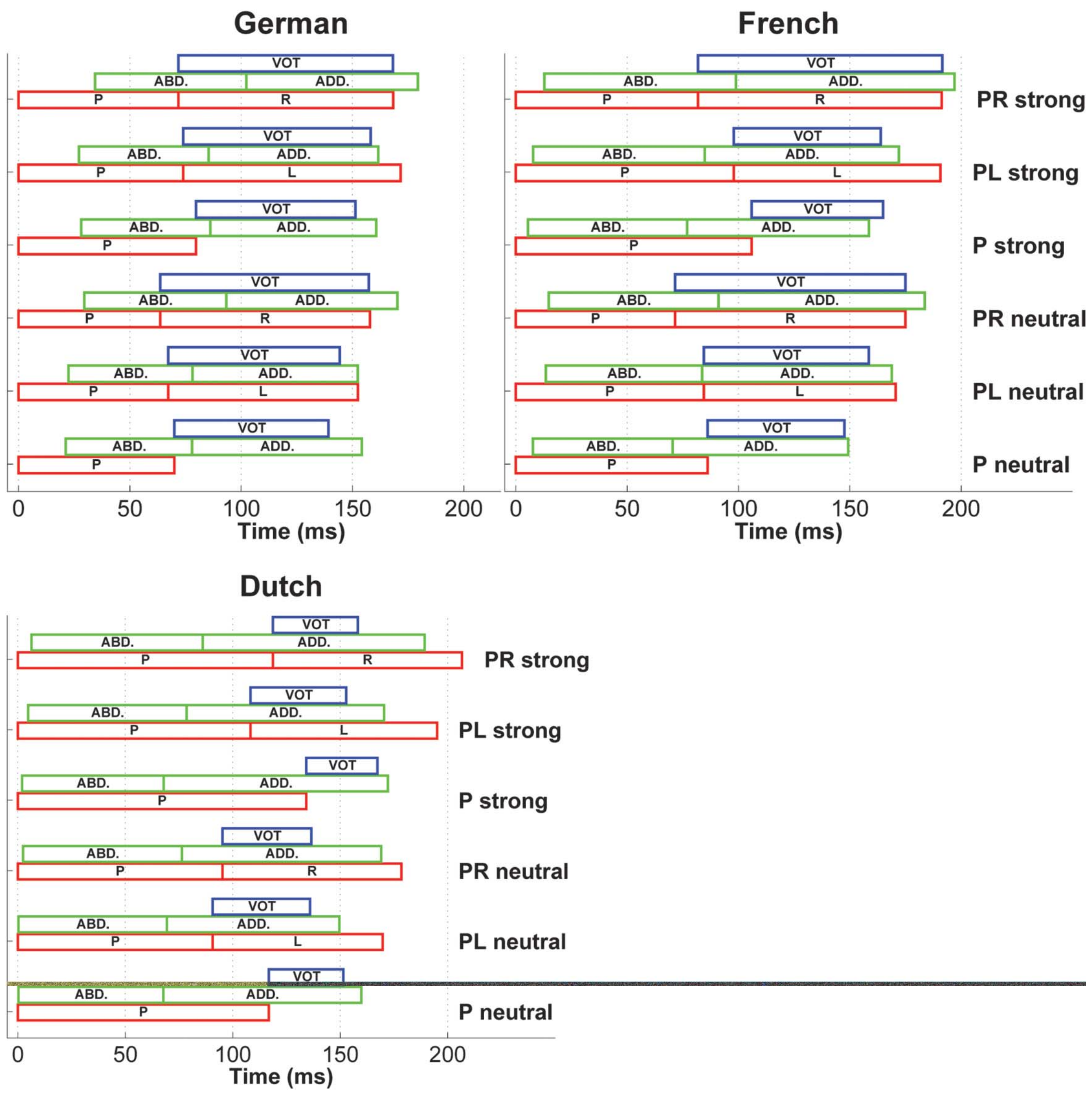
The results for the singleton and lateral cluster comparison were similar: syllable structure was significant at $p < .01$, $F(1, 10) = 20.42$, and the interaction between language and syllable structure was again marginally significant, $p < .05$, $F(2, 10) = 5.24$.

VOT

Turning to the other acoustic parameter, the first point to make about VOT is that in several respects, the results invert the pattern of the occlusion duration results (i.e., they increase from Dutch via French to German and also increase going from the singleton to cluster syllable-types; see Figure 4). Perhaps the most striking result is that French is actually closer to German than Dutch. Note also the particularly high values in the rhotic clusters for German and even more so for French (where French even exceeds German). Comparing the left and right blocks of Figure 4 does not reveal any obvious effect of prosodic condition for any of the languages.

In a three-factor ANOVA with syllable structure and place of articulation merged as above, there were trivially huge main effects of language and syllable onset (looked at in detail below) but no main effect of prosody and in particular no interaction of prosody and language, which would have been the specific prediction if German had increased and Dutch had decreased with prosodic strengthening, as might have been expected from the literature.

Figure 8. Schematic illustration of timing patterns for main utterance categories (data are averaged over speakers, repetitions, and place of articulation of C1). Same labeling of utterance types as in Figures 3–7. In each group of three bars, the vertical arrangement is as follows: The lowest bar(s) show oral timing of C1 and C2 on the basis of the acoustic signal (the left and right edges of the bar labeled *P* correspond in Figure 2 to OCCstart and OCCend, respectively); the middle bars show glottal timing (ABD = abduction phase, ADD = adduction phase; peak glottal opening occurs at the junction between these two phases; in Figure 2, the left edge of ABD corresponds to GESTstart and the right edge of ADD to GESTend). The top bar shows voice onset time (from release of C1 to voice onset). All patterns are aligned with the oral onset of C1 at zero. Separate panels for each language.



To shed more light on the status of French (once again the language in the middle), we compared French with German and Dutch in separate three-factor ANOVAs.

Interestingly, in the comparison of German and French, the main effect of language was indeed not significant. There was, however, a significant interaction between language and syllable onset ($p < .001$), probably because of the particularly strong increase in French for the rhotic clusters. If we treat the latter clusters potentially as a special case and look only at the other syllable onsets, this gives a total of 10 conditions (/p, t, k, pl, kl/ in two prosodic conditions). If these conditions are averaged over speaker for each language and the resulting 10 matched pairs are compared in a Wilcoxon signed ranks test, then the language difference is in fact significant, just reaching $p < .01$, but values for German are still longer than French by only about 12 ms.

For the comparison of Dutch and French, the main effect of language was significant in the ANOVA ($p < .01$). Because of the very different patterns on the rhotic clusters, the Language × Syllable Onset interaction was again highly significant ($p < .001$). Leaving out the rhotic clusters in a signed-ranks test as above, the language difference easily reaches $p < .01$, with French exceeding Dutch by 24 ms (i.e., quite a comfortable difference even without the possibly exaggerating effect of the rhotic clusters).

In any case, although there are some complexities associated with the interactions, the statistical tests confirm the unexpected finding visible in Figure 4 that French is closer to German than to Dutch (on the basis of previous findings mentioned in the introduction, we had expected that French might be further away from the unaspirated end of the VOT continuum than Dutch, but we had not expected it to actually be closer to German).

Looking in more detail at the syllable-structure factor (as before with separate four-factor ANOVAs for P vs. PR and P vs. PL) confirms that there is a very strong effect of syllable structure in the singleton-rhotic comparison, as well as a highly significant Language × Syllable-Structure interaction ($p < .001$ in both cases).

For the singleton-lateral comparison, the syllable-structure effect again showed significantly higher values for the clusters but less so than for the rhotics ($p < .01$). However, the Language × Syllable-Structure interaction was not significant.

Relative Timing of Peak Glottal Opening

Turning to direct measurements of laryngeal activity, we will look first at the main measure of laryngeal-oral coordination, namely, the timing of peak glottal opening relative to the oral occlusion.

The first main point is that, on the background of the values for occlusion duration and VOT, it was to be suspected that French would show a timing pattern intermediate between Dutch and German. This is indeed very clearly the case (see Figure 5; for the following observations, it may also be useful to refer to the overviews of temporal structure in Figure 8).

Illustrating this for the singleton stops, we find for Dutch a value of about 0.5 (for both the neutral and strong prosodic condition), which corresponds to the expectation that for a voiceless unaspirated stop, peak glottal opening is roughly in the middle of the oral occlusion. Similarly, for the aspirated stops of German, a wide-open glottis at the release of the oral occlusion is expected (i.e., values of about 1; in fact, values are generally slightly greater than 1, indicating the glottis has still not quite reached its maximum opening at oral release). The intermediate value of about 0.75 for French indicates that, unlike German, the glottis is already closing by the time of oral release but is still far enough from actual closure to allow for a substantial period of voicelessness after release. A general trend over all languages is that peak glottal opening is timed later in the clusters than the singletons. This is particularly striking for the rhotic clusters of German and French (values well above 1 for both languages and both prosodic conditions), and it will be recalled that it was precisely these items that had the longest VOTs.

In a three-factor ANOVA (language, prosody, syllable onset) with all three languages, there is a highly significant effect of language ($p < .001$). However, because this is a trivial result, we will look in more detail once again at ANOVAs comparing French, the middle language, with the more extreme languages. In the corresponding three-factor ANOVA including only French and Dutch, the effect of language is marginally significant at $p < .05$, $F(1, 6) = 10.7967$. There is, however, also a marginally significant Language × Syllable Onset interaction ($p < .05$), reflecting the fact that the differences between French and Dutch are particularly strong in the rhotic cluster (the main effect of syllable onset and prosody will be examined below). Nevertheless, even in the nonrhotic onsets, the difference between French and Dutch is consistent. In all 10 matched pairs of nonrhotic syllable onsets (constructed as above for the VOT results), the mean French value is greater than the Dutch value, giving $p < .01$ in the Wilcoxon test. In a similar comparison of French and German, the main effect of language is once again marginally significant at $p < .05$ in the ANOVA, $F(1, 7) = 7.3102$, but there is now no interaction between language and syllable onset. In the matched-pair comparison across syllable-onset conditions, the language difference is again completely consistent: German has higher values than French for all conditions (16 conditions including rhotics, 10 without).

Statistical tests of the influence of syllable structure once again compared P versus PL and P versus PR separately. Timing of the rhotic cluster differed from the singleton at $p < .001$, $F(1, 10) = 75.2859$ (i.e., later timing of peak glottal opening in the cluster). There was no interaction with language or prosody (nor with place of articulation of C1). The main effect of syllable structure was significant at the same level in the lateral cluster versus singleton comparison, $F(1, 10) = 72.7009$, but here the interaction with language was significant at $p < .01$, $F(2, 10) = 9.3890$, because, as can be seen in the figure, the tendency for later timing in the cluster is weaker in German than in the other two languages.

Concerning the prosodic factor, a comparison of data points matched for language and syllable structure in the left and right subblocks of Figure 5 reveals quite a consistent trend (both across languages and syllable types) for values to be somewhat lower in the strong prosodic condition. In other words, peak glottal opening is timed slightly earlier in the stronger prosodic condition. Accordingly, in the three-factor ANOVA with all three languages, prosody is significant at $p < 0.01$, $F(1, 10) = 11.6479$, and there is no interaction with either language or syllable onset.

Peak Glottal Opening Magnitude

Results for the magnitude of peak glottal opening are shown in Figure 6. This provides some evidence that it is not just in terms of timing but also in terms of movement amplitude that French occupies an intermediate position between German (largest amplitudes) and Dutch (smallest amplitudes). We recall that, because of technical difficulties with transillumination, the amplitude cannot be calibrated in absolute physical units but is here expressed as a normalized value relative to the average movement amplitude in target words with fricative onsets (this reference value being calculated separately for each subject and block of repetitions). It should also be taken into account that, as discussed in the Method section, this normalization procedure may have some limitations because of different distributions of male and female speakers across languages. In any case, at the output of the normalization, a value of 1 indicates a comparable glottal opening to the fricatives. For German, the grand average over all the data points in the figure is indeed of roughly this order (0.95), whereas for French and Dutch, the overall average is roughly half the fricative value, namely, 0.54 and 0.41, respectively.

Looking at the results in more detail, one of the main points of interest is that German and French have in common a particularly large glottal opening in the rhotic clusters (i.e., adding a rhotic to the syllable onset results in a larger glottal opening than in the singleton). Taken together with the timing measurements, this indicates that speakers are actively aiming for a substantial amount of glottal opening over a substantial part of the rhotic segment (i.e., they are aiming to ensure its realization as a clear voiceless fricative; this is the condition in which French speakers, in particular, are closest to the fricative reference value of 1).

The other main point of interest concerns the prosodically related differences in peak glottal opening; these reveal a further key area in which French patterns together with German, rather than lying between German and Dutch: German and French both show a clear increase in movement amplitude in the prosodically stronger condition, whereas the change for Dutch is absolutely negligible.

Filling in the statistical details, we look directly at the comparison of French with the other two languages (running an ANOVA with all three languages would here be particularly uninformative because the German-Dutch difference is once again essentially trivial and because the obvious interactions visible in the figure are easier to discuss if the material is subdivided).

In the comparison of German and French, the language factor is weakly significant in the ANOVA at $p < .05$, $F(1, 7) = 7.7519$, but as with some of the other findings above, the pattern is actually very consistent because in all 16 matched-pair comparisons over syllable onset and prosodic conditions, the value is higher in German ($p < .001$ in the Wilcoxon test). Also not surprising from perusal of the figure is that the prosody factor is clearly significant, $p < .01$, $F(1, 7) = 20.9764$, as well as the syllable onset factor, $p < .001$, $F(7, 49) = 4.7651$. Of more interest is the fact that there is no interaction between language and prosody nor between language and syllable onset, confirming the impression from the figure that prosodic strengthening and syllable onset complexity result in rather similar changes for these two languages. To give a supplementary idea of the consistency of the prosodic effect in the two languages (which will be particularly relevant in the comparison of French and Dutch immediately below), we computed a variant of the matched-pair procedure used for some of the language comparisons above. For this, we compare within each language strong-neutral prosodic pairs matched for speaker and syllable-onset condition. This gives for German a total of 40 pairs (five speakers × eight syllable-onset conditions), of which 35 had higher values in the strong prosodic condition. For French, there are 32 pairs (only four speakers), of which 30 are higher in the strong prosodic condition ($p < .001$ for both languages in the Wilcoxon test).

In contrast, in the comparison of French and Dutch, the main effect of language is not significant, but the interaction of language and prosody is marginally significant at $p < .05$, $F(1, 6) = 6.6386$. Thus, we cannot claim that French categorically has a larger glottal opening than Dutch (note the virtually identical values for P and PL in the neutral prosodic context). Rather, the difference between the languages lies in the fact that only in French is glottal opening sensitive to prosodic condition. This becomes strikingly clear with the matched-pair procedure: Dutch also has 32 pairs (four speakers, as in French), but in only 18 does the strong prosodic condition show higher values. The Wilcoxon test shows no significant effect of prosody for Dutch.

On the basis of Figure 6, one might also have expected a significant interaction between language and syllable onset (French appears to increase for PR relative to PL, whereas Dutch decreases). However, this was not in fact significant. This may be simply due to the inevitably small number of subjects in a transillumination/endoscopic study. But note the particularly long error bars for the French rhotic clusters. This reflects the fact that the overall increase in the rhotic clusters (compared with the singletons) actually consists of a very strong increase for two subjects but essentially no difference for the other two.

Glottal Gesture Duration

When considering the pattern of results for glottal gesture duration (see Figure 7), it is quite useful to compare this measure of the activity of the laryngeal articulator with the results for oral occlusion duration (see Figure 3 above) because the latter is the measure available to us that most closely reflects gestural duration of the oral articulators. Consideration of Figure 7 reveals some differences between the languages, although these are noticeably less clear-cut than for the oral data in Figure 3. Interestingly, Dutch and French have somewhat longer durations for the glottal gesture than German, which serves to indicate that longer duration of the abduction-adduction cycle does not have to lead automatically to longer VOTs. Possibly, the fact that Dutch and French have longer durations of both the laryngeal and oral gestures may reflect some slight differences in overall speech rate (slower for these two languages than for German) that we have not tried to factor out here. Similarly to the language effect, it is noticeable that the effect of prosodic strengthening goes in the same direction for glottal and oral gestures, but the lengthening of the glottal gesture with prosodic strengthening is clearly weaker than that for the oral occlusion, amounting to about 2.5 ms, 9 ms, and 15 ms for German, French, and Dutch, respectively, for the glottal gesture, compared with 8 ms, 14 ms, and 20 ms for the oral data. This expresses the lengthening in absolute terms. In relative terms (actually more relevant for discussion of timing below), the lengthening of the glottal gesture is even weaker, given the much longer overall duration of the glottal gesture compared with the oral occlusion. In a three-factor ANOVA, both language and prosody were marginally significant at $p < .05$; by comparison, as seen above, for oral occlusion duration, language and prosody were more clearly significant at $p < .01$ and $p < .001$, respectively. It will be noticed in Figure 7 that for French and Dutch, there is substantially more variability than for German. This may reflect somewhat lower accuracy in measuring gestural onsets and offsets when the amplitude of glottal opening is smaller. Syllable onset, the remaining factor in the three-factor ANOVA, was clearly significant ($p < .0001$). To look at this in more detail, we subdivided the material as before. In the comparison of the singletons with the rhotic clusters, there was a clearly significant effect ($p < .001$) for the rhotic clusters to have longer gestural durations than the singletons. There was a weak interaction ($p < .05$) of onset type with language, reflecting cross-language differences in the clarity of the increase in the rhotic clusters. Analyzing in terms of matched pairs as above, for French, 24 matching pairs of rhotics and singletons can be formed, in 23 of which gestural duration is longer (by about 30 ms) in the rhotic clusters ($p < .001$ in a signed-rank test). For German, 21 of 30 pairs are longer (average difference about 10 ms; $p < .01$ in the signed-rank test). For Dutch, the average difference between onset types was only slightly less than in German, but as already mentioned, Figure 7 shows more variability for Dutch than German (longer error bars), and the matched pairs test was not significant for Dutch (16 of 24 pairs longer for the rhotic clusters).

In contrast to the rhotic clusters, the comparison of the singletons with the lateral clusters showed no evidence at all for differences in gestural duration.

General Discussion

We start by giving a brief summary of the main trends in the results with respect to the three independent variables in the investigation (i.e., language, prosodic strength, and syllable structure). Figure 8 shows this summary in graphical form.

The effect of language is the easiest to summarize because of the very pervasive pattern for French to be located between German and Dutch with respect to the key experimental variables, that is, occlusion duration (German short, Dutch long), VOT (German long, Dutch short), relative timing of peak glottal opening (German roughly at the release of the occlusion, Dutch roughly at the midpoint of the occlusion), and peak glottal opening amplitude (German large [about the same amount of abduction as for voiceless fricatives], Dutch small [about half that of fricatives]).

Prosodic strengthening led to consistently longer durations of the oral occlusion but surprisingly had no effect on VOT in any language. It was associated with slightly but consistently earlier timing of peak glottal opening. In German and French, but not in Dutch, an increase in the magnitude of peak glottal opening was found.

With respect to syllable structure, there was quite a consistent trend for the occlusion duration of C1 to decrease from the singletons via the lateral clusters to the rhotic clusters, with the opposite pattern for VOT. The long duration of VOT in the rhotic clusters was especially pronounced in German and French. Not surprisingly, the timing of peak glottal opening patterned very similarly to VOT, being timed latest for the rhotic clusters. This was particularly noticeable for German and French, with the peak glottal opening often not being reached until well after the release of the C1 occlusion. The amplitude of glottal opening showed the least clear pattern, but there was some evidence for larger opening in the rhotic clusters of German and French (compared with lateral clusters and singletons).

We will now relate the results in more detail to the specific predictions made in the introduction, considering the mechanisms underlying the observed behavior and the implications for characterizing voicing specifications across languages.

It is convenient to start with a striking case of where one of the main initial predictions turned out to be too conservative: The expectation on the basis of previous findings that Dutch and French might not show identical patterns was clearly confirmed (in the expected direction of longer VOT in French). However, we had not specifically envisaged the possibility that French might be closer to German than Dutch. In terms of VOT, many of the realizations in French would be regarded as voiceless aspirated (i.e., the category found for German) rather than voiceless unaspirated as expected from traditional descriptions. This makes it instructive to consider more generally the mechanisms by which languages may vary VOT. Figures 3 and 4 show a striking complementary pattern across languages for occlusion duration and VOT, that is, longer occlusions associated with shorter VOT and vice versa (also across the syllable-structure condition, which will be discussed below). There may be a tendency both across languages and across contexts (such as place of articulation) for voiceless consonants to have a rather similar glottal gesture duration (leading to a rather similar total duration of voicelessness). Any variation in occlusion duration (across languages or across place of articulation) then effectively varies the point in the glottal abduction-adduction cycle at which release of the oral occlusion occurs, thus in turn directly affecting VOT. (This scenario can perhaps best be visualized by referring to the schematic timing patterns in Figure 8. For any given item, consider the effect of varying the position of the right edge of the P bar while keeping the total length of the ABD and ADD bars constant. For more background for such a scenario, see, e.g., Bombien & Hoole, 2013; Hutters, 1985.) The pattern for relative timing of peak glottal opening shown in Figure 5 follows the VOT pattern very closely, which is not surprising because the timing of peak glottal opening is the interarticulatory coordination relation most directly responsible for VOT at the acoustic surface. But the scenario just outlined above indicates that this key coordination relation may in turn be a by-product of durational control of the oral occlusion. In the simplest version of this scenario, there would indeed be constant duration of the glottal gesture over languages. For the present material, this turned out to be not strictly the case (refer back to Figure 7): The duration for Dutch and French was actually slightly longer than for German. However, the start of glottal abduction (GESTstart in Figure 2, left edge of ABD bar in Figure 8) was later in German by about the same amount (approximately 20 ms), so the net result is indeed that VOT duration has a very direct inverse relation to the duration of the oral occlusion.

These ballpark results for overall language differences indicate that the basic laryngeal abduction-adduction cycle (captured here in gestural duration) may actually be quite similar across languages; it is the coordination of oral events with this glottal cycle that can differ radically while at the same time allowing for a continuum of behavior between the possible extremes. As just observed, glottal gesture duration may not be the same in a hard and fast sense across languages; the more important point, as noted in the Results section, is that the variation across languages is much less systematic for this variable than for the other variables we have examined.[3]

We will now consider the implications of the similarities and differences between the languages with respect to the experimentally induced variation of syllable structure and prosodic strength.

It is convenient to turn first to the syllable structure because, as just mentioned, the results are at first sight similar to the cross-language results in that a pattern of decreasing occlusion duration, here from singleton via lateral cluster to rhotic cluster, is paralleled by increasing VOT (this pattern is clearest for French and German; for Dutch, the distinction is rather between the singleton and the two clusters). Following the scenario outlined above linking occlusion duration, VOT, and glottal gesture duration, this could mean that any devoicing in the second element of the clusters (/l, r/) is a passive coarticulatory effect of the reduction in C1 duration, whereas duration of the glottal devoicing gesture stays much the same. However, there are indications that specifically for the rhotic clusters of German and French, a more active account may be required. The timing of peak glottal opening is strikingly late in the rhotic clusters of these two languages (paralleled by particularly large VOT values), indicating that the increase in VOT (relative to the singletons) may not simply be accounted for by the reduction in occlusion duration. It might rather be the case that the late timing of peak glottal opening could also be brought about by a lengthening of the duration of the glottal gesture. Indeed, this was precisely what we found (see, e.g., Figure 7): a lengthening of the gesture in the rhotic clusters that was most robust for French and least robust for Dutch, while no language showed any lengthening in the lateral clusters. Taking these timing results together with the results for the magnitude of the glottal opening, which showed higher values for rhotic clusters than singletons in German and French (albeit with some speaker variability in the latter language), this indicates that speakers of these two languages are actively aiming to increase the amount of voicelessness in the rhotic clusters. Even though this confirms earlier findings by Jessen (1999) for German, this is nonetheless at first sight an unexpected result because /r/ would not normally be regarded as having an active specification for glottal abduction (and so a combination of, e.g., /p/ + /r/ should not result in two smaller glottal gestures blending into one larger one).[4]

However, this may be quite a natural process in cases where /r/ is realized with a dorsal constriction, which gives conditions that are very unfavorable for voicing at the release of the plosive (see Hoole & Bombien, 2014). Thus, speakers could be reinforcing the tendency toward voicelessness by strengthening the glottal abduction already present for the plosive itself. The more striking result is that a similar pattern occurred for French. The presence of a strong glottal abduction gesture in these clusters in French is mysterious if the voiceless plosives are assumed to be laryngeally unspecified or even specified as {–spread glottis} at the oral release, because such a representation would hardly predict amplification of the glottal abductory movement in specific contextual conditions. But the problem dissolves if, for many French speakers, voiceless plosives are assumed to be represented in terms of an active glottal abductory movement that is simply timed somewhat differently from German and English.

Turning finally to the prosodic condition, we have already mentioned that the lengthening of the oral occlusion under prosodic strengthening is a completely expected effect for all languages, confirming that the structure of the corpus and the elicitation procedures worked as intended. The first issue to consider in this section is how this result links

[3]A convenient way to drive home this point about glottal gesture duration is to refer briefly to place of articulation. This could not be dealt with in detail here, but the full results are available in the online supplemental materials (see Supplemental Material S1 and S2 for occlusion duration, S3 and S4 for VOT, and S9 and S10 for glottal gesture duration). These show very clearly the systematic effects of place of articulation (/p, t, k/) on duration of oral occlusion and VOT that are well known from the literature (e.g., Cho & Ladefoged, 1999). As in the cross-language case, there is an inverse relationship between occlusion and VOT (e.g., long occlusion but short VOT for labials). In contrast, glottal gesture duration shows no trace of a systematic influence of place of articulation.

[4]Clusters such as /pr/ are usually discussed under headings such as "sonorant devoicing" (e.g., Jessen, 1999) or "voicing assimilation in mixed voicing clusters" (e.g., Colantoni & Steele, 2007). Certainly, there have been few calls for a phonemic distinction between voiced and voiceless /r/ in English (and German) on the basis of minimal pairs such as "prat" and "brat," even though at the acoustic surface, the difference essentially involves the voicing properties of the rhotic!

up with other less expected results. In particular, for German, it was somewhat unexpected that VOT did not lengthen. For VOT to lengthen, it would be necessary for the speaker to time peak glottal opening later relative to the oral occlusion. Because the timing is actually slightly earlier in the strong prosodic condition, this indicates (following on from the kind of timing arguments already used in this Discussion section) that the glottal gesture duration is not lengthening in the strong prosodic condition, or at least not as much as the oral occlusion duration is. In any case, the resulting timing patterns indicate that the German speakers are not aiming to actively lengthen VOT, at least not in the prosodic context used in this experiment.[5]

In fact, all languages showed slight lengthening of the glottal gesture under prosodic strengthening, but as noted in the Results section, in all cases this was less than the lengthening of the oral occlusion both in absolute terms but even more clearly and more relevantly in relative terms given the longer overall duration of the glottal gesture than the oral occlusion (see Figure 7; see also Figure 8). Overall, this continues the picture that even if the duration of the glottal gesture does not remain constant in a hard and fast sense, it varies much less over languages and phonetic conditions than oral occlusion duration does. Note also that although we did not find the slight shortening of VOT in Dutch that might have been expected from the literature, the present results do provide enough information to indicate that there are grounds for caution in interpreting a shortening of VOT (which might be found in other prosodic conditions) as strengthening of a phonetic feature such as {–spread glottis}. The timing patterns observed here indicate that shorter VOT could just emerge as passive fallout of the more robust increase in the oral occlusion duration and could even be compatible with a slight increase in glottal gestural duration.

Although glottal timing analysis thus does not show notable effects of the prosodic condition, the situation with respect to the magnitude of the gesture is rather different.

The results for German conformed to expectations in that the magnitude of the laryngeal abduction gesture did increase under prosodic strengthening. Interestingly, this was the case for French as well, which fits in with the other results just discussed, suggesting that French voiceless plosives have an active specification for glottal spreading, with a timing pattern that is still different from German but nonetheless results in substantially positive VOT values. This active glottal spreading can then be targeted by the phonetic reinforcement processes forming part of prosodic strengthening (interestingly, Benguerel et al., 1978, also observed increased glottal abduction for French in a rather different case of prosodic strengthening, namely, in emphatic stress). The results for Dutch were different, because there was no tendency toward an increase in the magnitude of the glottal abductory movement in strong prosodic contexts. Accordingly, although French and Dutch are traditionally regarded as typologically similar with regard to the voicing distinction, we have found further confirmation for hints in the literature that the articulatory representation of the voiceless consonants may well have started to diverge, an interpretation consistent with the syllable-structure results already discussed.

Taking stock now of the overall patterns in the results, the consistent finding of French being located between Dutch and German (and sometimes closer to German) fits in well with the recent work of Torreira and Ernestus (2011) mentioned in the introduction, who compared various processes involving voiceless stops in French and Spanish. Specifically, they found more devoicing of vowels between voiceless stops in French than in Spanish and less lenition of intervocalic voiceless stops in French than Spanish, with both findings being consistent with a larger and/or more robust glottal abduction gesture in French.

Accordingly, in French, it is easy to find situations in which the release of voiceless plosives is followed by a substantial amount of devoicing, whether, for example, originally related to a propensity in the language for vowel devoicing or to the difficulty of reinitiating voicing in dorsal rhotics. Thus, this may have resulted in a general weakening of constraints in the language to avoid aspirated plosives. The fairly large range of variation for VOT in the French data of Figure 4 indicates that, currently, the language has a fair amount of latitude in timing of voice onset. This may also explain why the French VOT values found here are higher than those found, for example, in our own previous work (Bombien & Hoole, 2013). Because of the constraints of the transillumination technique, all of the target consonants were followed by a high vowel (whereas in our earlier experiment, which used electromagnetic articulography, we used predominantly low vowels). There is a reasonably clear tendency for VOT be longer in high-vowel contexts (e.g., Docherty, 1992; Nearey & Rochet, 1994). Thus, if French is currently showing a general shift toward longer VOT, this may be currently most evident in contexts where longer VOT simply involves “going with the flow.” Of course, in a transillumination experiment, we are not in a position to probe the full range of possibly relevant coarticulatory, prosodic, and especially sociophonetic contextual effects (and cannot categorically exclude some influence of the Germanic setting in which the experiments took place), but there are suggestive indications in the literature. For example, Nearey and Rochet (1994) found that VOT in French varied much more widely than in English over place of articulation and vowel context. Kirby and Ladd’s (2015) data mentioned in the introduction for voiceless stops in French and Italian show longer VOT for French and, perhaps more to the point, a much less compact distribution of VOT values for French, with the upper tail of the distribution reaching close to 100 ms.

From a more phonological perspective, the results here for French also have an interesting parallel to the work of Beckman et al. (2011) on Swedish mentioned in the

[5]There may also be a kind of ceiling effect at work with respect to VOT. As discussed in more detail below, the high-vowel context that had to be used in these experiments means that VOT tends to be at the upper end of its range anyway.

introduction. Their results, following a similar rationale to that followed here, led them to suspect that a nonredundant representation of voicing in terms of a single privative feature may not be appropriate for all languages. A similar situation may be starting to apply to French (because we are not aware of any evidence that the voicing lead in the voiced stops is weakening, i.e., it is probably not the case that the voicing distinction is shifting wholesale to more positive VOTs; cf. Kirby & Ladd, 2015).

However, as a general conclusion, we would prefer to avoid, if possible, any representation in terms of static, atemporal features. In line with a tradition going back at least to Löfqvist and Yoshioka (1981), it seems to us that a more gesturally oriented representation (i.e., a representation with an intrinsic representation of time and of speech as coordinated behavior) allows a much more natural account of contextual effects involving voicing and of how languages may change over time as a result of subtle shifts in intergestural timing. Many of the results shown here in effect flesh out some of the original insights of Browman and Goldstein (1986), in which they showed, for example, that the lack of plosive aspiration in English /sp/ clusters and the devoicing of /l/ in /pl/ clusters can be captured very naturally by means of gestural representations. The possibility that French may gradually be shifting toward later glottal timing is a very natural change in gestural terms; such gestural realignment clearly occurs diachronically, as, for example, recently documented for an ongoing change in some Spanish dialects from preaspiration to postaspiration (Parrell, 2012). Formulating such shifts in gestural terms avoids the necessity for categorical decisions as to where in its development a language may change from one typological pattern to another ([voice] to [spread glottis]) or where, as in Beckman et al.'s (2011) proposal for Swedish, the language may change from a typical nonredundant feature specification to a redundant one (with both [voice] and [spread glottis]).

Thus, categorical timeless representations do not seem to allow straightforward predictions as to the possible form of natural diachronic changes of the kind just sketched out and, relatedly, obscure the potential for a clear continuum of gestural behavior as seen in the present study from Dutch via French to German. As a concluding point here, we recall the reciprocal relationship across languages between occlusion duration and VOT. As we have seen, this can be formulated very naturally in gestural terms, but the relationship must seem purely coincidental in any representation that is blind to coordination relations.

A final point that remains unclear, and which will require further work, is why German and French speakers actually increase the magnitude of glottal opening in the strong prosodic condition. Even if the finding was not unexpected, at least for German, given the many articulatory correlates of prosody that have been found, the greater glottal opening cannot be part of a set of adjustments to increase the duration of voicelessness because we observed above that VOT was only weakly influenced by prosody. Future work will need to look in detail at the acoustic properties of the burst and aspiration phase of the plosives (as well as the frication phase of fricatives, which showed similar increases in glottal opening not presented here) for prosodically related differences that could be useful to the listener in recovering the prosodic structure of the utterances.

Acknowledgments

This research was supported by DFG Grant HO3271/3-2 in SPP1234 to Philip Hoole. The authors thank René Bombien and René Schramm for performing the endoscopy, Elizabeth Heller for help with segmentation and kinematic analysis, and Ruben van de Vijver for advice on the Dutch corpus. Above all, we thank our subjects for submitting to such an extensive endoscopic examination.

References

Abdi, H. (2007). The Bonferonni and Šidák corrections for multiple comparisons. In N. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 103–107). Thousand Oaks, CA: Sage.

Beckman, J., Helgason, P., McMurray, B., & Ringen, C. (2011). Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics, 39,* 39–49.

Benguerel, A.-P., Hirose, H., Sawashima, M., & Ushijima, T. (1978). Laryngeal control in French stop production: A fiberscopic, acoustic and electromyographic study. *Folia Phoniatrica, 30,* 175–198.

Bombien, L., & Hoole, P. (2013). Articulatory overlap as a function of voicing in French and German consonant clusters. *Journal of the Acoustical Society of America, 134,* 539–550.

Browman, C., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3,* 219–252.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics, 27,* 207–229.

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.

Colantoni, L., & Steele, J. (2007). Voicing-dependent cluster simplification asymmetries in Spanish and French. In P. Prieto, J. Mascaró, & M.-J. Solé (Eds.), *Segmental and prosodic issues in Romance linguistics* (pp. 109–129). Amsterdam, the Netherlands: John Benjamins.

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*. Berlin, Germany: Walter de Gruyter.

Fuchs, S., & Toda, M. (2010). Do differences in male versus female /s/ reflect biological or sociophonetic factors? In S. Fuchs, M. Toda, & M. Zygis (Eds.), *Turbulent sounds: An interdisciplinary guide* (pp. 281–302). Berlin, Germany: De Gruyter Mouton.

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* (p. 194–300). Cambridge, UK: Cambridge University Press.

Hoole, P., & Bombien, L. (2014). Laryngeal-oral coordination in mixed-voicing clusters. *Journal of Phonetics, 44,* 8–24. https://doi.org/10.1016/j.wocn.2014.02.004

Hutters, B. (1985). Vocal fold adjustments in aspirated and unaspirated stops in Danish. *Phonetica, 42,* 1–24.

Iverson, G., & Salmons, J. (2007). Domains and directionality in the evolution of German final fortition. *Phonology, 24,* 121–145.

Jessen, M. (1999). Redundant aspiration in German is primarily controlled by closure duration. In J. J. Ohala, Y. Hasegawa,

M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the XIVth International Congress of Phonetic Sciences* (pp. 993–996). Berkeley: University of California.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language, 60,* 286–319.

Keating, P., Cho, T., Fougeron, C., & Hsu, C. (2003). Domain-initial articulatory strengthening in four languages. *Laboratory Phonology, 6,* 143–161.

Kirby, J., & Ladd, D. R. (2015). Stop voicing and F0 perturbations: Evidence from French and Italian. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: University of Glasgow.

Löfqvist, A. (1980). Interarticulator programming in stop production. *Journal of Phonetics, 8,* 475–490.

Löfqvist, A., & Yoshioka, H. (1981). Interarticulator programming in obstruent production. *Phonetica, 38,* 21–34.

Nearey, T. M., & Rochet, B. L. (1994). Effects of place of articulation and vowel context on VOT production and perception for French and English Stops. *Journal of the International Phonetic Association, 24,* 1–18.

Parrell, B. (2012). The role of gestural phasing in Western Andalusian Spanish aspiration. *Journal of Phonetics, 40,* 37–45.

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

Torreira, F., & Ernestus, M. (2011). Realization of voiceless stops and vowels in conversational French and Spanish. *Journal of Laboratory Phonology, 2,* 331–353.

Tsuchida, A., & Cohn, A. C. (2000). Sonorant devoicing and the phonetic realization of [spread glottis] in English. *Working Papers of the Cornell Phonetics Laboratory, 13,* 167–181.

Yoshioka, H., Löfqvist, L., & Collier, R. (1982). Laryngeal adjustments in Dutch voiceless obstruent production. *Annual Bulletin Research Institute of Logpedics and Phoniatrics, 16,* 27–35.