The articulatory modeling of German coronal consonants using TADA

Manfred Pastätter, Marianne Pouplier

Institute of Phonetics and Speech Processing, Ludwig-Maximilians-University, Munich, Germany
`manfred@phonetik.uni-muenchen.de, pouplier@phonetik.uni-muenchen.de`

Abstract

We present a modeling study on German coronals using the task dynamic synthesizer TADA to provide new aspects to the discussion whether coronals should be generally specified for a tongue body target to control the tongue-shape globally. We argue that all coronals must be specified for a tongue body target in order to avoid inappropriate dominance of the vowel during the consonant. This target must be weighted more strongly compared to the temporally co-existing vowel gesture. We further study the effects of changing the consonant:vowel stiffness ratio on CV and VC transitions. The stiffness settings employed in TADA for American English lead to general vowel diphthongization, which is inappropriate for German. Better results are obtained with a smaller consonant:vowel stiffness ratio which gives rise to faster transitions. These results raise the possibility that global stiffness settings differ between languages.

Keywords: German coronals, task dynamics, gestural model, TADA, tongue body target, stiffness

1. Introduction

In this paper, we present a modeling study of German coronals /t, d, n, s, z, l/ in CV and VC contexts. Our work specifically addresses the question of tongue body (TB) control during coronals. Besides being relevant for our understanding of articulatory synergies in speech motor control, the issue of TB control has been causally linked to exceptional phonotactic and phonological properties of liquids. For instance, Proctor (2011) proposed that TB control is cross-linguistically the defining feature of liquids setting them apart from other coronals. He observed that both Russian dark /l/ and Spanish clear /l/ coarticulated less with the vowel than other coronals. However, Geumann and colleagues do not agree with Proctor's claim since their EMA data revealed no difference in vowel induced coarticulation for /t, d, n/ and clear German /l/ (Geumann et al. 1999). Also Mermelstein (1973) mentioned that coronal stops probably have a TB target. He observed that an intervocalic alveolar stop deflects the V-to-V TB trajectory, suggesting that alveolar stops exert TB control in a synergistic manner to support tongue tip closure. Therefore it is important to gain a better understanding of how laterals, in particular clear /l/, can be differentiated from other coronals.

For this purpose we used TADA (Nam et al. 2004), a modularized, MATLAB based implementation of the linguistic gestural model and the associated task dynamic model (Browman and Goldstein 1990, Saltzman and Munhall 1989). Based on an alphabetic string input, TADA simulates utterance planning using systems of coupled oscillators. Thereby dynamically parameterized gestures are coupled to one another in a pairwise fashion, either in-phase (e.g., CV) or anti-phase (e.g., VC). Gestures specify abstract constriction goals on the basis of mass spring equations, that is, each gesture is specified for a rest position (i.e., target), as well as a stiffness and damping parameter. The stiffness parameter serves to distinguish between classes of sounds with vowel gestures having a lower stiffness setting than consonant gestures (e.g. Fowler 1980, Perkell 1969, Roon et al. 2007). Each gesture hierarchically controls an ensemble of articulators which are yoked in a task-specific fashion in order to achieve a particular constriction goal. For instance, a lip closure gesture is associated with the articulators upper lip, lower lip, jaw. If temporally overlapping gestures call on the same articulators with conflicting demands, weighting parameters specify the degree to which one gesture may dominate in its control over a given articulator. This effectively implements coarticulatory resistance. The result of the gestural planning process is a gestural score which specifies the gestures' constriction goals and their relative timing. This gestural score provides the input to the task dynamic model which calculates the articulatory trajectories of the vocal tract variables. These are in turn the basis for the computation of time-varying area functions and formant frequencies by means of the vocal tract model CASY (The Haskins Configurable Articulatory Synthesizer; see Iskarous et al. (2003) and Rubin et al. (1996)). Finally, the CASY parameters are used to drive HLsyn (Hanson and Stevens 2002) which generates acoustic output consisting of the fundamental frequency and the first four formants.

The currently released version of TADA provides a full gestural dictionary comprising parameter settings for American English phonemes. In the American English model, coronals are generally implemented by a tongue tip (TT) constriction gesture. Manner differences are implemented as follows: coronal stops /t, d, n/ are characterized by a complete closure at the alveolar ridge; additionally, nasality of /n/ is rendered by a velum lowering gesture. Lateralization of /l/ cannot be addressed directly since CASY is a two-dimensional model of the vocal tract. Thus, as an approximation to lateralization the distance between TT and palate is narrowed, however, without producing frication. The coronal constriction is accompanied by a pronounced tongue body gesture, a well known characteristic of American English dark /l/ (Sproat and Fujimura 1993). To realize frication typical for sibilants /s, z/, the constriction degree is even more narrowed compared to the lateral. In addition, the gestural specifications of /s, z/ include a TB target at the articulator (as opposed to the gestural) level.

For our current work on German coronals, we generally adapted all parameter settings of TADA to reflect the Standard High German phoneme inventory. Below we discuss three articulatory manipulations used to model German coronal consonants in CV and VC contexts appropriately.

2. Modeling of German coronals using a TB articulator target

2.1. Coronal sibilants /s, z/

As already mentioned, for American English TADA, the sibilants are by default specified with a TB target at the articulator level. This active TB control assures that the TB retains its position during the production of sibilants, thus TB is less affected by V-to-C coarticulation during sibilants compared to other coronals (Recasens et al. 1997, Stone et al. 1992). Since the same coarticulatory resistance was found for German sibilants (Geumann et al. 1999), we adopted the same specifications from the American English sibilant specifications (i.e., tongue body constriction location (TBCL) = 110 ° and tongue body constriction degree (TBCD) = 10mm).

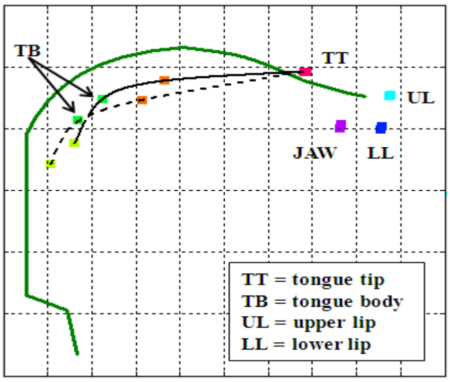

Figure 1: *The tongue configurations at the moment of maximum constriction for the stop in /ta/ before (dashed line) and after (solid line) the introduction of the TB target.*

2.2. Coronal stops and nasal /t, d, n/

Following the American English TADA assumptions, we initially specified no TB target for /t, d, n/. Modeling CV coarticulation for German coronals /t, d, n/ we found that especially in low vowel contexts the absence of a TB target resulted in an unsatisfactory acoustic and articulatory quality for all coronals alike. As to the nasal and stops, it is possible that due to the simultaneous activation of the nasal/stop and vowel gestures (i.e., in-phase coordination (Löfqvist and Gracco 1999, Browman and Goldstein 2000), the underlying synthesizer interpreted the pharyngeal constriction of the vowel (which is achieved during the consonantal closure) as the primary constriction instead of the alveolar one of the coronal. For stops, this led to a simulated supraglottal decrease in pressure. The remaining intraoral pressure was then not sufficient to cause an appropriate burst and voice onset time at the release for /t, d/. For /n/ this misinterpretation of the primary constriction led to an /N/-like percept.

The synthesis issue observed for CV held also for VC contexts: due to the absence of a consonantal TB target, the TB remained in the vowel's position while the tongue tip moved towards its alveolar place of articulation. In the task dynamic model, an articulator returns to its rest position when no longer actively controlled (Saltzman and Munhall 1989). This neutral attractor, however, has next to no effect in terms of causing the TB to move away from its preceding vowel target during the final consonant. Thus the unreduced persistence of the vowel during the consonant severely compromises the consonant's acoustic and auditory identifiability across vowel contexts.

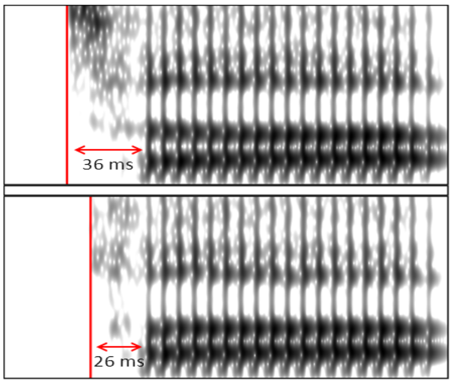

Figure 2: *Spectrograms for /ta/ with (top) and without (bottom) the TB target. Closure release is marked by the solid red line.*

In order to reduce the dominance of the vowel gesture during the consonant we introduced an articulatory TB target for /t, d, n/. As a first approximation, we adopted the parameterization for the TB target specified in the American English TADA for /s/. Since then both the vowel and the consonant call on the TB articulator, we could use the weighting parameter of TADA to specify the degree of vocal tract control. Observe in Figures 1 and 2 how the addition of a TB target during the consonant leads to a wider pharyngeal opening during the stop, a stronger burst, and a longer VOT.

As mentioned earlier, Mermelstein (1973) argued for a TB target associated with coronal stops, since he found the TB configuration actively adjusted for the coronal stop in VCV sequences. He further stated that a consonantal TB target (in conjunction with jaw position control) avoids undue extension of the tongue blade during tongue tip raising towards the alveolar ridge. We observed exactly this effect of a reduced vertical tongue stretching during tongue tip closure in TADA when a TB target was introduced (Figure 1). Mermelstein treated nasals identically to stops which we also do in our simulations.

2.3. Coronal lateral /l/

Within the gestural model it is assumed that TB control for laterals is part of their gestural specification, i.e., both the tongue tip (TT) and the TB gesture form a part of the lexically specified coupling graph (Proctor 2011). German, as opposed to American English, has a clear /l/ which is characterized by a fronted, raised TB position rather than the post-dorsal retraction typical for dark /l/ (Ladefoged and Maddieson 1996).

Recall that we initially specified no TB target for /t, d, n/. Since Geumann et al. (1999) found that /l/ and the aforementioned stops showed no differences in coarticulatory variability we suspended the gestural TB target for the German lateral in the first instance. However we observed the same acoustic and articulatory issues as earlier described for /t, d, n/, caused by an overly dominant vowel gesture. To achieve a clear quality typical for German /l/ a TB target turned out to be necessary even though with a different parameterization than for the other coronals. We based our parameter choice on the descriptions in the literature which suggest a lower TB positioning for German /l/ compared to /t, d, n, s, z/ (Ladefoged and Maddieson 1996, Wängler 1961) by setting the TB target to TBCL = 110 ° and TBCD = 13mm.

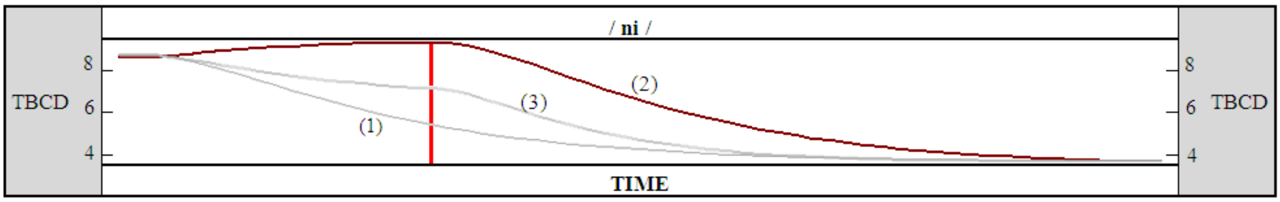

Figure 3: *Three TBCD trajectories are illustrated for /ni/ over time. The lower the value, the closer the TB is to the palate. The vertical line indicates the point in time of the constriction release. (1) The trajectory before any manipulations were done corresponding to American English TADA. (2) A TB target is specified for /n/ with the original stiffness parameter setting. (3) The final trajectory of TB constriction degree with new stiffness settings and blending of TBCD(Cons) and TBCD(Vow) to avoid diphthongization.*

3. Weighting of the TB target

The introduction of a TB target for all coronals allowed us to specify dominance relations between consonantal and vowel TB target for both CV and VC sequences. For the case of conflicting demands on the same articulator (i.e., CV), the blending parameter α is part of a gesture's specification in the task dynamic model (Saltzman and Munhall 1989). To achieve complete dominance of, e.g., the vowel TB target over the consonant the blending parameter α would be set to α(Cons) = 1 and α(Vowel) = 100, for an equal blending between the two parameter specifications, one would set α = 1 for both consonant and vowel. Based on extensive auditory evaluation by the authors, α was adjusted for TBCL(Cons) and TBCD(Cons) to 10 and 1, respectively. For vowels, α was set to 1 for both parameters. This relative greater weighting of the consonantal TBCL suppresses the vocalic TBCL during the consonant resulting in a wider pharyngeal space (Figure 1). A dominance of consonantal specification over the vowel is again in accordance with Mermelstein's (1973) observation that the V-to-V trajectory of the TB seemed to be perturbed by a consonantal target.

We further varied the specification of the blending/dominance parameter for coronals as a function of syllable position. In contrast to CV, for VC, there are no synchronous and thus no conflicting TB targets, since vowel and consonant are coordinated anti-phase. Thus the dominance relationship may remain well-balanced between the coronal consonant and the vowel, i.e., α = 1 for both consonants and vowels.

4. The role of articulator stiffness

We found that the introduction of a consonantal target for the TB necessitated a modification of the consonant:vowel stiffness ratio. Recall that the relative speed with which a gesture reaches or moves away from its target is specified by the stiffness parameter of a gesture's mass-spring equation. Consonants are differentiated from vowels based on this parameter: as an approximation, TADA employs globally two stiffness values for American English consonants and vowels with 8Hz and 4Hz, respectively; this means 8, respectively 4 closure-release cycles per second (Browman and Goldstein 1990, Nam et al. 2012). Hence, the consonant:vowel stiffness ratio is 2:1. Note that the stiffness parameter is a specification at the gestural task level and does not refer to muscular stiffness.

The introduction of a consonantal TB target and the concomitant decrease of anticipatory vowel coarticulation caused an increase in articulatory distance between the consonantal and vowel TB targets in CV (Figure 3). Particularly in a high front vowel context (e.g., /ni/) TB movement was then too slow to cover the distance between the /n/ and /i/ positions (see trajectory (2) in Figure 3). Thus TB passed through an articulatory position similar for an /e/ after the stop closure has already been released (see also Mermelstein (1973) on diphthongization resulting from a relatively lower peak velocity). This resulted in an audible, for German inappropriate diphthongization of the vowel (i.e., /nei/).

In the context of a low vowel the problem did not arise due to a lesser articulatory distance between the consonantal TB target and the low vowel target and a greater variability in producing an /a/ compared to /i/ (Hoole and Kühnert 1995). For American English, the problem might not occur at all, since American English vowels are generally diphthongized, in contrast to German. Since the diphthongization arose due to the CV transition being covered too slowly we increased stiffness for both consonants and vowels, resulting in an overall reduced consonant:vowel stiffness ratio (i.e., 1.6:1).

The stiffness of all tract variables associated with vowels (including TB and lips) was increased to 6Hz (Table 1). The stiffness of TT gestures for sonorants and sibilants was set to 10Hz and for stops 12Hz, respectively. The higher stiffness for stops evoked a considerably better burst quality; the possibility of a higher stiffness value for stops than for fricatives has been previously considered by Browman and Goldstein (1990). For the consonantal TB target, we specified a vowel-like stiffness level due to its hypothesized 'vocalic' property and its greater mass (Roon et al. 2007).

Table 1: *Modified stiffness settings.*

Class	Class	Sounds	Articulator	Stiffness
vowel			TB Lips	6Hz 6Hz
coronal	stop	/t, d/	TT TB	12Hz 6Hz
	sonorant	/n, l/	TT TB	10Hz 6Hz
	sibilant	/s, z/	TT TB	10Hz 6Hz

5. Discussion and conclusion

We propose in the current paper that an adequate modeling of German coronal consonants requires the specification of a TB target for all coronals to provide appropriate consonant-vowel coarticulation patterns. Our modeling work shows that the tongue-shape behind the primary tongue tip constriction of the consonant cannot be entirely dominated by the following or preceding vowel. Therefore, the consonantal TB target needs to be weighted more strongly than the vowel target. Further, we modified the relative stiffness difference between consonants and vowels in order to avoid vowel diphthongization. Our work raises the possibility that the relative stiffness of consonants and vowels and the concomitant speed of CV-VC transitions may be a language-specific setting, leading to the characteristic vowel diphthongization of American English. A language like German in which there is no general diphthongization of vowels, a different global consonant:vowel stiffness ratio may hold (cf. Laver (1978) on global language-specific articulatory settings). Generally, it is likely that stiffness should be distinguished in a much more fine-grained fashion on an articulator basis (Perkell 1969, Roon et al. 2007). Further it has been shown that the stiffness (and hence the velocity) of an articulator may vary over time and is not necessarily constant during an articulator's motion. Thus to assign for each articulator movement a single stiffness value for each (as proposed by the task dynamic model) is a simplifying assumption (Fuchs et al. 2011). These aspects clearly remain an issue for future research.

Overall, our modeling results receive support from the literature in which it has been argued that V-to-V trajectories show evidence for a consonantal TB target for coronal nasal/stops (Mermelstein 1973). A question not addressed in our work so far is how to account for differences in TB variability among the coronals as typically induced by vowel coarticulation. It is well known for instance, that sibilants coarticulate very little with the vowel, while laterals show much less coarticulatory resistance (Recasens et al. 1997, Stone et al. 1992). For German, Geumann et al. (1999) studied based on EMA data to what extent coronals differ in vowel-context induced variability. Unsurprisingly, they found less variability for the TB for sibilants relative to other German coronals. Interestingly, TB during /l/ varied with vowel-context similarly to the other coronals /n, d, t/ and would support our hypothesis that all coronals have a TB target. This contrasts with findings in Proctor (2011) that in Spanish and Russian coronal stops show more dorsal vowel-conditioned variability compared to the liquids in both languages (Spanish has a clear /l/ while Russian has a dark /l/). Proctor assumes that all liquids are specified for a TB gesture and that this differentiates coronal liquids from other coronals. At least for German our modeling results and the data of Geumann et al. (1999) speak against generalizing Proctor's findings for Spanish to other languages. While we agree that German /l/ should be specified for a TB target, this may not be a characteristic differentiating the lateral from other coronals in German. Whether the TB target for the lateral as much as for the other coronals should be considered to be controlled at the articulatory level or the gestural level (and thus participating in coupling relations at the planning level) is a question for future research. Conceivably, this differ between coronals as well as between languages.

6. Acknowledgements

Work supported by the ERC under the EU's 7th Framework Programme (FP/2007-2013)/Grant Agreement n. 283349-SCSPL. We thank Julian Baumann and Julia Stiegert for their help with the evaluation of TADA.

7. References

Browman, C. and L. Goldstein (1990). "Gestural specification using dynamically-defined articulatory structures". In: *Journal of Phonetics* 18, pp. 299–320.

— (2000). "Competing constraints on intergestural coordination and self-organization of phonological structures". In: *Bulletin de la Communication Parlée* 5, pp. 25–34.

Fowler, C. (1980). "Coarticulation and theories of extrinsic timing". In: *Journal of Phonetics* 8, pp. 113–133.

Fuchs, S., P. Perrier, and M. Hartinger (2011). "A critical evaluation of gestural stiffness estimations in speech production based on a linear second-order model". In: *Journal of Speech, Language and Hearing Research* 54, pp. 1067–1076.

Geumann, A., C. Kroos, and H. G. Tillmann (1999). "Are there compensatory effects in natural speech?" In: *Proceedings of the 14th International Congress of Phonetic Sciences*, pp. 399–402.

Hanson, H. and K. Stevens (2002). "A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn". In: *J. Acoust. Soc. Am.* 112, pp. 1158–1182.

Hoole, P. and B. Kühnert (1995). "Patterns of lingual variability in German vowel production". In: *Proceedings of the 13th International congress of Phonetic Sciences*, pp. 442–445.

Iskarous, K., L. Goldstein, D. H. Whalen, M. Tiede, and P. Rubin (2003). "CASY: The Haskins Configurable Articulatory speech synthesizer". In: *Proc. 15th ICPhS*, pp. 185–188.

Ladefoged, P. and I. Maddieson (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.

Laver, J. (1978). "The concept of articulatory settings: an historical survey". In: *Historiographia Linguistica* V, pp. 1–14.

Löfqvist, A. and V. L. Gracco (1999). "Interarticulator programming in VCV sequences: Lips and tongue movements". In: *J. Acoust. Soc. Am.* 105.3, pp. 1864–1876.

Mermelstein, P. (1973). "Articulatory model for the study of speech production". In: *J. Acoust. Soc. Am.* 53.4, pp. 1070–1082.

Nam, H., L. Goldstein, E. Saltzman, and D. Byrd (2004). "TADA: An enhanced, portable task dynamics model in MATLAB". In: *J. Acoust. Soc. Am.* 115.5,2, p. 2430.

Nam, H., V. Mitra, M. Tiede, M. Hasegawa-Johnson, C. Espy-Wilson, E. Saltzman, and L. Goldstein (2012). "A procedure for estimating gestural scores from speech acoustics". In: *J. Acoust. Soc. Am.* 132.6, pp. 3980–3989.

Perkell, J. (1969). *Physiology of Speech Production*. Cambridge, Mass.: MIT Press.

Proctor, M. (2011). "Towards a gestural characterization of liquids: Evidence from Spanish and Russian". In: *Laboratory Phonology* 2.2, pp. 451–485.

Recasens, D., M. D. Pallarès, and J. Fontdevila (1997). "A model of lingual coarticulation based on articulatory constraints". In: *J. Acoust. Soc. Am.* 102, pp. 544–561.

Roon, K. D., A. I. Gafos, P. Hoole, and C. Zeroual (2007). "Influence of articulator and manner on stiffness". In: *Proc. 16th ICPhS*, pp. 409–412.

Rubin, P., E. Saltzman, L. Goldstein, M. McGowan, Tiede M., and C. Browman (1996). "CASY and extensions to the task-dynamic model". In: *SPM–1996*, pp. 125–128.

Saltzman, E. and K. Munhall (1989). "A Dynamical Approach to Gestural Patterning in Speech Production". In: *Ecological Psychology* 1.4, pp. 332–382.

Sproat, R. and O. Fujimura (1993). "Allophonic variation in English /l/ and its implications for phonetic implementation". In: *Journal of Phonetics* 21.3, pp. 291–311.

Stone, M., A. Faber, L. J. Raphael, and T. H. Shawker (1992). "Cross-sectional tongue shape and linguopalatal contact patterns in [s], [sh], and [l]". In: *Journal of Phonetics* 20, pp. 253–270.

Wängler, H. H. (1961). *Atlas deutscher Sprachlaute*. Berlin: Akademie Verlag.