The effect of coarticulatory resistance and aerodynamic requirements of consonants on syllable organization in Polish

Manfred Pastätter

München 2017

The effect of coarticulatory resistance and aerodynamic requirements of consonants on syllable organization in Polish

Manfred Pastätter

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Philosophie
an der Ludwig–Maximilians–Universität
München

vorgelegt von
Manfred Pastätter
aus München

München, den 23. März 2017

Erstgutachter: PD Dr. Marianne Pouplier
Zweitgutachter: Prof. Dr. Philip Hoole
Tag der mündlichen Prüfung: 14.07.2017

Contents

List of Figures

List of Tables

Zusammenfassung auf Deutsch

Einleitung

Die vorliegende Arbeit befasst sich mit potenziellen Faktoren, die unterschiedliche Koartikulationsmuster in Konsonantenverbindungen im Silbenanlaut bedingen und wie diese Faktoren innerhalb der Artikulatorischen Phonologie (z.B. Browman & Goldstein, 1992) implementiert werden können. Zu diesem Zweck werden experimentelle Untersuchungen anhand des Polnischen durchgeführt. Dank der reichhaltigen Konsonantenphonotaktik können – anders als in vielen anderen Sprachen – mehrere Konsonanten im Silbenanlaut annähernd frei und systematisch variiert werden (z.B. Piotrowski, 1992; Rubach & Booij, 1990), um grundlegende Fragen und relevante Gesetzmäßigkeiten der Silbenorganisation zu untersuchen.

Die Aussagen, die die Artikulatorischen Phonologie hinsichtlich artikulatorischer Korrelate der Silbenstruktur trifft, sind zentrale Ausgangspunkte für diese Dissertation. Die Grundannahme besteht darin, dass sich Silbenstruktur in bestimmten vorhersagbaren Koordinationsmuster in den artikulatorischen Bewegungsabläufen (*Gesten*) niederschlägt. Für Silben mit komplexen Anlautverbindungen (CCV) wurde die „C–Center Hypothese“ als allgemein gültiges Prinzip ausgewiesen (z.B. Browman & Goldstein, 2000), die besagt, dass die Gesten der Konsonanten im Silbenanlaut *phasengleich* mit der Vokalgeste koordiniert sind, während sie untereinander jedoch *antiphasisch* koordiniert sind (Nam, Goldstein & Saltzman, 2009). Diese sich widersprechenden Phasenbeziehungen ergeben in der Theorie kontextunabhängige zeitliche Koordinationsmuster zwischen den jeweiligen Gesten der Anlautkonsonanten und der Vokalgeste (C-V) sowie zwischen den Gesten der Anlautkonsonanten zueinander (C-C). Allerdings konnte durch zahlreiche empirische Untersuchungen gezeigt werden, dass diese zeitliche Koordinations- oder Koartikulationsmuster durchaus kontextabhängige Variabilität aufweisen können (CCV Koordination: z.B. Marin (2013) und Pouplier (2012); CCV Koordination: z.B. Bombien, Mooshammer & Hoole (2013) und Chitoran & Cohn (2009)), nämlich genau dann, wenn artikulatorische und aerodynamische Anforderungen erfüllt werden müssen. Da der Einfluss dieser Anforderungen auf die Silbenorganisation bislang nicht systematisch untersucht wurde, setzt hier die vorliegende Dissertation an und bringt die C–Center Forschung mit segmenteller Koartikulationsresistenz zusammen. Dies ist ein seit Jahrzehnten bekanntes Phänomen, das unterschiedliche Grade kontextbedingter Beeinflussbarkeit von Sprachlauten beschreibt (z.B. Öhman, 1966). Konkret werden in drei experimentellen Untersuchungen die Koartikulationsresistenz sowie die aerodynamische Kompatibilität aufeinanderfolgender Silbenanlautkonsonanten systematisch variiert, um die Auswirkung dieser Faktoren auf die raum-zeitliche Organisation der Artikulationsabläufe innerhalb des Silbenanlautes zu beleuchten.

Variation der Zungen-Koartikulationsresistenz des vokalnahen Konsonanten (CCV)

Im Kern handelt es sich in dem ersten Forschungskapitel um eine klassische C–Center Studie (z.B. Browman & Goldstein, 2000) von komplexen CCV Anlautverbindungen (Clustern) des Polnischen. Ausgehend von der modelltheoretischen Annahme, dass der C–Center Effekt für den vokalnahen Konsonanten (CCV) eine steigende Vokalüberlappung bewirkt im Vergleich zu CV, erscheint es plausibel, dass die Koartikulationsresistenz dieses Konsonanten mit dem Grad der Cluster-Vokal Überlappung interagiert. Die in dieser Studie verwendeten Konsonantenverbindungen unterscheiden sich hinsichtlich der Zungen-Koartikulationsresistenz des vokalnahen Konsonanten (CCV), um drei Gruppen von CCV Anlautverbindungen miteinander verglichen zu können, eine mit *Sibilanten* (hoch koartikulationsresistent), eine mit *alveolaren Sonoranten* und eine mit *Labialen* (wenig koartikulationsresistent) in vokalnaher Position. Inwiefern diese drei Gruppen unterschiedliche Koartikulationsmuster zwischen Silbenanlaut und Silbennukleus bedingen, wurde anhand

zwei verschiedener Maße evaluiert. Bei dem ersten Maß handelt es sich um ein bereits etabliertes zeitliches Vergleichsmaß zwischen simplen CV und komplexen CCV Anlautverbindungen. Zusätzlich wird ein methodisch innovatives Maß verwendet, dass den Grad der Vokalüberlappung über den Einfluss des vokalfernen Konsonanten (CCV) auf die Zungenposition des Vokals bestimmt. Die zentrale Frage bei diesem Maß lautet wieviel Einfluss des vokalfernen Konsonanten (CCV) auf die Zungenposition des Vokals vom vokalnahen Konsonanten zugelassen wird.
Die Ergebnisse beider Maße bestätigen übereinstimmend die Hypothese, dass bei steigender Koartikulationsresistenz des vokalnahen Konsonanten die zu erwartende zunehmende Vokalüberlappung in CCV gegenüber CV eingeschränkt wird. Da der Grad der Cluster-Vokal Überlappung über die Gruppen *Labiale* > *alveolare Sonoranten* > *Sibilanten* hinweg sukzessive abnimmt, beweist, dass der Faktor Zungen-Koartikulationsresistenz des vokalnahen Konsonanten (CCV) unterschiedliche Koartikulationsmuster zwischen Silbenanlaut und Silbennukleus bedingen kann.

Variation der Kiefer-Koartikulationsresistenz des vokalnahen Konsonanten (CCV)

Dieses Kapitel schließt mit der Frage an, inwiefern unterschiedliche C-C Koartikulationsmuster auf die segmentspezifische Koartikulationsresistenz des vokalnahen Konsonanten (CCV) zurückzuführen sind. Anders als im vorangegangen Kapitel liegt hier das Hauptaugenmerk allerdings auf der Kiefer-Koartikulationsresistenz. Der theoretischen Rahmen für diese Untersuchung leitet sich von einem Modell ab, das den Kiefer als eine Art Grundbaustein der silbischen Organisation des Sprechens darstellt (z.B. Lindblom, 1983; MacNeilage & Davis, 2000). In diesem Zusammenhang wurde für CCV Silben die Hypothese formuliert, dass die Kieferhöhe vom vokalfernen Konsonanten (CCV) bis hin zum Vokal abnimmt („geschlossen-offen Zyklus“), was für die intrinsische Kieferhöhe des vokalnahen Konsonanten (CCV) eine prinzipielle Anpassung an den Grundzyklus impliziert (z.B. Redford, 1999).

Die erste Analyse in diesem Kapitel greift die Kieferzyklushypothese auf, da eine Anpassung der intrinsischen Kieferhöhe des vokalnahen Konsonanten als zunehmend unwahrscheinlich erscheint, je höher die kieferspezifische Koartikulationsresistenz des vokalnahen Konsonanten ist. Im Hinblick auf die zu erwartende Koartikulationsresistenz des Kiefers wird aus der Literatur eine Ordnung hergeleitet, in der der Grad der Kiefer-Koartikulationsresistenz (parallel zur segmentspezifischen Kieferhöhe) von /ʃ/ > /t/ > /n/ > /l/ abnimmt. Die Bewegungsbahnen des Kiefers in /mʃa/, /pta/, /pna/ und /pla/ Silben wurden anhand einer DCT-Analyse verglichen, um der Hypothese nachzugehen, dass der Krümmungsgrad (3. DCT-Koeffizient) in den Kieferzeitreihen in Abhängigkeit der Kiefer-Koartikulationsresistenz variiert (d.h. /mʃa/ < /pta/ < /pna/ < /pla/). Tatsächlich zeigen die Ergebnisse, dass sich /pla/ statistisch klar von den restlichen Silben unterscheidet (d.h. /pla/ konvexer als /mʃa, pta, pna/) und als einzige Silbe die Kieferzyklushypothese bestätigt. Ferner besitzt /pna/ eine signifikant konvexere Kieferbewegung als /mʃa/, wohingegen /pta/ sich nicht signifikant von /mʃa, pna/ unterscheidet.

In den weiteren Analysen wird auf Lindblom's (2011) (bislang ungetestete) Hypothese eingegangen, dass sowohl C-C Überlappung als auch das Phänomen der akustischen Konsonantenkürzung ($\underline{\text{C}}$V > C$\underline{\text{C}}$V) in kausalem Zusammenhang mit der Kieferzyklushypothese stehen. Ausgehend von den gegebenen Kieferzeitreihen wird erwartet, dass mit zunehmender Konvexität mehr artikulatorische C-C Überlappung (d.h. Überlappung von labialen und koronalen Bewegungen) mit einhergehender akustischer Konsonantenkürzung entsteht. Dementsprechend wird ein ähnliches Muster wie in der ersten Analyse beobachtet: /pl/ zeigt den erwartungsgemäß höchsten Überlappungsgrad, während zwischen /mʃ/, /pt/ und /pn/ (generell wenig Überlappung) kein signifikanter Unterschied nachzuweisen ist. Akustisch erfahren die vokalnahen Konsonanten /t, l/ mit zunehmender Anlautkomplexität eine Kürzung (/$\underline{\text{t}}$/ > /p$\underline{\text{t}}$/, /$\underline{\text{l}}$/ > /p$\underline{\text{l}}$/), während /ʃ, n/ statistisch gesehen nicht von akustischer Kürzung betroffen sind (/$\underline{\text{ʃ}}$/ ≈ /m$\underline{\text{ʃ}}$/, /$\underline{\text{n}}$/ ≈ /p$\underline{\text{n}}$/).

Abschließende Korrelationsanalysen bestätigen die Vorhersage, dass die Form der Kieferbewegung positiv mit C-C Überlappung und negativ mit dem Grad der akustischen Konsonantenkürzung korreliert ist. Dies sind eindeutige Hinweise darauf, dass segmentspezifischen Kieferspezifikationen (z.B. intrinsische Kieferhöhe, Kiefer-Koartikulationsresistenz) für artikulatorische und akustische Korrelate der Silbenstruktur verantwortlich sind. Einzig die erwartete Korrelation zwischen artikulatorischer Überlappung und akustischer Kürzung wird nicht bestätigt.

Aerodynamische Kompatibilität von Anlautkonsonanten (<u>CC</u>V) unter prosodischer Variation

Im Hinblick auf die biomechanisch motivierten Hypothesen des vorangegangenen Kapitels zeigt das Cluster /pn/ ein unerwartetes Ergebnis, nämlich deutlich weniger Überlappung der Anlautkonsonanten im Vergleich zu /pl/. Aus diesem Grund wird im letzten Forschungskapitel der einschlägigen Annahme nachgegangen, dass aerodynamische und damit letztendlich perzeptuelle Faktoren unterschiedliche Koartikulationsmuster in Silben mit komplexen Anlautverbindungen bedingen können. Dass Plosiv-Nasal Cluster (z.B. /pn/) systematisch weniger koartikulieren als Plosiv-Lateral Cluster (z.B. /pl/) wurde mit dem aerodynamischen Konflikt begründet, der für Plosiv-Nasal Verbindungen besteht. Ein Konflikt käme demzufolge zustande, wenn eine zu starke Überlappung der Velumsöffnung von /n/ mit dem Plosiv den erforderlichen intraoralen Überdruck beeinträchtigen und in der Konsequenz den perzeptuell wichtigen Verschlußlösungsimpuls gefährden würde (z.B. Hoole, Pouplier, Beňuš & Bombien, 2013; Kühnert, Hoole & Mooshammer, 2006).
Um verschiedene Koartikulationmuster zwischen Anlautkonsonanten (<u>CC</u>V) methodisch auszulösen, hat sich in den vergangenen Jahren die Variation der Satzprosodie etabliert. In diesem Zusammenhang lautet die globale Annahme, dass zwei Anlautkonsonanten in deakzentuierter Position stärker miteinander überlappen, als in akzentuierter Position. Allerdings – so die Kernhypothese dieses Kapitels – sollte dieser prosodische Effekt blockiert werden, wenn durch die Überlappungszunahme wesentliche akustische Eigenschaften des ersten oder zweiten Konsonanten zerstört würden (aerodynamische Konflikte).
Diese Hypothese wird hier für zwei Gruppen von Anlautverbindungen getestet (**pC**: /pl/, /pn/, /pʃ/; **mC**: /ml/, /mɲ/, and /mʃ/), in denen sich jeweils *ein* Beispiel befindet, für

die theoretisch aerodynamische Konflikte anzunehmen wären. Konkret wird angenommen, dass ausschließlich bei /pn/ und bei /mʃ/ unter prosodischer Variation eine zu starke Überlappung der Velumabsenkung mit /p/ oder /ʃ/ den erforderlichen intraoralen Überdruck beeinträchtigen könnte. Tatsächlich weist /pn/ eine qualitativ geringere prosodische Variation auf als /pl, pʃ/ (vgl. Bombien et al., 2013). Allerdings wird entgegen der Erwartungen für /mʃ/ beobachtet, dass die beiden Konsonanten in deakzentuierter Position deutlich mehr überlappen, als in deakzentuierter Position. Dass weiterhin etwa für /ml/ praktisch keine Auswirkung der prosodischen Bedingungen gefunden wurde, könnte darauf hindeuten, dass aerodynamische Kompatibilität von Anlautkonsonanten nicht allein bestimmte Koartikulationsmuster beeinflussen kann.

Schlussbemerkungen

Insgesamt bestätigen die Ergebnisse der drei experimentellen Untersuchungen, dass die Organisation der Artikulationsabläufe innerhalb von Silben mit komplexen CCV Anlautclustern davon abhängt, aus welchen Konsonanten der Silbenanlaut zusammengesetzt ist. Dabei scheinen zwei Faktoren von besonderer Bedeutung zu sein, wie die jeweiligen Anlautkonsonanten zueinander und zusammen (als globales Anlautgebilde) relativ zum Silbennukleus koordiniert sind, nämlich die Zungen- sowie die Kiefer-Koartikulationsresistenz des vokalnahen Konsonanten (C<u>C</u>V).

Der große Potenzial der Artikulatorischen Phonologie besteht darin, dass koartikulatorische Variabilität <u>und</u> Resistenz zwischen benachbarten Segmenten (z.B. in simplen CV Silben) integrale Bestandteile in diesem Modell sind. Jedoch gab es bislang keine umfassenden Ansätze, wie dieses Potenzial genutzt werden kann, um unterschiedliche Koartikulationsmuster in komplexen Anlautsilben zu modellieren (z.B. CCV Silben mit verschiedenen vokalnahen Konsonanten). In den jeweiligen Untersuchungen dieser Dissertation wird daher vorgeschlagen, das bereits in der Artikulatorischen Phonologie existierende Konzept der Bindekraft (*bonding* oder *coupling strength*) neu zu interpretieren, d.h. Bindekraft explizit als einen Wert für Koartikulationsresistenz zu verstehen. Obwohl diese Hypothese in ersten Fallbeispielen mithilfe des modelleigenem artikulatorischen Synthesizer TADA (Nam, Goldstein, Saltzman & Byrd, 2004) gestützt werden konnte, sollte in zukünftigen Untersuchungen eine exakte Evaluation und Ermittlung dieser Bindekraftwerte vorgenommen werden. Die inhärente Motivation besteht darin, jedem Konsonanten (und in der Konsequenz auch jedem Vokal) einen individuellen, kontextunabhängigen Bindekraftwert empirisch zu ermitteln, um unterschiedliche Koartikulationsmuster in Konsonantenverbindungen im Silbenanlaut synthetisieren zu können. Dies wäre eine vielversprechende Innovation im Feld der Artikulatorischen Phonologie und generell ein Fortschritt in unserem Verständnis von artikulatorischen Koordinationsmustern und Sprachproduktion.

Literatur

Bombien, L., Mooshammer, C., & Hoole, P. (2013). Articulatory coordination in word-initial clusters of German. *Journal of Phonetics*, *41*, 546–561.

Browman, C. & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*, 155–180.

Browman, C. & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée*, *5*, 25–34.

Chitoran, I. & Cohn, A. (2009). Complexity in phonetics and phonology: Gradience, categoriality, and naturalness. In F. Pellegrino, E. Marisco, I. Chitoran & C. Coupé (Hrsg.), *Approaches to Phonological Complexity* (S. 21–46). Berlin: Mouton de Gruyter.

Hoole, P., Pouplier, M., Beňuš, Š. & Bombien, L. (2013). Articulatory coordination in obstruent-sonorant clusters and syllabic consonants: Data and modelling. In L.Spreafico & A. Vietti (Hrsg.), *Rhotics. New Data and Perspectives* (S. 81–97). Bolzano University Press.

Kühnert, B., Hoole, P. & Mooshammer, C. (2006). Gestural overlap and C-center in selected French consonant clusters. In *In Proceedings of the 7th International Seminar on Speech Production* (S. 327–334).

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Hrsg.), *The Production of Speech* (Vol. 55, S. 217–245). New York: Springer.

Lindblom, B. (2011). The units of speech – A bio-cultural perspective. In *PERILUS 2011, Symposium on Language Acquisition and Language Evolution* (S. 1893–1896). Stockholm, Sweden.

MacNeilage, P. F. & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, *288*, 527–531.

Marin, S. (2013). The temporal organization of complex onsets and codas in Romanian: A gestural approach. *Journal of Phonetics*, *41*, 211–227.

Nam, H., Goldstein, L., Saltzman, E. & Byrd, D. (2004). TADA: An enhanced, portable task dynamics model in MATLAB. *Journal of the Acoustical Society of America*, *115*(5,2), 2430.

Nam, H., Goldstein, L. & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. In F. Pellegrino, E. Marisco, I. Chitoran & C. Coupé (Hrsg.), *Approaches to Phonological Complexity* (S. 299–328). Berlin: Mouton de Gruyter.

Piotrowski, M. (1992). Polish yers in non-linear phonology. In W. Dressler, H. C. Luschutzky, O. E. Pfeiffer & J. R. Rennison (Hrsg.), *Phonetica 1988* (S. 215–227). Cambridge: Cambridge University Press.

Pouplier, M. (2012). The gestural approach to syllable structure: Universal, language- and cluster-specific aspects. In S. Fuchs, M. Weirich, M., D. Pape & P. Perrier (Hrsg.), *Speech planning and dynamics* (S. 63–96). Peter Lang.

Redford, M. A. (1999). *An articulatory basis for the syllable* (Dissertation). University of Texas, Austin.

Rubach, J. & Booij, G. (1990). Syllable structure assignment in Polish. *Phonology*, *7*(1), 121–158.

Öhman, S. E. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, *39*(1), 151–168.

Chapter 1

Introduction

This dissertation presents an experimental investigation of various constraints of particular consonants and consonant sequences which are hypothesized to condition the spatiotemporal organization syllables with complex onsets. Previous research on articulatory correlates of syllable structure suggests that the temporal organization of syllables varies as a function of the segmental make-up of the onset cluster. But there is currently little understanding of the details that might condition this variation. Aiming for a thorough understanding of potential biomechanical and aerodynamic constraints on spatiotemporal syllable organization, we systematically vary the lingual (i.e tongue body) and mandibular coarticulation resistance of the consonant adjacent to the vowel, and the aerodynamic requirements of the consonants composing an onset cluster. Since the current work is framed within the theory of Articulatory Phonology and its computational relatives (i.e. task dynamics and coupled oscillator model of syllable organization) we will briefly introduce these models and how they are related to our research questions

1.1 The Articulatory Phonology framework

Articulatory Phonology (e.g. Browman & Goldstein, 1986, 1989, 1992, 1995, 2000) is a theoretical framework of phonology which specifies phonological atoms of speech in terms of vocal tract actions – the so-called 'gestures'. Thus, gestures are understood to be simultaneously *information units* and *action units* which are used in a unitary fashion to capture phonological contrasts and physical properties of speech production. Gestures control the spatiotemporal coordination of different articulators which act in a synergistic manner to fulfill an articulatorily characteristic constriction task in the vocal tract. From this perspective it is crucial that gestures are considered to be abstract, i.e. they specify articulatory tasks, but *not* for individual movements. For instance, the labial stop /p/ specifies a closure of the lips, but *not* moving the lower lip towards the upper lip. Further, gestures are inherently temporal and spatial since they are activated for a particular temporal interval to achieve individually specified constriction degrees and/or locations in the vocal tract. Another property of the gestural approach is that two individual gestures can be coordinated by means of pair-wise coupling relations to form larger phonological units (e.g. syllables). These coupling relations (specified in terms of particular phasing relationships; more on this below) resolve theoretically in either synchronously or sequentially initiated gestures. Particularly in the former case (i.e. synchronous coordination) the gestures involved overlap extensively in time, i.e. coproduction. In cases where coproduced gestures share the same articulators, coproduction might result in competing articulatory goals which in turn creates coarticulatory effects (cf. Fowler & Saltzman, 1993; Saltzman & Munhall, 1989; Öhman, 1966).

1.2 The task dynamics model

An approach that aims to model articulatory movements in terms of articulatory gestures was proposed by Saltzman and colleagues (e.g. Saltzman, 1986; Saltzman & Munhall, 1989). The task dynamic model of speech production suggests that articulatory gestures emerge from dynamical systems with two functionally distinct but interacting levels. At the *intergestural coordination level* a set of coupled activation intervals is specified in the form of a gestural score (cf. bottom panel in Figure 1.1) This gestural score is in turn applied at the *interarticulator coordination level*, at which both tract variables (e.g. for /p/: lip aperture

(LA) and lip protrusion (LP)) and articulator variables (e.g. for /p/: upper/lower lip and jaw) are specified. This theoretical approach suggests that coordinated speech gestures are predictable if both tract and articulator variables are available.

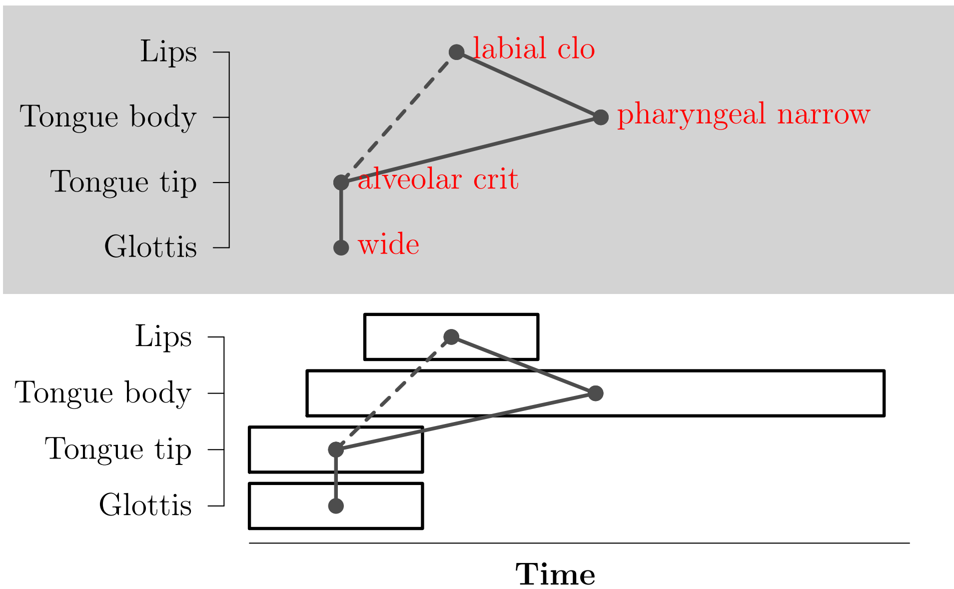

Figure 1.1: Top: The coupling graph for the syllable [spɑ] (as in Amercian English pronounciation of "spot") in which the consonantal tongue tip (fricative) gesture and the lip closure gesture are coupled in-phase to the vocalic tongue body gesture, while they are also coupled to each other in an anti-phase coupling mode. Bottom: The gestural score that results from the planning model is also shown. Lines indicate coupling relationships between pairs of gestures – solid and dashed are different in-phase and anti-phase coupling modes, respectively (cf. Goldstein et al., 2006, p. 227).

In order to model articulatory movements, the task dynamic model computationally implements articulatory gestures as sets of dynamic systems on the basis of gestural scores and mass-spring equations. More specifically, this means that once a gestural score is input into the task dynamic model, gestures are quantitatively modeled by individually assigned target specifications in the tract-variable space: stiffness and (critical) damping parameter. Eight tract variables are considered which represent the constriction location/degree for the lips, the tongue tip and the tongue body, the constriction degree for the velum (i.e.

velic aperture) and the constriction degree for the glottis (i.e. glottal aperture). Table 1.1 shows that each tract variable involves a set of different articulators which act in a synergistic manner to attain a gestural target. (e.g. the labial closure gesture for /p/ involves the upper/lower lips and the jaw).

Organ	**Tract variables**	**Articulators**
Lips	Lip protrusion (PRO)	Upper/lower lip, jaw
	Lip aperture (LA)	Upper/lower lip, jaw
Tongue tip	Tongue tip constriction location (TTCL)	Tongue tip/body, jaw
	Tongue tip constriction degree (TTCD)	Tongue tip/body, jaw
Tongue body	Tongue body constriction location (TBCL)	Tongue body, jaw
	Tongue body constriction degree (TBCD)	Tongue body, jaw
Velum	Velum (VEL)	Velum (NA)
Glottis	Glottis (GLO)	Glottis (GW)

Table 1.1: List of constriction organs, tract variables, and associated articulators used in the task dynamic model.

1.3 Coupled oscillator model of syllable organization

Further developments of the gestural approach (Goldstein et al., 2006; Nam, Goldstein, & Saltzman, 2009) provide an opportunity for intergestural timing, which is determined by the *planning oscillators* associated with the set of individual gestures, i.e. pair-wise coupling relations of two individual planning oscillators (or clocks). These coupling relations are specified for particular phasing relationships, for which only two modes of coupling are hypothesized (cf. Turvey, 1990): either they are coupled in an in-phase relationship (i.e. 0°; results in synchronously initiated gestures) or in an anti-phase relationship (i.e. 180°, results in sequentially initiated gestures) with each other. The observation that in consonant-vowel sequences (henceforth referred to as C_AV, where 'C_A' denotes the vowel-adjacent consonant), consonant and vowel gestures show a high degree of C_AV overlap and near-synchronous movement initiation (de Jong, 2003; Löfqvist & Gracco, 1999; Nam et al., 2009) is captured in the model in terms of an in-phase relationship between onset consonant and vowel.

However, the account of C-V in-phase coupling cannot be simply adapted to the intergestural coordination of syllables with a consonant cluster preceding the vowel (henceforth referred to as C_RC_AV, where 'C_R' denotes the vowel-remote consonant). If all consonants were initiated synchronously with the vowel (i.e. C_R-V and C_A-V in-phase coupling), this would result in the simultaneous production of C_R and C_A. Therefore, to prevent the perceptual masking of one of the consonants, a compromise is needed in order to model cluster-vowel timing patterns. From this perspective, Goldstein and colleagues (Browman & Goldstein, 2000; Goldstein, Nam, Saltzman, & Chitoran, 2009; Nam et al., 2009) suggested that multiple, competing coupling relations can be specified in the network of oscillators in the coupling graph. For example, in the case of syllable [spɑ], both onset consonants are coupled in-phase with the vowel, and anti-phase with each other (as shown in the coupling graph in Figure 1.1). In terms of intergestural coordination, competing phase relationships condition the consonant gestures (associated with C_R=/s/ and C_A=/p/) to shift relative to the vowel gesture, so that the onset of the vowel gestures coincides with the temporal midpoint of both consonant gestures (i.e. C-center timing pattern).

The gestural approach of coupled oscillators makes clear predictions concerning the intergestural coordination of syllables with complex onsets (C_RC_AV syllables), i.e. consonants are coupled in-phase with the vowel, and anti-phase with each other. Taken together, the competing coupling constraints should in theory render context-independent patterns of C_R-V, C_A-V, and C_R-C_A temporal coordination.

While there has been general support for the C-center model, previous research has also revealed that for some clusters, the timing of the vowel-adjacent consonant remains constant under increasing onset complexity (e.g. Brunner, Geng, Sotiropoulou, & Gafos, 2014;

Marin, 2013; Peters & Kleber, 2014; Pouplier, 2012). Yet there is currently no clear picture of the factors which may condition these unexpected patterns. There is some indication in the literature that these differences may be accounted for by taking the coarticulation resistance of the vowel-adjacent consonant into consideration (Brunner et al., 2014; Marin, 2013; Pouplier, 2012), but a systematic investigation is missing.

As to the gestural C_R-C_A timing, previous research (e.g. Bombien, Mooshammer, & Hoole, 2013; Byrd, 1996; Chitoran & Cohn, 2009; Hoole, Pouplier, Beňuš, & Bombien, 2013; Kühnert, Hoole, & Mooshammer, 2006) also hints at different factors which have been repeatedly shown to affect C_R-C_A coordination, e.g. manner of articulation of both C_R and C_A. Consonant clusters in which C_R was a stop showed relatively less gestural overlap compared to clusters with a fricative in C_R, and consonant clusters in which C_A was the nasal /n/ showed less gestural overlap compared to clusters with /l/ or /s/ in the vowel-adjacent position (e.g. Bombien et al., 2013; Kühnert et al., 2006; Yip, 2013). Hoole and colleagues suggested that the respective aerodynamic requirements of stop+nasal clusters seem to constitute an additional constraining factor on intra-cluster coordination.

1.4 Modeling coarticulation (resistance)

As pointed out earlier, due to the notion that coarticulation results from the interaction of spatially and temporally overlapping gestures (e.g. Saltzman & Munhall, 1989; Öhman, 1966), the gestural model is in principle ideally suited to model coarticulatory effects. Fowler and Saltzman (1993) considered two possible scenarios for coarticulatory interactions of overlapping gestures, depending on whether the overlapping gestures share the same articulators or not. For example, since the tongue is not a shared articulator between vowel and consonant in /apa/, there will be maximal consonant-vowel coarticulation. This contrasts with cases like /asa/ in which vowel and consonant impose conflicting demands on the same articulator (here: tongue body) with the results of gestural blending. And indeed, blending coefficients are an integral part of the task dynamic model (Fowler & Saltzman, 1993; Saltzman & Munhall, 1989), they specify the degree to which a given gesture dominates the vocal tract, i.e. coarticulation resistance and aggression. For instance, the lingual coarticulation resistance of a sibilant is implemented by a relatively greater weighting of the consonantal compared to the vocalic tongue body gesture, limiting vowel-induced variability of the sibilant (i.e. a low degree of V-to-C coarticulation). In the case of coarticulatorily least resistant labials, the blending parameter is irrelevant since the labial and the vowel do not share the same (lingual) articulator, hence there is, in terms of the tongue, no constraint on V-to-C coarticulation. While the concept of coarticulation resistance is part of the gestural and associated task dynamic model in terms of dominance coefficients for articulators (Fowler & Saltzman, 1993; Iskarous, McDonough, & Whalen, 2012; Saltzman & Munhall, 1989), there has been virtually no work so far that has looked at the interaction of temporal coordination and spatial dominance factors.

1.5 Research aims and outline of this thesis

At the heart of this thesis is the question of what Polish onset cluster can teach us about biomechanical and aerodynamic constraints of consonants on syllable-related timing patterns. We choose Polish for this work, since the great variety of consonant combination in Polish onset clusters allows us to provide systematic studies of how tongue body/jaw coarticulation of C_A and conflicting aerodynamic requirements of C_R and C_A interacts with the temporal organization of syllables. Previous research on the temporal coordination of syllable structure gives rise towards the assumption that both consonant-to-vowel and consonant-to-consonant coordination patterns vary as a function of such articulatory constraints, but systematic investigation is missing thus far.

In order to provide a theoretical foundation of how these constraints can be integrated into the gestural model of syllable organization, we use the task dynamic application (TADA, Nam, Goldstein, Saltzman, & Byrd, 2004) which is a modularized implementation of the linguistic gestural model and the associated task dynamic model (Browman & Goldstein, 2000; Saltzman & Munhall, 1989). Accordingly, this is a perfect tool to probe hypotheses which arise in the course of this dissertation.

In Chapter 2, the degree of onset-vowel overlap will be examined in terms of different degrees of lingual coarticulation resistance of the vowel-adjacent consonant (C_A = sibilants vs. alveolar sonorants vs. labials). In Chapter 3, the interaction of the jaw cycle and coarticulation resistance of the vowel-adjacent coronal will be examined by varying the segmental make-up of onset clusters. This will be done in order to estimate the effects of jaw movement on labial-coronal coarticulation and incremental shortening of the vowel-adjacent coronal. In Chapter 4 we investigate labial+alveolar clusters as to whether conflicting aerodynamic requirements prevents increasing gestural overlap under deaccentuation. Finally, in Chapter 5 we will bring together the present result and discuss them with respect to the research aims outlined here.

Chapter 2

Articulatory mechanisms underlying onset-vowel organization[1]

2.1 Introduction

Finding phonetic correlates in syllable organization has engaged phonetic studies for many years since it is difficult to tease apart universals in speech planning from language-specific and segmental composition effects. The goal of the present study is to shed light on how the segmental make-up of a consonant cluster may affect cluster-vowel organization. In particular we will focus on the coarticulation resistance of the consonant adjacent to the vowel (i.e. #C<u>C</u>V) as one possible factor conditioning differences in the timing of complex syllable onsets relative to the syllable nucleus. Throughout this chapter, the consonants immediately preceding the syllable nuclei in the singleton and cluster condition will be referred to as vowel-adjacent consonant (abbr. C_A); in the cluster condition, the first member of the cluster will be referred to as vowel-remote consonant (abbr. C_R).

Regarding the articulatory organization of the syllable, systematic timing differences have been found depending on whether the consonant precedes (#CV) or follows the vowel (VC#). In a CV sequence, the movement onset for the vowel will occur before the con-

[1] A version of this chapter has been published in the Journal of Phonetics (Pastätter & Pouplier, 2017).

sonant has reached its target, leading to a considerable degree of overlap, while a coda consonant will – with some exceptions – show considerably less vowel overlap (de Jong, 2003; Krakow, 1999; Löfqvist & Gracco, 1999). There is also some evidence that the organization of consonantal gestures with respect to the vowel differs as a function of syllable complexity (e.g. Browman & Goldstein, 1988; Byrd, 1995; Honorof & Browman, 1995). By comparing the timing of a cluster onset relative to a corresponding singleton onset, it has been demonstrated that at least in certain circumstances onset-vowel timing reorganizes dynamically when onset complexity increases (i.e. $\mathrm{C_A V} \rightarrow \mathrm{C_R C_A V}$). What is meant by reorganization is illustrated schematically in Figure 2.1: The vowel-adjacent /m/ in the cluster condition (bottom panel) starts later in time compared to /m/ in the singleton condition (top panel). The dashed line indicates the temporal midpoint of the singleton (top) and the cluster (bottom) onset, while the solid line indicates the constant anchor point relative to which the timing of the onset consonants is typically evaluated. This relative temporal shift between singleton and cluster induces the /m/ in *szmata* [ʃm̲ata][2] to overlap more with the vowel compared to the /m/ in *mata* [m̲ata]. This has been interpreted as the onset being coordinated to the vowel as a single prosodic unit: While the temporal relationship to the vowel of each individual consonant changes with increasing complexity, the timing of the onset as a whole remains the same. Within the gestural model of syllable structure, this 'global' onset-vowel organization has been termed the "C-center" effect (Browman & Goldstein, 1988, 2000).

The C-center effect is by hypothesis a universal correlate of syllable structure (Goldstein et al., 2006) and, empirically, it has by and large been confirmed for several clusters and languages (American English: Browman & Goldstein, 1988; Byrd, 1995; Honorof & Browman, 1995; Marin & Pouplier, 2010; German: Pouplier, 2012; Italian: Hermes, Mücke, & Grice, 2013; Romanian: Marin, 2013; Marin & Pouplier, 2014; but see Brunner et al., 2014). However, some onset clusters showed other timing patterns than the expected C-center,

[2]It has to be noted that different IPA symbols are used in the literature for the Polish post-alveolar sibilant in question, i.e. /ʃ/ (e.g. Gussmann, 2007; Jassem, 2003) and /ʂ/ (e.g. Bukmaier & Harrington, 2016; Hamann, 2004; Żygis & Hamann, 2003). In this paper, we use the symbol /ʃ/ throughout.

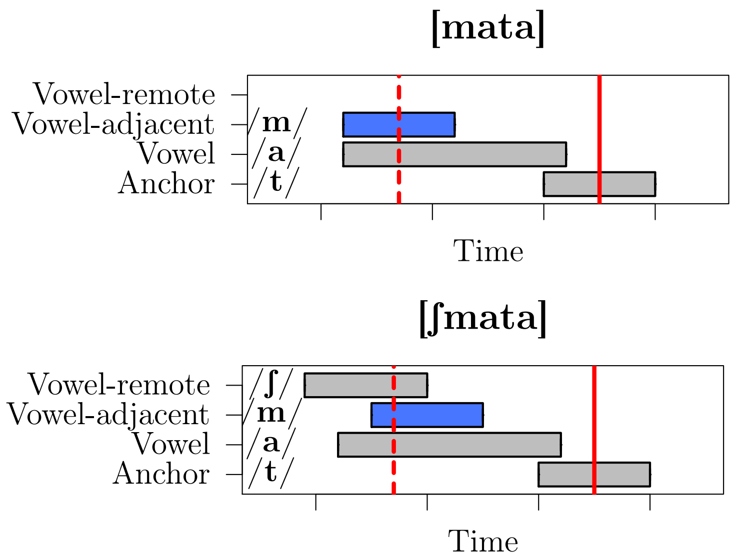

Figure 2.1: Schematic representation of the predicted "C-center" organization of singleton (top) and cluster (bottom) onsets in [mata] and [ʃmata], respectively, relative to a constant anchor point, in this case /t/. Each box represents the temporal interval during which a given structure is active.

suggesting an interaction of cluster composition and cluster-vowel organization in ways not accounted for by the gestural model of syllable structure. In particular, certain clusters fail to show the characteristic increase of onset-vowel overlap as a function of onset complexity. Yet it remains unclear which factors exactly may condition these deviating patterns. In the present paper, we propose that the coarticulation resistance of the vowel-adjacent consonant may be one of these factors. Since the concept of C-center organization is based on the observation that generally $C_A V$ overlap increases with increasing onset complexity, it is conceivable that this may be blocked if C_A is highly coarticulation resistant. However, current articulatory models of syllable structure do not foresee such an interaction between articulator (spatial) dominance and degree of temporal overlap. We focus on Polish in our current study, since Polish phonotactics allow for a great variety of consonant combinations in onsets. This enables us to carry out a systematic investigation of articulatory mechanisms which may interact with cluster-vowel timing. Specifically, we compare onset clusters with vowel-adjacent consonants which are known to differ in their coarticulatory resistance: sibilants, alveolar laterals (which tends to be a 'clear' [l][3] in Polish; Rochoń, 2000) and nasals, and labial stops. These analyses allow us to determine how different degrees of coarticulation resistance affect the temporal organization of C_AV sequences in singleton and cluster onsets. The focus of the present study is on coarticulation resistance of C_A since this is the locus at which we can expect to find the greatest effect on onset-vowel overlap. In the following we will summarize the gestural model of syllable structure as proposed within Articulatory Phonology and review previous studies that have investigated cluster-vowel timing in a similar fashion to the current study.

2.1.1 Gestural coupling model

Articulatory Phonology models phonological representations of speech sounds in terms of the spatiotemporal coordination of articulatory gestures (e.g. Browman & Goldstein, 1992). Articulatory gestures specify vocal tract actions for the production of speech sounds (e.g. a lip closure for /p/ or the combination of lip closure and velum lowering for /m/). Larger phonological units such as syllables arise from coupling relations between gestures. These coupling relations are specified for particular phase relationships. The observation that in C_AV sequences, consonant and vowel gestures show a high degree of C_AV overlap and near-synchronous movement initiation (de Jong, 2003; Löfqvist & Gracco, 1999; Nam et al., 2009) is captured in the model in terms of a 0° phase relationship between onset consonant and vowel. In the case of onset clusters, not all consonants can be produced synchronously (or 'in-phase') with the vowel, since this would result in a simultaneous production of C_R and C_A. The C-center organization (Figure 2.1) represents a compromise solution to prevent the perceptual masking of one of the consonants: figuratively, the vowel-adjacent

[3]Laterals are characterized by a tongue tip and a tongue body constriction. Depending on the target position of the tongue body movement laterals are characterized auditorily as 'clear' or 'dark': clear laterals have a fronted and raised tongue body whereas 'dark' /l/ features a retracted and lowered tongue body (e.g. Browman & Goldstein, 1995; Sproat & Fujimura, 1993). Clear laterals are for instance typical for German, Romanian, and Polish (Recasens, 2012a; Rochoń, 2000).

consonant shifts towards the vowel (i.e. /m/ overlaps more with the vowel in [ʃm̲ata] than in [m̲ata]) while the vowel-remote consonant (i.e. /ʃ/) shifts away from the vowel. This is so because by hypothesis onset clusters have competing phase relationships: while both consonants are coupled in-phase with the vowel, they are coupled anti-phase with each other (Browman & Goldstein, 2000; Goldstein et al., 2009; Nam et al., 2009). These conflicting underlying phase specifications cannot be satisfied at the same time. The result is a blended output giving rise to the C-center timing pattern described above (Figure 2.1). To infer the underlying coordination (or phase) relationship of syllables with complex onsets, cluster-vowel timing has often been investigated relative to a corresponding singleton condition, as already discussed in the context of Figure 2.1. The relative change of the temporal lag between the vowel-adjacent consonant and a constant anchor point (usually a consonant following the vowel, i.e. the /t/ in [mat̲a] and [ʃmat̲a]) from singleton to cluster condition is used as an index to consonant-vowel overlap. If the temporal distance between the vowel-adjacent consonant and the anchor point is shorter in the cluster compared to the singleton condition, this is taken to mean that the overlap between consonant(s) and vowel increases with increasing onset complexity (i.e. C-center organization; Figure 2.1). That this temporal shortening of the anchor distance is truly due to increasing consonant-vowel overlap rather than just vowel shortening has been confirmed in Peters and Kleber (2014). While there has been general support for the C-center model, previous research has also revealed that for some clusters, the timing of the vowel-adjacent consonant remains constant under increasing onset complexity, yet there is currently no clear picture of the factors which may condition these unexpected patterns. There is some indication in the literature that these differences may be accounted for by taking the coarticulation resistance of the vowel-adjacent consonant into consideration (Brunner et al., 2014; Pouplier, 2012), but a systematic investigation is missing. We will now review the previous literature from this perspective.

Generally, across languages, onset clusters with a labial in the vowel-adjacent position (e.g. /sp/, /sm/) have been shown to predominantly agree with the timing patterns hypothesized by the gestural model in that they showed more $C_A V$ overlap in cluster than in singleton onset condition (American English: Browman & Goldstein, 1988; Honorof & Browman, 1995; Marin & Pouplier, 2010; Romanian: Marin, 2013; German: Pouplier, 2012). Labials are known to have a very low degree of coarticulation resistance (see Section 2.1.2 below). Two notable exceptions are Italian (Hermes et al., 2013) and Slovak (Pouplier & Beňuš, 2011), but here there is relatively clear evidence for this being a very particular language-specific effect in both cases connected to other structural factors in these languages. For Italian, /sp/ clusters are traditionally described as heterosyllabic ('*impure-s*') and there are a number of morpho-phonological dimensions along which /sp/ clusters have an exceptional status in this language. In Slovak, the systematic presence of lexical syllabic consonants (e.g. *smrk*) may require further coordination patterns for complex onsets apart from predicted patterns introduced above.

A number of studies have investigated stop+sonorant clusters, but here a more complex picture emerges. For German, a C-center effect was confirmed for /kn/ (Peters & Kleber, 2014) and /gl/ (Brunner et al., 2014), but not for /pl/ (Brunner et al., 2014; Pouplier, 2012). In other languages, however, stop+lateral onsets consistently showed the expected coordination pattern (American English: Browman & Goldstein, 1988, Honorof & Browman, 1995; Marin & Pouplier, 2010; Polish: Bruni, 2011; Italian: Hermes et al., 2013; Romanian: Marin & Pouplier, 2014). While numerous methodological differences between studies may contribute to these mixed results, in our present context it is important to note that, in all of these studies, clusters with vowel-adjacent labials (e.g. /sp/) overlapped gradually more with the vowel than clusters with a vowel-adjacent lateral (e.g. /pl/; cf. Goldstein et al., 2009; Marin & Pouplier, 2010; Pouplier, 2012). This is in line with our suggestion that the degree to which onset-vowel overlap may increase varies as a function of the coarticulation resistance of C_A, since labials are independently known to be less coarticulation resistant than alveolar sonorants (e.g. Recasens & Espinosa, 2009; Recasens, Pallars, & Fontdevila, 1997; see Section 2.1.2 for a detailed discussion of coarticulation resistance). In this view, also cross-linguistic differences in the nature of laterals (clear vs. dark [l]) could lead to variation in the degree of onset-vowel overlap, since dark laterals are by and large more resistant to coarticulation than clear laterals (see Section 2.1.2). The relatively strongest evidence so far for the role of coarticulation resistance stems from Marin's (2013) work on Romanian which showed no change in overlap for clusters in which the vowel-adjacent consonant is, among others, a sibilant (Marin, 2013). Sibilants are among the highest ranking segments the coarticulation resistance scale. In Marin's work the clusters failing to increase overlap were /ps, ks, kn, and kt/. Therefore she concluded that this was due to C_R being a stop without going further into detail why a vowel-remote stop might block a C-center effect. She also considered frequency as a conditioning factor, a point we will take up again in the Discussion.

In sum, previous results concerning consonant cluster timing reveal that the organization of onset clusters cannot be described with a single timing pattern as originally proposed by Browman and Goldstein (1988, 2000). Instead, the onset-vowel timing seems to vary gradually with cluster composition. To the extent that previous results are comparable across studies, there is some evidence for coarticulation resistance of C_A being one of the conditioning factors. The concept of coarticulation resistance and why exactly it is expected to interact with onset-vowel timing will be discussed in detail in the following Section 2.1.2.

2.1.2 Coarticulation resistance

The term 'coarticulation' was originally introduced by Menzerath and de Lacerda (1933) to describe articulatory interaction of neighboring speech sounds (see Kühnert and Nolan (1999) for an outline of historical coarticulation research). More specifically, coarticulation results from temporally and spatially overlapping gestures (e.g. Saltzman & Munhall, 1989; Öhman, 1966). Numerous studies have revealed that the degree of coarticulation varies with the particular consonants and vowels involved, a phenomenon known as coarticulation resistance (e.g. Bladon & Al-Bamerni, 1976; Farnetani, 1990; Fowler & Brancazio, 2000; Recasens & Espinosa, 2009; Recasens et al., 1997; Öhman, 1966). Coarticulation resistance has often been evaluated based on VCV sequences with varying consonants and vowel contexts. The articulation of intervocalic labial consonants has been shown to vary substantially as a function of vowel context in that the tongue body positioning during a labial is largely determined by the surrounding vowels (low degree of coarticulation resistance). Sibilants, on the other hand, have been found to retain their tongue shape robustly against the coarticulatory force of adjacent vowels (high degree of coarticulation resistance). Therefore, for instance, all else being equal, there will be more vowel-conditioned variability during the intervocalic consonant between /ipi/ vs. /apa/ than between /isi/ vs. /asa/ (cf. Recasens & Espinosa, 2009; Recasens et al., 1997).

In a series of studies Recasens and colleagues came to the conclusion that specifically the degree of active tongue body control during the production of a given consonant conditions its coarticulatory resistance to the influence of adjacent segments (e.g. Recasens & Espinosa, 2009). This means the degree of vowel-induced variability increases when the consonantal demand on the tongue body (or the articulatory constraint) decreases. At the same time this also means that the manipulative strength of a consonant or coarticulatory aggression on neighboring consonants and vowels increases with the consonant's inherent degree of articulatory constraint (Farnetani, 1990; Fowler & Saltzman, 1993).

Given that coarticulation results from the interaction of temporally overlapping gestures and that complex onsets show a comparatively high degree of C_AV overlap, coarticulation resistance of the vowel-adjacent consonant can be expected to interact with onset-vowel timing. While the concept of coarticulation resistance is part of the gestural and associated task dynamic model in terms of dominance coefficients for articulators (Fowler & Saltzman, 1993; Iskarous et al., 2012; Saltzman & Munhall, 1989), there has been virtually no work so far that has looked at the interaction of temporal coordination and spatial dominance factors. Brunner et al. (2014) did not address coarticulation resistance specifically, but they related the temporal C-center measure to coarticulatory effects. For instance, they observed that due to the influence of a C_R=/k/ the C-V movement path of the tongue in /k<u>ve</u>/ is shorter compared to /<u>ve</u>/. Due to the positive correlation of this measure and their temporal C-center measure, Brunner et al. suggested that the temporal lag measure is informative about relative distance of constriction location between C_R and vowel in a

C_RC_AV cluster rather than underlying syllable organization. Yet their study did not target this hypothesis specifically and it is difficult to glean conclusive results from their paper.[4] In the current study, we capitalize on the range of phonotactic variation permitted in Polish onset clusters in order to provide a systematic study of how coarticulation resistance of C_A interacts with onset-vowel timing. We use onset clusters with vowel-adjacent consonants known to differ in their coarticulation resistance and predict a negative correlation between the degree of C_AV overlap change when onset complexity increases and the degree of coarticulation resistance of C_A.

For our current study, we define for Polish consonants three discrete categories of coarticulation resistance which are primarily based on the presumed degree of tongue body control of the respective consonants (e.g. Recasens & Espinosa, 2009). Our work thereby faces the problem that there is, to our knowledge, no independently established coarticulation resistance hierarchy for Polish. We thus confine ourselves to a relatively coarse-grained hierarchy using sibilants, labials, and alveolar sonorants. Despite possible language-specific effects in coarticulation resistance, it can generally be assumed that sibilants are cross-linguistically among the most highly coarticulation resistant consonants, while labials are least coarticulation resistant. For alveolar stops, laterals and nasals, the situation is more complex. In particular the coarticulation resistance of alveolar stops relative to alveolar sonorants is unclear (Geumann, Kroos, & Tillmann, 1999; Hoole, Gfroerer, & Tillmann, 1990; but see Iskarous et al., 2013). In the absence of independent work on the patterning of these stops in Polish, we decided to include /n, l/ in an alveolar sonorant category for which we can relatively safely expect to range between sibilants and labials on the coarticulation resistance scale (see Recasens & Espinosa, 2005 for the similarity of clear /l/ to /n/ in terms of coarticulation resistance). In sum, the consonants used in this work are categorized as outlined in the coarticulation resistance hierarchy (1):

[4]Also note that other C-center studies have contrasted $VC_R\#C_AV$ with $V\#C_RC_AV$, thus controlling for distance changes between conditions (e.g. Marin, 2013; Marin & Pouplier, 2010; Pouplier, 2012).

(1) **(post-)alveolar sibilants > alveolar laterals/nasals > labial consonants**[5]

We predict that the characteristic increase in onset-vowel overlap will vary according to this hierarchy: there should be more C_AV overlap in a cluster compared to a corresponding singleton condition if the vowel-adjacent consonant is prone to coarticulation (labial consonants), but no change in C_AV overlap when a highly resistant consonant is adjacent to the vowel (sibilants) and an intermediate change in C_AV overlap when C is an alveolar sonorant. We evaluate C_AV overlap changes both in terms of a temporal lag analysis as done in the previous studies outlined above, as well as in terms of tongue body position measurements.

Parallel to our temporal hypotheses, we expect different degrees of consonant tongue body position changes in $\underline{\text{C}}$V vs. C$\underline{\text{C}}$V as a function of the coarticulation resistance of the vowel-adjacent consonant. For instance, a less coarticulation resistant vowel-adjacent consonant should have a different tongue body position in the singleton compared to the cluster condition due to its increased overlap with the vowel (e.g. [$\underline{\text{m}}$ata] vs. [ʃ$\underline{\text{m}}$ata]). However, C_R-C_A coarticulation prevents us from isolating the effect of increased vowel overlap in the vowel-adjacent consonant across clusters: When increasing onset complexity, the influence exerted by C_R on tongue body position during C_A confounds the ability to directly show a stronger presence of the vowel in C$\underline{\text{C}}$V compared to $\underline{\text{C}}$V. We will therefore assess onset-vowel overlap indirectly by the degree to which C_R (the vowel-remote consonant) exerts influence on the vowel. If C_A blocks an increase in onset-vowel overlap we should see less influence of C_R during the vowel compared to cases in which C_A allows for increasing onset-vowel overlap.

Summing up, our current paper brings together the two strands of research on cluster-vowel organization and on the lingual coarticulation resistance of consonants by comparing C_AV overlap patterns in singleton and cluster conditions for different onset compositions. The

[5] In this context the symbol '>' means 'coarticulatory more resistant than' and 'coarticulatory more aggressive than' at the same time.

reason to expect that cluster-vowel organization should interact with consonantal coarticulation resistance arises from the observation that $\mathrm{C_AV}$ overlap increases with increasing onset complexity. We use Polish to investigate the articulatory mechanisms underlying cluster-vowel timing and to understand the interplay of cluster composition and syllable organization. Finally, we will briefly examine the effect of voice onset time (VOT) on cluster-vowel organization for onsets with vowel-adjacent voiceless stop consonants, since this has been claimed to be a possible confounding factor in the performed timing measures (cf. Brunner et al., 2014).

2.2 Method

2.2.1 Speakers and data collection

We acquired data from three female and three male native speakers of Polish. Prior to the experiment subjects were familiarized with the recording setup, the reading task, and a list of utterances. Kinematic data were recorded by means of electromagnetic articulography (EMA; AG501, Carstens Medizinelektronik). We attached four sensors to the tongue, approximately equally spaced to the tongue tip (TT), anterior tongue mid (TM1), posterior tongue mid (TM2), and to the dorsal region of the tongue body (TB). Further sensors were placed on the upper and lower lips (UL, LL), and on the lower incisor to capture jaw movement (JAW). Reference sensors (nose ridge, upper incisor and on the mastoid process behind the ears) were used for the correction of head movements. Kinematic signals were sampled at 250 Hz. At the same time we recorded acoustic data at a sampling rate of 25 600 Hz. Articulatory and acoustic data were post-processed using an algorithm toolbox developed by Carstens Medizinelektronik and Philip Hoole. While the kinematic signals of the references sensors were filtered at 5 Hz, we chose a filter frequency of 40 Hz for TT and 20 Hz for the remaining sensors (TM1, TM2, TB, UL, LL, and JAW). The first derivative of the signals was filtered at 24 Hz. Finally, the data were corrected for head movement and rotated to each subject's occlusion plane.

2.2.2 Corpus setup

We used sets of target words with singleton and cluster onsets which allow for the direct comparison of onset-vowel timing over increasing onset complexity, e.g. set **ʃm-** consists of target words [ʃmata] and [mata]. Due to lexical constraints, we consistently employed disyllabic target words. Since in Polish the primary word stress falls predominantly on the penultimate nucleus (Gussmann, 2007), all syllables of interest bore word-level stress. Within each set we kept the phonemic environment as consistent as possible to preserve the comparability between singleton and cluster target words. However, we had to include some non-words or proper names in cases where no appropriate lexical word exists. In these few cases we included phonologically well-formed items as similar as possible to lexical words of Polish (e.g. **pnaci* [pnat͡ʂ̧i] similar to *pnący* [pnɔ̃nt͡ʂ̧ɨ] '*climbing (plant)*'). For this study we assembled 14 sets of singleton and cluster target words. Table 2.1 gives the stimuli grouped by coarticulation resistance of the vowel-adjacent consonant. The choice of onset clusters and the apparent variability of the vowel-remote consonant (C_R) is conditioned by the intended measurements (see Section 2.2.3 below) since for homorganic clusters it is hardly possible to identify two separate consonantal gestures in articulography data.

sibilants			*alveolar sonorants*			*labials*		
SET	**singleton**	**cluster**	**SET**	**singleton**	**cluster**	**SET**	**singleton**	**cluster**
ps-	[sɔtʲɲa]	[psɔtʲɲe]	**ml-**	[lekaʃ]	[mlɛkax]	**sp-**	[pɔdʲɲɛt]	[spɔdʲɲɛ]
pʃ-	[ʃerek]	[pʃɛraʑ]	**vl-**	[lit͡ʃi]	[vlit͡ʃɨ]	**ʃp-**	[pɛrɔ̃n]	[ʃpɛrat͡ɕ]
mʃ-	[ʃalik]	[mʃalik]	**pl-**	[latom]	[plat͡som]	**ʃm-**	[mata]	[ʃmata]
ks-	[zɛrɔ]	[ksɛrɔ]	**kl-**	[lud͡ʑik]	[klut͡ʃɨk]	**kp-**	[pʲinom]	[kpʲinom]
			pn-	[nat͡ʂ̧i]	[pnat͡ʂ̧i]			
			kn-	[nurek]	[knurɨ]			

Table 2.1: Sets of singleton and cluster onset target words. The stimuli are grouped according to the coarticulation resistance of the vowel-adjacent consonant.

During the recording session, speakers sat in a semi-anechoic room and produced four randomized blocks of accented target words embedded in a carrier phrase. The carrier phrases varied slightly across sets to avoid monotony throughout the experiment; however, we used the identical carrier phrases within each set. The target words were always flanked by low vowels, e.g. "*Kasia powiedziała <target word> automatycznie.*" ('*Kasia said <target word> automatically.*'). The targeted data set comprised four repetitions per target word and subject (14 sets × 2 complexity conditions × 4 repetitions × 6 subjects = 672 utterances). We had to exclude 36 utterances due to undetected misreadings and technical difficulties during the recording session, leaving 636 items for analysis.

2.2.3 Measurements

Labeling

The articulatory movement time series obtained for each articulator were labeled by means of the MATLAB-based *mview* algorithm developed by Mark Tiede at Haskins Laboratories. The consonantal gestures were identified in the movement time functions of the relevant articulator. Labial consonants (/m, p, v/) were defined on the basis of lip aperture (LA), i.e. the Euclidean distance between the sensors attached to the upper and the lower lips. Consonants having a primary constriction in the (post-)alveolar (/t, d, n, s, ʃ, l, r/) or velar region (/k/) were defined on the basis of the sensors attached to the tongue tip (TT) or the tongue body (TB), respectively. Once a consonant gesture was manually determined in a given articulator time series (Figure 2.2, solid line), *mview* detected automatically the maxima in the tangential velocity profile (Figure 2.2, dotted line) corresponding to the articulator movement towards (PVEL1) and away (PVEL2) from the constriction. These landmarks were chosen since they can most reliably be identified algorithmically across conditions.

Temporal organization measurements

Following previous work (e.g. Browman & Goldstein, 1988; Marin & Pouplier, 2010), we applied articulatory timing measurements as a diagnostic tool to indirectly determine C_AV

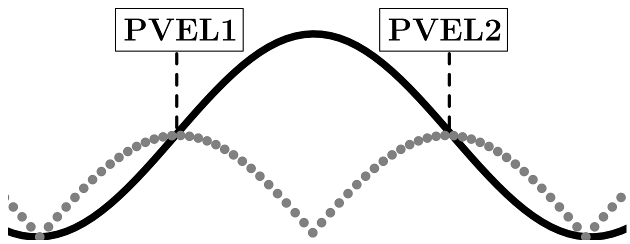

Figure 2.2: Time functions of tangential velocity (dotted line) and position (solid line) are schematically displayed. Articulatory landmarks are identified for the peak velocity of the constriction formation (PVEL1) and the release movement (PVEL2).

overlap as the distance between PVEL1 of the vowel-adjacent consonant and a constant anchor point for each singleton and cluster target word (cf. Figure 2.3). Here, the timepoint of PVEL1 of the consonant following the vowel was defined as the anchor point. That means C_AV overlap was defined as the temporal lag /m/ ↔ /t/ in [m̲at̠a] and [ʃm̲at̠a]. Decreasing temporal lag values thus indicate increased overlap.

We then computed lag ratios for each cluster in order to quantify the relative change in C_AV overlap between singleton and cluster condition. We averaged all lag measurements of a given singleton target word (e.g. onset: /m/ ↔ /t/ in [m̲at̠a]) and related then the lag value of each occurrence of the corresponding cluster target word (e.g. onset: /m/ ↔ /t/ in [ʃm̲at̠a]) to the singleton mean value. Finally, we centered the lag ratios to 0. Positive lag ratios represent a decrease in C_AV overlap with increasing onset complexity. Negative lag ratios indicate an increase in C_AV overlap with increasing complexity (cf. Figure 2.1). Lag ratios around 0 suggest no change in C_AV overlap in the cluster compared to the singleton condition.

Considering our hypotheses, we expect that onset clusters with a less coarticulation resistant vowel-adjacent labial stop should show more C_AV overlap in the cluster than in the singleton condition (i.e. negative lag ratios) while we predict no change in C_AV overlap for

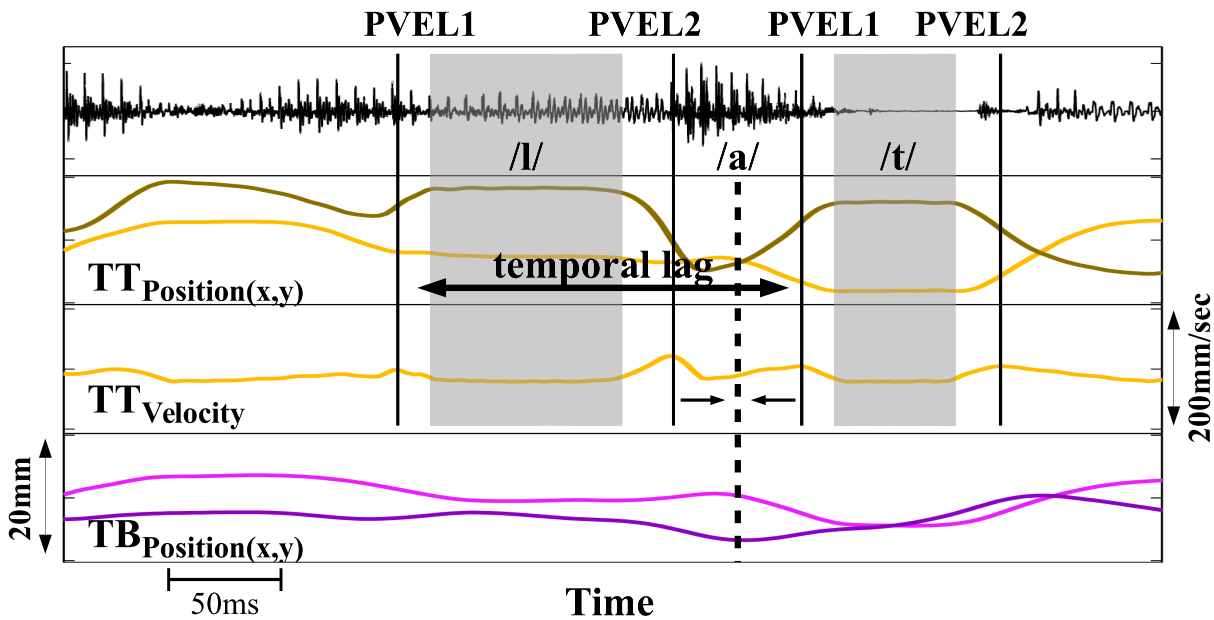

Figure 2.3: This plot illustrates the temporal lag measurements between, here, /l/ ↔ /t/ in a given utterance of target word [latom]. Movement and velocity profiles are given for tongue tip (TT) and tongue body (TB) in correspondence to the oscillogram in the upper panel. The vertical solid lines in the TT panels identify the time points of peak velocity of the particular closing (PVEL1) and opening (PVEL2) movements; the shaded boxes indicate the constriction plateaus for /l/ and /t/. The temporal lag is defined as the distance between PVEL1 of /l/ and /t/, respectively. The vertical dashed line indicates the measurement time point for the vowel which is defined as temporal midpoint of PVEL2 of /l/ and PVEL1 of /t/. At this point in time we extracted the tongue body (TB) position for the vowel variability analysis (see Section 2.2.3).

onset clusters with highly coarticulation resistant vowel-adjacent sibilants (i.e. lag ratios around 0). Vowel-adjacent alveolar nasals and laterals should show an intermediate degree of $C_A V$ overlap change (cf. hierarchy (1) in the Introduction).

One part of our analyses also considers the possible confounding role of VOT in our measurements, a point raised by Brunner et al. (2014). They argue that decreasing temporal lags in the case of vowel-adjacent stops may not reflect changes in $C_A V$ overlap but merely a shortened VOT between singleton and cluster conditions. We thus test for our labial condition which contains mostly stops as vowel-adjacent consonants whether the lag decrease is equivalent in size to the reduction of VOT duration between singleton and cluster. VOT was measured from the acoustic signal as the interval between the closure release of the stop and the beginning of periodicity for the following vowel. Then we subtracted for each token this VOT duration from the respective temporal lag gained from the articulatory analysis and recomputed the lag ratios without VOT.

Tongue body position measurements

As mentioned earlier, the temporal lag between the vowel-adjacent consonant and a constant anchor point is an indirect measure to indicate changes in C_AV overlap over increasing onset complexity. We complement the temporal measures by also looking at tongue body position changes which should be concomitant with increased overlap. Since it is difficult to reliably segment vowel movements in EMA data across different consonantal contexts, we defined the vowel as the temporal midpoint of PVEL2 of the vowel-adjacent consonant and PVEL1 of the anchor consonant in the tongue body (TB) trajectory. This time point matches the acoustic midpoint of the vowel fairly well (cf. dashed line in Figure 2.3). The TB sensor position should reflect the respective tongue position of the vowels.

To measure the C-to-V coarticulation in singleton and cluster condition, we adapted the Euclidean Distance Ratio (EDR) methodology from Harrington (2006). He used this method to quantify the relative proximity of a particular vowel to two other vowel categories. Analogously, for most of the cluster target words in our corpus we assigned two singleton target words to figure out the relative proximity of vowel tongue body position between matched cluster and singleton target words (e.g. [spɔ̠dʲɲɛ] compared to [pɔ̠dʲɲɛt] and [sɔ̠tʲɲa]). This analysis could only be performed for cases in which we had matching triplets of target words with similar segmental compositions (Table 2.2). This was the case for only parts of our corpus since the design of our study was primarily aimed at singleton-cluster target pairs.

sibilants			*alveolar sonorants*			*labials*		
$\mathbf{C_RV}$	$\mathbf{C_RC_AV}$	$\mathbf{C_AV}$	$\mathbf{C_RV}$	$\mathbf{C_RC_AV}$	$\mathbf{C_AV}$	$\mathbf{C_RV}$	$\mathbf{C_RC_AV}$	$\mathbf{C_AV}$
[pɔdʲɲɛt]	[psɔtʲɲe]	[sɔtʲɲa]	[mata]	[plat͡som]	[latom]	[sɔtʲɲa]	[spɔdʲɲɛ]	[pɔdʲɲɛt]
[pɛrɔ̃n]	[pʃɛraʑ]	[ʃerek]	[vit͡sɛ]	[vlit͡ʃɨ]	[lit͡ʃi]	[ʃerek]	[ʃpɛrat͡ɕ]	[pɛrɔ̃n]
[mata]	[mʃalik]	[ʃalik]	[mata]	[pnat͡ʂi]	[nat͡ʂi]	[ʃalik]	[ʃmata]	[mata]

Table 2.2: Triplets of cluster target word (C_RC_AV) and corresponding singleton target words (C_RV and C_AV). Triplets are grouped with respect to the coarticulation resistance of the vowel-adjacent consonant of the cluster condition. The onset consonant in C_RV target words corresponds to the vowel-remote consonant in C_RC_AV target words (e.g. [sɔtʲɲa] and [spɔdʲɲɛ]). The onset consonant in C_AV target words corresponds to the vowel-adjacent consonant in C_RC_AV target words (e.g. [pɔdʲɲɛt] and [spɔdʲɲɛ]).

For each subject and triplet of target words, we identified for both singleton target words (e.g. /ɔ/ in [pɔdʲɲɛt] and [sɔtʲɲa]) the respective centroid of the tongue body position during the vowel. Then we determined for each singleton and cluster token of the particular triplet the Euclidean distance to the centroid of C_AV (e.g. /ɔ/ in [pɔdʲɲɛt]) and the centroid to C_RV (i.e. /ɔ/ in [sɔtʲɲa]). We thus derived for each singleton and cluster token the Euclidean distances to the respective centroids, i.e. E1 and E2. Finally, we calculated the logarithmic Euclidean distance ratio (EDR = log(E1/E2)) to quantify the relative proximity of each token to these two singleton centroids (cf. Figure 2.4).

For the singleton conditions, positive EDR values correspond to the singleton target word beginning with the vowel-remote consonant of the cluster condition (C_RV); negative EDR values correspond to the singleton target word beginning with the vowel-adjacent consonant of the cluster condition (C_AV). To the extent that C_R in C_RC_AV cluster exerts influence on the vowel, EDR values are expected to fall between the EDR values of $C_R\underline{V}$ and $C_A\underline{V}$ (here: [sɔtʲɲa] and [pɔdʲɲɛt]). An EDR of 0 would indicate that a cluster vowel token (C_RC_AV) is equidistant to the singleton centroids. If a given pair of $\underline{C_AV}$ and $C_R\underline{C_AV}$ tokens has similar EDR values, this indicates that the vowel tongue body position did not change

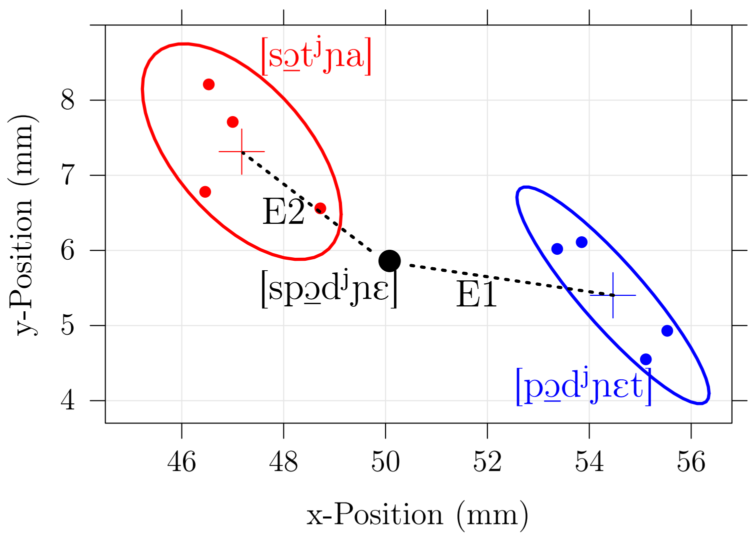

Figure 2.4: Illustration of the Euclidean Distance Ratio procedure by means of tongue body position data for speaker PL3. Ellipses represent the distribution of the respective singleton vowels with the determined centroids indicating 'prototypical' vocalic tongue body positions for, here, [sɔtʲɲa] and [pɔdʲɲɛt]. The distance measures E1 and E2 indicate the Euclidean distances of one particular cluster vowel (i.e. [spɔdʲɲɛ]) to the centroids. The Euclidean Distance Ratio (EDR) was then computed as log(E1/E2).

with increasing onset complexity, as to be expected if C_A blocks increasing overlap with the vowel. If, however, $\underline{C_A V}$ and $C_R\underline{C_A V}$ tokens have clearly diverging EDR values, this indicates the vowel-remote consonant (C_R) exerted coarticulatory influence on the vowel in $C_R C_A V$ target word.

We hypothesize that EDR values for the vowel in the cluster condition differ as a function of the coarticulation resistance of the vowel-adjacent consonant in parallel to the temporal lag measure: Since the less coarticulation resistant labials (group *labials*) allow for more $C_A V$ overlap in the cluster compared to the singleton condition, we expect EDR values of $C_R\underline{C_A V}$ tokens to differ from $\underline{C_A V}$ tokens, tending towards $C_R V$ tokens. For target words of group *sibilants* the tongue body position of the vowel is not assumed to change with increasing onset complexity (i.e. similar EDR values for $C_R\underline{C_A V}$ and $\underline{C_A V}$) since the vowel-adjacent sibilant is expected to prevent an increase of onset-vowel overlap. Target words of the group *alveolar sonorants* should consequently show an intermediate pattern due to the moderate resistance to coarticulation of vowel-adjacent alveolar sonorants.

2.2.4 Statistical analyses

For statistical analyses, we used the R environment (R Core Team, 2013) to carry out linear mixed models (*lme4* package: Bates, 2010; Bates, Mächler, Bolker, & Walker, 2015). P-values were obtained by comparing, for example, one model with and one without the fixed factor/without the interaction of the fixed factors. Unless stated otherwise, we excluded the potential random factor Repetition since it was not required by the particular model. Tukey post-hoc tests (*multcomp* package: Hothorn, Bretz, & Westfall, 2008) were carried out to perform pairwise comparisons.

2.3 Results

2.3.1 Temporal measurements on onset-vowel overlap

In this section we present the results for the relative change in C_AV overlap as a function of cluster group. Figure 2.5[6] gives the mean lag ratios for each group in the order of decreasing coarticulation resistance of the vowel-adjacent consonant from left to right. Negative lag ratios indicate a C_AV overlap increase with increasing onset complexity. Positive lag ratios suggest a C_AV overlap decrease. No change is indicated by lag ratios around 0.

The linear trend in Figure 2.5 confirms our hypothesis that the relative change of C_AV overlap between singleton and cluster condition is affected by the degree of coarticulation resistance of the vowel-adjacent consonant: for highly coarticulation resistant sibilants there is no change in C_AV overlap between cluster and singleton condition (i.e. ratios around 0) while *labials* group with low coarticulation resistance shows the greatest change in C_AV

[6] The results of the by-set analysis are given in the Appendix to allow for comparisons to previous studies (Figure A.1).

overlap between singleton and cluster conditions; vowel-adjacent alveolar nasals and laterals (group *alveolar sonorants*) show as predicted an intermediate pattern of relative C_AV overlap. We carried out a mixed model analysis to test whether the dependent variable Lag Ratio differed between Groups (three levels: *sibilants*, *alveolar sonorants*, *labials*); Speaker and Set served as random factors. The relative change of C_AV overlap between singleton and cluster onsets differed significantly as a function of Group ($\chi^2[2]=10.0$, $p<0.01$). This corroborates our hypothesis that the coarticulation resistance of the vowel-adjacent consonant interacts with cluster-vowel timing. Post-hoc Tukey tests revealed a significant differences between *sibilants* and *labials* ($p<0.01$) and *alveolar sonorants* and *labials* ($p<0.05$). However, the comparison between groups *sibilants* and *alveolar sonorants* was not significant.

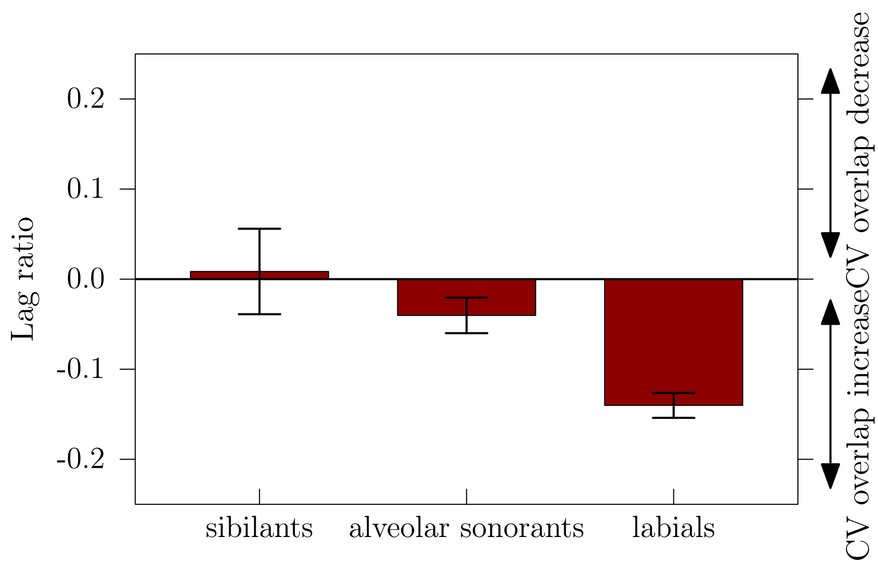

Figure 2.5: Lag ratios for onset clusters with sibilants, alveolar laterals/nasals and labial stop consonants in vowel-adjacent position indicating the relative change of C_AV overlap. Error bars: ± 1.00 SD.

In addition to the between-group differences concerning the relative C_AV overlap change, we ran a second mixed model to evaluate whether C_AV overlap changed significantly between singleton and cluster condition within each group. This analysis is informative as to whether we see the expected C-center effect within each group. We tested the interaction of the fixed factors Group (three levels: *sibilants*, *alveolar sonorants*, and *labials*) and Complexity (two levels: singleton and cluster) on the dependent variable Temporal Lags. As random factors we used Subject, Set, and Word. While the interaction turned out be significant ($\chi^2[2]=7.2$, $p<0.05$), the subsequent Post-hoc Tukey tests revealed only for group *labials* (i.e. sets **kp-**, **ʃm-**, **ʃp-**, and **sp-**) significantly more C_AV overlap in the cluster than in the corresponding singleton condition ($p<0.001$). For the remaining groups *sibilants* and *alveolar sonorants* we could not observe a significant increase in C_AV overlap with increasing onset complexity (*sibilants*: p=0.99; *alveolar sonorants*: $p=0.16$). In sum the C_AV overlap results confirm the hypothesized effect of coarticulation resistance of the vowel-adjacent consonant on onset-vowel overlap (cf. successive decline of lag ratios in Figure 2.5). It is particularly interesting that only labial consonants overlapped more with the vowel with increasing onset complexity, although Figure 2.5 shows that qualitatively the effect is in the predicted direction for *alveolar sonorants* also (lag ratio < 0).

The results suggest a clear quantitative difference in lag ratios between groups *labials* and *sibilants*/*alveolar sonorants* (cf. Figure 2.5). We now turn to the question whether this result could have been conditioned by other phonetic properties than consonantal coarticulation resistance, in particular aspiration differences between singleton and cluster stops might confound the lag measures (Brunner et al., 2014). We therefore subtracted each token's VOT duration from its temporal lag for all clusters but **ʃm-** in the group *labials* (**kp-**, **sp-**, and **ʃp-**; see Method). The comparison of the two available data sets of temporal lag measurements allows us to investigate how VOT changes of the vowel-adjacent stops in singleton and cluster condition affects the onset-vowel timing. If the VOT difference between singletons and clusters is the main factor that conditioned our results, then we would expect to see clearly divergent lag ratios for the genuine (+VOT) and the simulated (-VOT) data, i.e. the lag ratios for the simulated sets should be around 0 (cf. groups *sibilant* and *alveolar sonorants* in Figure 2.5).

Table 2.3 contrasts the lag ratios determined for the genuine data set (+VOT) and the simulated data set (-VOT). The overview confirms that the relative C_AV overlap increase is slightly diminished for all sets when VOT duration was excluded from the temporal lags, i.e. mean lag ratios are consistently closer to 0 in -VOT than +VOT columns. The mixed

	Lag ratios	
	+VOT	**-VOT**
kp-	-0.17	-0.14
sp-	-0.10	-0.06
ʃp-	-0.15	-0.14
mean	-0.14	-0.11

Table 2.3: Comparison of the lag ratios for the genuine (+VOT) and the simulated (-VOT) data of **kp-**, **sp-**, and **ʃp-**.

model carried out similarly to the previous analyses revealed an interaction of Group (three levels: *sibilants*, *alveolar sonorants*, and *labials*) and Complexity (two levels: singleton and cluster) only at trend level ($\chi^2[2]=4.7$, p=0.095), but both fixed factors were significant (Group: $\chi^2[2]=11.9$, $p<0.01$; Complexity: $\chi^2[1]=7.1$, $p<0.01$). Post-hoc Tukey tests showed a significantly increasing C_AV overlap with increasing onset complexity only for Group *labials* ($p<0.05$), but not for Groups *sibilants* and *alveolar sonorants*. Evaluating the question whether this reduced relative C_AV overlap change has affected the inter-group differences, we used again the Lag Ratios as the dependent variable, Groups (three levels: *sibilants*, *alveolar sonorants* and *labials*) as fixed factor and Speaker and Set as random factors. This mixed model revealed a significant effect of Group on Lag Ratios ($\chi^2[2]=7.6$, $p<0.05$), but the Post-hoc Tukey confirmed only between the groups *labials* and *sibilants* a significant difference ($p<0.05$). In sum, the analysis of the simulated sets with vowel-adjacent labials confirms that VOT difference between singletons and clusters is not the driving factor for our temporal results on C_AV overlap.

2.3.2 Coarticulatory measurements on onset-vowel overlap

Up to this point we determined C_AV overlap indirectly by comparing onset-vowel timing under increasing onset complexity. We will now present the results of the analysis on tongue body position data as outlined in Section 2.2.3. In Figure 2.6 we present for each group of onset triplets the Euclidean Distance Ratios (EDR) by cluster group. For the singleton conditions, positive EDR values correspond to the singleton target word beginning with the vowel-remote consonant of the cluster condition (C_RV); negative EDR values correspond to the singleton target word beginning with the vowel-adjacent consonant of the cluster condition (C_AV). For the cluster conditions (C_RC_AV), an EDR of 0 would indicate that the vowel the in cluster condition is equidistant to the singleton centroids. In this analysis the comparison of $C_R\underline{C_AV}$ and $\underline{C_AV}$ is particularly meaningful since it demonstrates how vowel tongue body position varies with increasing onset complexity. If $C_R\underline{C_AV}$ and $\underline{C_AV}$ target words (i.e. target words with the same vowel-adjacent consonant) have similar EDR values, this indicates that the position of the tongue body during the vowel did not change with increasing onset complexity. If, however, $C_R\underline{C_AV}$ and $\underline{C_AV}$ target words have clearly diverging EDR values, then the vowel is coarticulated by the vowel-remote consonant.

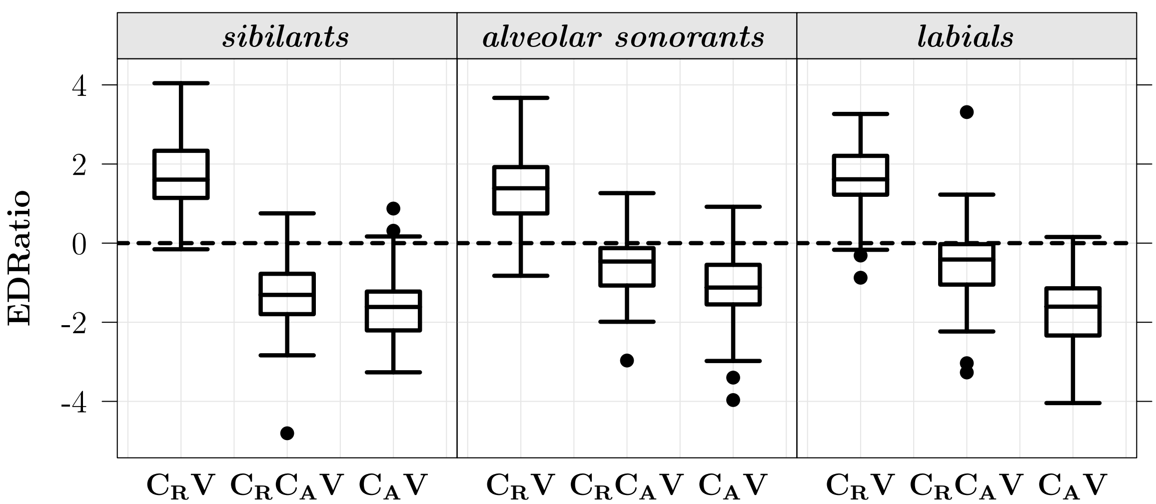

Figure 2.6: Logarithmic Euclidean Distance Ratios by cluster group (C_A = *sibilants*, *alveolar sonorants*, and *labials*) and target word type (C_RV, C_RC_AV, and C_AV).

Considering first the C_RC_AV conditions in Figure 2.6, a difference between *alveolar sonorants/labials* and *sibilants* is evident: while the cluster condition of group *sibilants* has clearly negative EDR values (*sibilants*: mean=-1.24), the cluster conditions of the remaining two groups show EDR values closer to 0 (*alveolar sonorants*: mean=-0.51; *labials*: mean=-0.57). This implies that the tongue body position changes during the vowel of the cluster tokens relative to both of the singleton centroids is roughly the same for groups *alveolar sonorants* and *labials*. A mixed model tested whether the relative tongue body position changes (i.e. EDR values of C_RC_AV tokens) varied as a function of cluster group (i.e. Group; three levels: *sibilants*, *alveolar sonorants*, *labials*); Speaker and Triplet served as random factors. Results revealed a significant effect of Group on the relative change of vowel tongue body position ($\chi^2[2]=11.397$, $p<0.01$). A post-hoc Tukey test further confirmed a significant difference between *sibilants* and *labials* ($p<0.01$) and *sibilants* and *alveolar sonorants* ($p<0.01$); *alveolar sonorants* and *labials* did not differ significantly from each other. This suggests that as coarticulation resistance of the vowel-adjacent consonant decreases, the coarticulatory influence of the vowel-remote consonant on the vowel increases (i.e. C_R-to-V coarticulation: *sibilants* < *alveolar sonorants*, *labials*).

Since this comparison does not actually quantify the degree of tongue body position change within each group, we now look at differences between singleton and cluster conditions by cluster group focusing only on tokens with the same vowel-adjacent consonant (i.e. C_AV vs. C_RC_AV). By visual inspection, for group *sibilants* the corresponding boxes are fairly similar (Figure 2.6, left), i.e. the tongue body position of the vowel did not change with increasing onset complexity. This conforms with our hypothesis for *sibilants* since vowel-adjacent sibilants apparently prevent increasing onset-vowel overlap. Group *labials* on the other hand shows clearly different EDR values for C_AV and C_RC_AV (Figure 2.6, right) which indicates that the C_R sibilant in the C_RC_AV condition exerts influence on the vowel (i.e. C_R-to-V coarticulation). This again corroborates our hypothesis that less coarticulation resistant labials allow for onset-vowel overlap increase from singleton to cluster condition. Finally, for group *alveolar sonorants* the EDR values slightly differ between C_AV and C_RC_AV tokens. We conducted a mixed model to test for an interaction of Group (three levels: *sibilants*, *alveolar sonorants*, *labials*) and Onset Complexity (two levels: C_AV, C_RC_AV) on the dependent variable EDR with Speaker, Set, and Word as random factors. The interaction of the fixed factors was significant ($\chi^2[2]=10.716$, $p<0.01$). A post-hoc Tukey test revealed a significant difference between cluster and singleton condition for the groups *alveolar sonorants* (p<0.01) and *labials* (p<0.001), but not for *sibilants* (p=0.13). The statistical analysis thus confirms that the vowel coarticulated with vowel-remote consonants (i.e. C_R-to-V coarticulation) in onset target words with a vowel-adjacent /n, l, m, p/, but not in those with vowel-adjacent sibilants.

2.4 Summary and Discussion

Since previous research suggests that models of articulatory correlates of syllable structure do not sufficiently take the segmental composition of the syllable into account, we systematically investigated in this study to what extent cluster-vowel timing depends on cluster composition and particularly on the coarticulation resistance of the vowel-adjacent consonant. We applied the articulatory timing measures previously used in C-center studies by comparing onset-vowel timing in singleton and cluster onsets (i.e. $\underline{C_A V}$ vs. $C\underline{C_A V}$). We further analyzed the articulatory variability of vowels with respect to onset complexity and used the degree of contextual vowel variability conditioned by C_R as an index of onset-vowel overlap change. For both the temporal and coarticulatory measurements we expected an effect of the coarticulation resistance of the vowel-adjacent consonant on onset-vowel overlap, i.e. relatively more $C_A V$ overlap change for group *labials* compared to *sibilants* and an intermediate pattern for group *alveolar sonorants*. We will summarize first the results presented in this paper before we turn to the discussion of the how to account for these findings within the gestural model of syllable structure.

The analysis of the temporal lag ratios revealed significant inter-group differences as indicated in Figure 2.5 in terms of a successive increase of the relative change of $C_A V$ overlap with decreasing coarticulation resistance of the vowel-adjacent consonant. More precisely, we found no change in $C_A V$ overlap for vowel-adjacent sibilant onsets (i.e. lag ratios close to 0 for group *sibilants*) but the largest increase in $C_A V$ overlap for onsets with vowel-adjacent labials (i.e. clearly negative lag ratios for group *labials*). In addition, we observed an intermediate degree of $C_A V$ overlap change for vowel-adjacent alveolar sonorant onsets. The quantitative difference between vowel-adjacent alveolar sonorant and labial onsets is in accordance with recent data for American English and German (cf. Goldstein et al., 2009; Marin & Pouplier, 2010; Pouplier, 2012). Overall, the outcome of our study confirms our hypothesis that onset-vowel timing interacts with the coarticulation resistance of the vowel-adjacent consonant.

We further investigated the role of VOT as a possible confounding factor (cf. Brunner et al., 2014). Subtracting VOT values from the temporal lag measures attenuated the effect but did not change the overall pattern of our results. The simulated lag ratios of group *labials* revealed a significant $C_A V$ overlap increase from singleton to cluster condition even when VOT duration was factored out. This provides clear evidence against Brunner et al.'s claim that VOT is one of the main factors conditioning C-center type results. In addition, our findings confirm previous results for Romanian: Marin (2013) found a relative increase of $C_A V$ overlap for /sp, sk/ compared to /p, k/ onsets even though Romanian stops are generally unaspirated.

Complementary to the temporal lag measurements we applied a second analysis that gauged onset-vowel overlap from tongue body position changes during the vowel. These results are in close agreement with the temporal lag results in that the relative change in tongue body position of a given vowel (conditioned by C_R) varied as a function of coarticulation resistance of C_A. In target words in which a coarticulatorily aggressive sibilant was adjacent to the vowel, tongue body position during the vowel was similar in singleton ($C_A V$) and cluster ($C_R C_A V$) condition, there was no influence of C_R. We found for group *alveolar sonorants* a statistical similar pattern as for group *labials*. While the pattern for *labials* is as expected, *alveolar sonorants* did not quite pattern as predicted (i.e. vowel-adjacent /l, n/ allowed for a relative onset-vowel overlap increase), even if Figure 2.6 suggests a slightly smaller tongue body position change compared to group *labials*.

In sum, the temporal and position measurements applied in this study revealed parallel results and substantiate the hypothesized interaction of coarticulation resistance of the vowel-adjacent consonant and temporal onset-vowel organization. More precisely, we found in both measurements that the degree of lingual coarticulation resistance of the vowel-adjacent consonant conditions the degree to which onset-vowel overlap changes under increasing onset complexity: onset-vowel overlap gradually increases as coarticulation resistance of the vowel-adjacent consonant decreases. However, with respect to the clusters with vowel-adjacent alveolar sonorants, there is some inconsistency across analyses: in the temporal measurements group *alveolar sonorants* patterned statistically with group *sibilants*, while in the coarticulatory measurements group *alveolar sonorants* patterned with group *labials*. To some extent, such a variable behavior can arguably be expected from an 'intermediate' category. Moreover, the second (coarticulatory) analysis was carried out on a subset of clusters used in the first (temporal) analysis, i.e. the difference in patterning may simply be conditioned by differences in statistical power. An alternative explanation concerns inherent limitations of the coarticulation resistance categories proposed in hierarchy (1). Polish consonants have thus far not been investigated in terms of coarticulation resistance, therefore the hierarchy applied here relies primarily on an external coarticulation resistance scale derived from consonants of other languages than Polish. That we adopted this scale for our study on Polish consonants might constitute a limitation of our results since the degree of a consonant's coarticulation resistance is conceivable to vary cross-linguistically and the underlying qualitative resistance measure (i.e. degree of articulatory constraint (DAC); e.g. Recasens & Espinosa, 2009; Recasens et al., 1997) might be too rough for an adequate distinction between the alveolar nasals, laterals, and sibilants. At the same time, as laid out in the Introduction, we believe that our choice of consonants was such that we can relatively safely assume that we have sampled three different levels along the coarticulation resistance scale. Establishing an independent, language-specific coarticulation resistance scale for Polish was outside the scope of this experiment but will nonetheless remain an important issue for future research.

In the following, we will discuss how the results of the temporal and coarticulatory measurements on onset-vowel overlap may be integrated into the gestural model of syllable structure which motivated our study. We will also consider whether a syllable model built on the jaw cycle (Lindblom, 2011; MacNeilage & Davis, 2000; Redford, 1999b) may add further insights to the interpretation of the results. Finally, the effects of frequency on coarticulatory patterns is taken into consideration (Marin, 2013).

Coarticulation is generally seen to be the result of spatially and temporally overlapping gestures and hence the gestural model is in principle ideally suited to model these effects. Fowler and Saltzman (1993) considered two possible scenarios for coarticulatory interactions of overlapping gestures, depending on whether the overlapping gestures share the same articulators or not. For example, since the tongue is not a shared articulator between vowel and consonant in /apa/, there will be maximal consonant-vowel coarticulation. This contrasts with cases like /asa/ in which vowel and consonant impose conflicting demands on the same articulator (here: tongue body) with the results of gestural blending. And indeed, blending coefficients or articulator weights are an integral part of the task dynamic model (Fowler & Saltzman, 1993; Saltzman & Munhall, 1989), they specify the degree to which a given gesture dominates the vocal tract, i.e. coarticulation resistance and aggression. For instance, the lingual coarticulation resistance of a sibilant is implemented by a relatively greater weighting of the consonantal compared to the vocalic tongue body gesture, limiting vowel-induced variability of the sibilant (i.e. a low degree of V-to-C coarticulation). In the case of coarticulatorily least resistant labials, the blending parameter is irrelevant since the labial and the vowel do not share the same (lingual) articulator, hence there is, in terms of the tongue, no constraint on V-to-C coarticulation.

The task dynamic application (TADA; Nam et al., 2004) is a computational implementation of task dynamics and the gestural coupling model (Browman & Goldstein, 2000; Goldstein et al., 2009; Nam et al., 2009; Saltzman & Munhall, 1989) in which gestural parameters (e.g. articulator target and weights) as well as intergestural coordination parameters are freely configurable. Using TADA, recent studies (Iskarous et al., 2012; Pastätter & Pouplier, 2014) demonstrated the effectiveness of the blending parameter to model different coarticulation effects in C_AV sequences: by means of a systematic alteration of the blending parameter these studies successfully modeled for a variety of consonants the particular vowel-induced variability patterns as observed in real articulatory data. Although the task dynamic mechanism of blending is in principle capable of capturing the spatial effects of V-to-C coarticulation in C_RC_AV sequences observed in our study, it cannot predict the temporal effects of coarticulation resistance of the vowel-adjacent consonant on cluster-vowel organization. When synthesizing, for instance, /ps/ and /sp/ clusters in TADA, the C-center effect invariably emerges, despite sibilants being modeled with a greater dominance over tongue body control compared to the vocalic gesture. The model output is thus in accordance with the gestural syllable model which maintains that onsets are, independently of their segmental composition, specified for the same coupling relations (see Introduction). That is, while the blending parameter influences the spatial configuration of the vocal tract, it has no influence on the degree of temporal overlap. This, however, is precisely the interaction uncovered in our study.

We propose in the following that coupling or bonding strength (Goldstein & Fowler, 2003) is a parameter that allows us to model the present results in a principled fashion. Interestingly, the concept of coupling strength was first introduced in the context of competing phase relationships which are the basis for the C-center effect (Browman & Goldstein, 2000). Coupling strength is a modeling device to give different blending weights to competing phase relationships, parallel to the spatial dominance parameter (articulator weights) used to implement coarticulation resistance. Browman and Goldstein (2000) originally argued that the C-center effect required a greater coupling strength for C-C anti-phase coupling relative to each consonant's C-V in-phase coupling (more on this below). This, however, has not been pursued in subsequent publications nor has it been implemented this way in TADA. The authors further proposed that different coupling strength settings between gestures would allow for continuous variation in gestural overlap due to prosodic factors and speaking style, but generally the function of this parameter has not received any deeper elaboration in the gestural model and hence at present lacks predictive power. By default in TADA all coupling strength settings are equal. We argue here that coupling strength can be used as a means to model the interaction of coarticulation resistance with different degrees of onset-vowel overlap. Specifically, we assume a correlation between the degree of coarticulation resistance of C_A and the degree of coupling strength.

Recall that in the gestural model cohesion among gestures is achieved via pair-wise coupling relations. As laid out in the Introduction, $\#C_RC_AV$ onsets are modeled in terms of three competing phase relations: C_R-V, C_A-V, both specified as in-phase, and C_R-C_A, specified to be anti-phase (e.g. Browman & Goldstein, 2000). While each of these coupling relations is associated with a coupling strength value, all coupling strength values are assumed to be equal in the current implementation, meaning the competing phase relations are blended with equal weights. Similar to the spatial dominance values which specify the degree to which a given gesture's parameters dominate articulator position / vocal tract shape, a lower or higher degree of coupling strength determines the degree to which the surface timing of the articulators reflects the underlyingly specified phase relation associated with a given coupling relation. Thereby a higher coupling strength value means greater weight or dominance is given to that coupling strength value in the case of blending. For instance in Goldstein et al. (2009) coupling strength is manipulated in order to model between speaker variability in terms of how closely individual speakers' onset cluster productions match the C-center pattern manifest in the cross-population mean. Thereby the focus is on modeling variation in the behavior of either C_R or C_A in $\#C_RC_AV$ onsets. If both C-V coupling relations are equally strong (the default), both consonants will deviate to an equal degree from their respective singleton C_AV timing pattern. Decreasing one of the consonant's coupling strength parameter gives greater dominance of the C-V (in-phase) coupling of the *other* consonant. For instance, when decreasing the coupling strength of C_R-V, C_A-V will remain closer to its singleton timing pattern (less shift), since the competing phase relationships will be resolved at the relative greater expense of C_R-V.

We propose here that each gesture should be associated generally with a characteristic coupling strength value representing its degree of coarticulation resistance. This is reminiscent of the numeric DAC scale of coarticulation resistance introduced by Recasens and colleagues (e.g. Recasens & Espinosa, 2009; Recasens et al., 1997). A ramification of this is that also singleton C-V coupling relations will be associated with different coupling strength values depending on the inherent value contributed by each specific C and V. But as with the spatial dominance parameters, this will only come into play if there are mutually incompatible phase relationship as is the case for the C-center model. For singletons, in-phase is the only specified phase relationship for C-V. For example, if we arbitrarily assign a value of 0.8 to coarticulatory resistant /s/, but a value of 0.2 for less resistant /p/, the in-phase coupling relationship is implemented in any case for syllables /sɔ/ and /pɔ/, since it is the only phase relationship specified. This does, however, make the prediction that C_AV timing in /sɔ/ should be less susceptible to variation than the C_AV timing in /pɔ/ if competing demands should arise, as might be conceivable through higher-level prosodic grouping (e.g. Smith, 2004). Yet prosodic grouping beyond the syllable has hardly ever been addressed in the gestural model, therefore elaborating and testing this prediction will have to remain a topic for future research. In order to test the idea that coupling strength can be used to model the interaction between coarticulation resistance and onset-vowel organization evident for clusters, we manipulated the coupling strength parameter in TADA and tried to replicate our results for /psɔ/ and /spɔ/.

In TADA two coupling strengths are specified for each pairwise coupling relation (e.g. C_A-V), one contributed by the consonant and one by the vowel. Again, by default all values are the same. In order to probe the hypothesized correlation of coarticulation resistance and coupling strength, we assigned an arbitrarily higher coupling strength value for coarticulatory resistant /s/ (α=0.8) compared to less coarticulatory resistant /p/ (α=0.2). The vowel contributed α=1.0 in all cases. First, C_AV singleton onsets (/pɔ, sɔ/) were modeled using these parameters, then the complex onsets /spɔ/ and /psɔ/ were generated. In the latter case, the C_R-C_A coupling relation was given the same value (α=1.0) for both /ps/ and /sp/, since the clusters combine the same consonants. These parameter settings ensure that in all singleton/cluster cases, there is a higher coupling strength associated with the sibilant-vowel coupling relation compared to the labial-vowel coupling relation. The model output shows first of all that, as predicted, timing in the singleton C-V output is not affected by differences in coupling strength settings between /pɔ/ and /sɔ/, i.e. both labial and sibilant start synchronously with the vowel. For /ps/ the vowel-adjacent sibilant starts near-synchronously with the vowel since the relative stronger coupling strength of the sibilant-vowel phasing causes a tight bonding of the vowel-adjacent sibilant and the vowel. For /sp/ however the relatively weaker coupling strength of the labial-vowel phasing allows for a relative increased overlap of the vowel-adjacent labial and the vowel. This indicates that the relative strength of C_R-V and C_A-V coupling relations is the determinant of how far C_A overlaps with the vowel in the cluster condition. Hence we suggest that the concept of coarticulation resistance in the gestural model should be expanded from a purely spatial dominance factor into a factor that affects both spatial and temporal relations among overlapping gestures (see also Iskarous et al. (2013) for a proposal of a more complex approach to the spatial dominance parameter). In sum, we demonstrated a way of how the results of the present study can be integrated into the gestural model of syllable organization in a principled fashion, without affecting the underlying coupling modes, i.e. in-phase (C_R-V, C_A-V) and anti-phase (C_R-C_A) coordination.

Our study was to a large degree motivated by Marin's (2013) observation that some clusters in Romanian did not show the C-center effect (see Introduction), some of which had a sibilant in vowel-adjacent position (/ps, ks/), but some of which did not (/kn/ and /kt/). Our proposal that coarticulation resistance is at the heart of that result would explain the patterning of /ps, ks/. Firstly, this highlights that a given segment's coarticulation resistance is not solely determined by the degree of dorsal control, but is also a function of stringent production requirements on other articulators such as the jaw. Again sibilants are a case in point since they allow for very little contextual variation in jaw position (e.g. Iskarous, Pouplier, Marin, & Harrington, 2010). This may be extended to /kt/, since it has been argued independently that voiceless coronal stops are for aerodynamic reasons quite restrictive in their requirement for a high jaw position (Mooshammer, Hoole, & Geumann, 2007). Therefore the patterning of /kt/ with the sibilants may again be due to the coarticulation resistance of the jaw of the vowel-adjacent consonant. This generally raises the point about the role of the jaw in syllable organization and hence for our current results. It is generally recognized that the jaw has a foundational basis for the syllable (e.g. MacNeilage & Davis, 2000; Redford, 1999b; Rochet-Capellan & Schwartz, 2007), even though it is also clear that it is not a sole determinant of syllable patterns (Nam, Goldstein, Giulivi, Levitt,

& Whalen, 2013; Redford & van Donkelaar, 2008). In the frame/content approach formulated by MacNeilage and colleagues (e.g. MacNeilage, 1998; MacNeilage & Davis, 2000), it is sometimes assumed that the segments of a syllable have to accommodate to a more or less constant jaw frame which serves as a basic timer in speech production (cf. Lindblom, 2011). In this view, increasing onset complexity therefore leads to decreased consonant duration in clusters compared to singleton consonant productions and/or increased consonantal overlap since more consonants have to be fitted into a more or less constant jaw frame. This would capture similarly to the C-center model the 'default' pattern of increased C-V overlap in singleton compared to cluster productions. However, this model has not been developed to account for differential C_AV (nor CC) overlap patterns depending on cluster composition.

Apart from the main focus of this study, we also consider lexical cluster frequency as a plausible confound in our data. Marin (2011, 2013) points out that the low frequency of Romanian clusters (e.g. /kn/, /ps/, and /ks/) may condition an unexpected syllable timing pattern at odds with the C-center model, but she also points out that in another study on German (Pouplier, 2012), other lexically marginal, very low frequency clusters displayed the expected C-center organization. In general, the role of frequency in speech production is widely recognized (among many others, Gahl, 2008; Jurafsky, Bell, Gregory, & Raymond, 2001; Levelt & Wheeldon, 1994). More frequent words tend to be reduced durationally and in terms of articulator position, and also fare better on accuracy, speed of production, and fluency metrics. Important in the current context is the fact that there is some evidence that frequency affects coarticulation specifically (Munson & Solomon, 2004; Pouplier, Marin, Hoole, & Kochetov, in press; Scarborough, 2004; Tomaschek, Wieling, Arnold, & Baayen, 2013; but see Wright, 1997, 2004) with higher frequency conditioning an increase in coarticulation (i.e. overlap). For our current study we assume that token frequency of each cluster was the relevant level to evaluate. We thus assembled freely available samples of the National Corpus of Polish (transcripts of governmental sessions; Przepiórkowski, 2013), the PELCRA corpus (spoken language; Pęzik, 2011), and a third corpus which includes subtitles of movies and TV series (Dave, 2012). Taken together, the total corpus size amounts 64 431 358 tokens (742 216 types). For each of our stimulus groups (*sibilants*, *alveolar sonorants*, and *labials*), we summed the token frequencies of the relevant clusters in absolute initial position and calculated the log probabilities as listed in Table 2.4.

Cluster Probability		
Rank	**C_A Group**	**Log(Prob)**
1	*sibilants*	-1.70
2	*labials*	-2.24
3	*alveolar sonorants*	-2.67

Table 2.4: Logarithmic probabilities of summed cluster frequencies by C_A group. The closer the logarithmic probability values are to 0, the higher the cluster frequency.

According to the cluster probabilities given in Table 2.4, onset clusters with vowel-adjacent sibilants (n = 1 274 032) occurred more frequently than clusters with vowel-adjacent labials (n = 368 866) and alveolar sonorants (n = 136 734). If frequency determines the degree of onset-vowel overlap, our results would predict the *sibilants* group to show the lowest frequency and the *labials* group to have the highest frequency. This clearly is not the case: against expectations, the *sibilants* group has the highest token frequency. However, this effect is largely carried by the cluster /pʃ/, which by itself counts 1 249 321 occurrences in our corpus. If we factor out this cluster, a probability ranking emerges that matches the results found in this study in that the higher frequency clusters are the ones which showed increasing onset-vowel overlap (log probability: *labials* (-2.24) < *alveolar sonorants* (-2.67) < *sibilants* (-3.42)). Yet it is unclear how to motivate the exclusion of /pʃ/. Also when taking a more detailed look at the frequency patterns, individual cluster frequencies are problematic for a pure frequency account. For instance, clusters /kp/ and /ʃp/ showed similar timing patterns (see Figure A.1), yet these clusters differed drastically in their frequency of occurrence in our corpus (n = 10 vs. n = 15 866, respectively). Also note that /pʃ/, which is by far the highest frequency cluster in the corpus has lag ratio values close to zero (no increase in overlap between singleton and cluster, see Figure A.1). Altogether, frequency may thus account for some of the qualitative differences observed between the groups *sibilants*, *alveolar sonorants*, and *labials* in Figures 2.5 and 2.6, but it is unlikely to be the major factor conditioning our results. To gain a deeper understanding of the role of frequency in articulatory timing, an experiment targeting frequency specifically will have to be done in future research.

To conclude, in this study we provided evidence that the temporal organization of onsets and vowel depends on segmental make-up of the consonantal onset and particularly the consonant adjacent to the vowel. In this respect the degree of tongue body control is a crucial parameter which determines the consonant's coarticulation resistance. The higher the degree of tongue body dominance of the vowel-adjacent consonant, the less is the relative $C_A V$ overlap increase. We argue that coarticulation resistance within the gestural model has to be expanded in terms of the coupling strength parameter in order to allow for coarticulation resistance to affect the temporal overlap of gestures in the case of competing phase relationships.

Chapter 3

Coarticulatory overlap and the jaw cycle: Evidence from Polish labial-coronal clusters

3.1 Introduction

In this chapter we continue to investigate the gestural timing of syllable constituents. In Chapter 2 we examined the effect of lingual coarticulation resistance of the vowel-adjacent consonant on onset-vowel organization and in particular on CV overlap. The results confirmed that the relative onset-vowel overlap increases from singleton to cluster onset condition as a function of the readiness of the vowel-adjacent consonant to coarticulate with the following vowel. This means that the relative change in CV overlap was larger when the vowel-adjacent consonant was a labial (e.g. /ʃm̲-/) than a coronal sonorant (e.g. /pn̲-, pl̲-/). In addition, there was overall no change in overlap when the vowel-adjacent consonant was a sibilant (e.g./mʃ̲-). As we pointed out in Section 2.4, the resistance of a segment to coarticulation concerns not only the tongue body but also the jaw. This raises the question as to whether the jaw plays a decisive role in determining the inter-gestural organization of adjacent syllable constituents, i.e. in the CCV sequence, the organization of C-V and C-C gestures. Lindblom (2011) suggested recently that the jaw (as a foundational basis of the syllable) could be the major determinant for gestural overlap and hence for coarticulation. To investigate further the articulatory basis of syllable organization, we take this assumption as a starting point for investigating the predictive power of jaw movements for the coarticulatory overlap of vowel-remote and vowel-adjacent consonant gestures (henceforth C_R and C_A, respectively) and how this in turn conditions incremental changes of C_A duration.

3.1.1 The jaw: subordination or determination

Although it is widely agreed that the jaw participates in the production of consonants and vowels (e.g. Mermelstein, 1973; Perkell, 1969; Wood, 1979), there is at the same time conflicting opinions as to the importance of the jaw during speech production. Some speech production theories assign only a subordinate role to the jaw, assuming that it merely assists the primary articulators (lips or tongue tip) in achieving their intended place of articulation (e.g. Browman & Goldstein, 1990; Saltzman & Munhall, 1989). However, there is some evidence that the jaw does more than just guide the primary articulator roughly into a particular position from which the primary articulator may execute its articulatory task. For example, different target jaw positions have been observed for /t/ and /d/ (but also other coronal consonants), although the respective places of articulation of the tongue are supposed to be the same (e.g. Keating, Lindblom, Lubker, & Kreiman, 1994; Kühnert, Ledl, Hoole, & Tillmann, 1991). In a systematic investigation of German coronal consonants, Mooshammer et al. (2007) revealed that the jaw's contribution to coronal constrictions varies with manner of articulation. Accordingly, a binary distinction can be made between /s, ʃ, t/ having higher jaw targets than /d, n, l/. They attributed these differences to different strategies concerning aerodynamic requirements for /s, ʃ/ and /t/ in order to produce high frequency noise (see also Iskarous, Shadle, & Proctor, 2011; Lee, Beckman, & Jackson, 1994; Shadle, 1985) and a salient burst, respectively. For coronals /d, n, l/, however, Mooshammer et al. (2007) argued for a lower jaw target in order to enlarge the oral cavity to maintain voicing. Mooshammer et al.'s categorization in terms of jaw height differences is in line with the results of other studies on this topic: sibilants /s, ʃ/ have been shown to invoke the comparatively highest jaw position, with /ʃ/ sometimes articulated with a higher jaw than /s/ (Amerman, Daniloff, & Moll, 1970; Hoole et al., 1990; Recasens, 2012b). Although /t/ was repeatedly reported for a slightly higher jaw than /d/, both alveolar stops have a targeted jaw position lower than that of the sibilants (Hoole et al., 1990; Keating et al., 1994; Lindblom, 1983), followed by the nasal /n/ (Hoole et al., 1990; Keating et al., 1994; Kühnert et al., 1991; Lindblom, 1983) and the lateral /l/ (Kühnert et al., 1991; Lindblom, 1983; Recasens, 2012b).

In relation to the question as to whether the jaw contributes to the production of consonants in a subordinate or determinant fashion, previous studies looked at the contextual jaw variability (e.g. Mooshammer et al., 2007; Recasens, 2012b). Related to the assumption that increasing coarticulation resistance is attributable to increasing coarticulatory constraint (e.g. Fowler & Brancazio, 2000; Recasens & Espinosa, 2009), Mooshammer et al. hypothesized that smaller contextual variability of the consonant's jaw height would provide evidence for more relevance of the jaw contribution to the consonant's production. Indeed there is cross-linguistic evidence that jaw height co-varies with degree of coarticulatory variability of the jaw, i.e. the contextual jaw variability is strongly related to the intrinsic jaw height of the consonant: least jaw variability was observed for consonants having a high jaw (i.e. /s, ʃ/ and /t/) while consonants with a comparatively low jaw (i.e. /d, n, l/) showed the highest degree of jaw variability (Geumann et al., 1999; Keating et al., 1994; Lee et al., 1994; Lindblom, 1983; Mooshammer et al., 2007; Recasens, 2012b; M. Stone & Vatikiotis-Bateson, 1995). These results suggest the following considerations concerning the jaw's importance for consonant productions: first, at least for consonants that demand a high and coarticulatory resistant jaw (i.e. /s, ʃ/ and /t/), the jaw plays a decisive role in coronal consonant production; second, the comparatively large degrees of contextual jaw variability reported for /d, n, l/ indicates a passive (or subordinate) function of the jaw, which is just helping the tongue tip to produce coronal constriction (cf. Mooshammer et al., 2007).

3.1.2 The jaw and the syllable

Other researchers assign even more importance to the jaw and argue that it exercises a substantial function in speech development, speech production, and syllable phonotactics (e.g. Davis & MacNeilage, 1995; Lindblom, 1983, 2011; MacNeilage, 1998; MacNeilage, Davis, Matyear, & Kinney, 2000; Redford, 1999a). It is generally assumed that vowels and consonants differ in their jaw position; a more open jaw (i.e. low jaw position) is expected for vowels compared to consonants (e.g. Keating et al., 1994; Lindblom, 1983). For the fundamental syllable pattern of consecutive consonants and vowels (i.e. CVC sequence), this means that jaw motions are characterized by an close-open-close jaw cycle (i.e. opening movement from the first consonant to the vowel; closing movement from the vowel to the second consonant). This observation led to the Frame/Content theory formalized by MacNeilage and colleagues. This is a biomechanical account of speech production that considers the jaw cycle as the phylogenetic and ontogenetic basis of syllables (‘frame’) onto which phonetic segments (‘content’) are superimposed. The Frame/Content theory emphasizes the structural function of the jaw cycle for syllable production, which means that the jaw cycle determines a mandibular opening-closing movement throughout a CVC sequence independent of the segmental make-up of the CVC sequence (e.g. Davis & MacNeilage, 1995; MacNeilage & Davis, 2000, 2001).

Carrying on the ideas of MacNeilage and colleagues, it has been hypothesized that the jaw cycle principle could be extended to syllables with complex consonant structures preceding a vowel (e.g. CCV syllables; Redford, 1999a; Redford & van Donkelaar, 2008, see also Lindblom, 2011). Syllable phonotactics and particularly the ordering of consonants within a syllable onset cluster is often said to be governed by the Sonority Sequencing Principle (SSP; e.g. Clements, 1990; Hooper, 1972; Selkirk, 1982). The SSP hypothesizes that onset consonants typically arrange such that sonority rises towards and falls away from the syllable nucleus (i.e. the vowel). In this perspective the syllable /bla/ would be a well-formed syllable. Even if the SSP is able to account for many instances of syllable patterns, conventional accounts of what generally constitutes sonority often fail to explain why consonant cluster pattern as they do (e.g. /bla/'s reverse sonority cluster /lba/). Since Lindblom (1983)[1] suggested the degree of jaw opening as one of the major determinants for sonority, Redford (1999a) hypothesized that the mandibular cycle could be an articulatory factor which determines consonant patterning such as /bla/ vs. /lba/ in Russian. However, her data provide only limited support for this hypothesis, i.e. although she could show a lower jaw position during obstruents than sonorants in sonorant+obstruent clusters (e.g. /lba/; jaw cycle), the respective consonantal jaw positions did not differ in obstruent+sonorant clusters (e.g. /bla/; no jaw cycle). Accordingly she stated that the weak correspondence between the mandibular cycle and the syllable patterning does not sufficiently explain sonority sequencing in terms of a biomechanical account of speech production (Redford & van Donkelaar, 2008).

3.1.3 The jaw and coarticulatory overlap

Even though the few studies discussed above showed at best limited support for the correspondence between the jaw cycle and syllable phonotactics, recently some suggestions have attributed particular syllable organization patterns to the jaw (e.g. Brunner et al., 2014; Pouplier, 2015; Rochet-Capellan & Schwartz, 2007). Relatedly, an interesting approach was put forward by Lindblom (2011), who aimed to explain coarticulation processes within a syllable in terms of a biomechanical perspective. This theoretical framework basically aggregates observations and assumptions of his earlier work (e.g. Keating et al., 1994; Lindblom, 1983). In Lindblom (1983) he could show for Swedish consonants that degree of jaw opening is positively correlated with the consonant's propensity to vowel coarticulation, i.e. degree of vowel coarticulation increases as a function of increasing jaw opening. From this result he suggested for cluster phonotactics that consonants with the smallest degree

[1]It should be mentioned that sonority hierarchies vary between Clements (1990) and Lindblom (1983): while Clements differentiates between manner of articulation (sonority: stop < fricative < nasal < liquid < glide < vowel), Lindblom considers sonority in terms of consonant-specific degrees of jaw opening (sonority: /s, ɧ/ < /p, t, k, b, d, g, f/ < /m, n, ŋ/ < /j, v/ < /r, l/) which resemble the commonly used sonority hierarchy of Swedish consonants.)

of jaw opening (e.g. /s, ʃj/) are phonotactically preferred in a vowel-remote position due to the articulatory incompatibility with the vowel, while consonants with moderate or considerable jaw opening (e.g. /n, l/) are suitable to appear in a position close to the vowel. However, as Keating (1983) critically noted, since only singleton consonants are considered in Lindblom (1983), his data are not capable to provide empirical support towards a biomechanical account of cluster phonotactics.

In order to explain coarticulatory processes in syllables with cluster onsets, Lindblom's (2011) approach starts from the core idea of the Frame/Content theory that the mandibular cycle constitutes the syllable with closed (i.e. high jaw position) and open (i.e. low jaw position) phases of the jaw related to consonants and vowels, respectively (e.g. MacNeilage, 1998; MacNeilage & Davis, 2000). According to this paradigm, as onset complexity of the syllable increases (i.e. $\mathrm{C_A V} \rightarrow \mathrm{C_R C_A V}$), the syllable constituents have to adapt their intrinsic jaw height to the mandibular cycle so that the jaw height continuously descends from the first consonant to the vowel, with an intermediate jaw height for the second consonant (i.e. jaw height: $\mathrm{C_R} > \mathrm{C_A} > \mathrm{V}$; cf. Redford, 1999a). More crucial to Lindblom's account, however is that this increase in onset/syllable complexity means that – figuratively speaking – more segments ('content') have to fit into the jaw cycle ('frame'). Applying the working assumption of a temporally invariant jaw cycle, he predicted substantial coarticulatory overlap of neighboring speech gestures (i.e. $\mathrm{C_R}$-$\mathrm{C_A}$ and $\mathrm{C_A}$-V overlap) due to the segmental density within the frame (cf. Redford, 2004). To illustrate this causality, Figure 3.1 shows how the jaw cycle might govern coarticulatory overlap. The two panels represent differing onset complexity (top: singleton $\mathrm{C_A V}$; bottom: cluster $\mathrm{C_R C_A V}$), while the corresponding jaw opening phases (red dashed demi-cycles) are invariant in the temporal dimension. In addition, hypothesized activation intervals associated with consonants and vowels (black solid lines)[2] are superimposed onto the mandibular demi-cycles. Comparing the activation intervals of the vowel-adjacent consonant ($\mathrm{C_A}$) in both complexity conditions

[2]Even though the temporal organization of the respective activation intervals is reminiscent of simplified gestural scores used in the Articulatory Phonology framework (AP; e.g. Browman & Goldstein, 1992), Lindblom (2011) does not explicitly integrate ideas of AP into his biomechanical model of coarticulation.

is particularly informative regarding Lindblom's (2011) account. On the one hand, since C_A is predicted to allow for more jaw coarticulation, C_A accommodates its jaw position with increasing onset complexity (i.e. a comparatively lower jaw in the cluster than in the singleton condition). On the other hand, with the emergence of the vowel-remote consonant (C_R) in the cluster condition, in Lindblom's view C_A should be subject to coarticulatory overlap with C_R (consider also the waxing and waning phases of C_R and C_A).

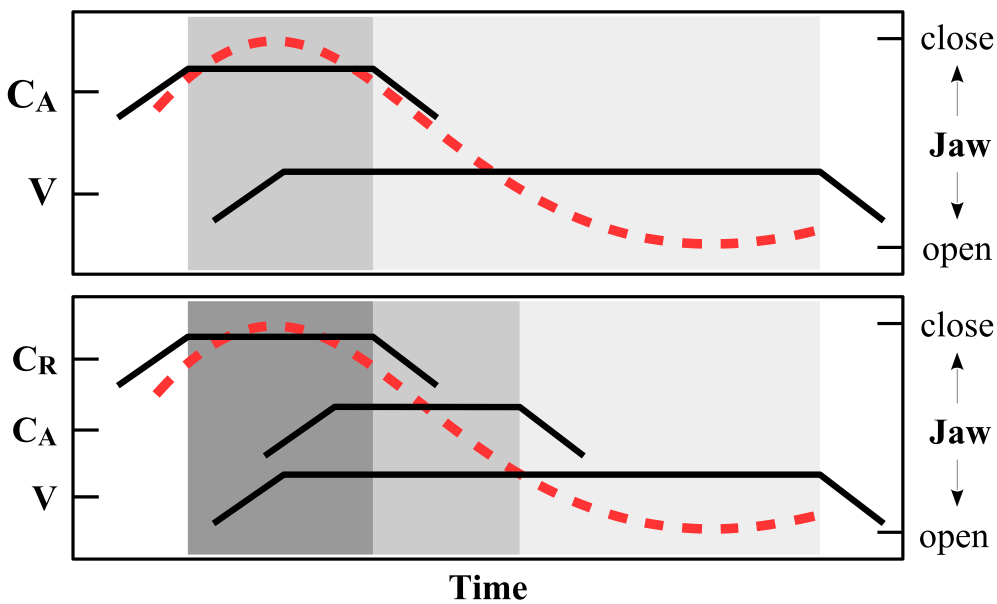

Figure 3.1: Two mandibular demi-cycles (i.e. opening phase) are illustrated for C_AV (top) and C_RC_AV (bottom) syllables which are layered with a hypothesized gestural score (black lines indicating activation intervals associated with C_R, C_A, and V gestures). The illustration shows that when the onset is a cluster (bottom), the C_A accommodates its targeted jaw position to the jaw cycle (lower jaw position compared to the top panel) and is temporally overlapped by the C_R gesture. In addition, the mid-gray portion (indicating the acoustic C_A duration) is shorter in the cluster than in the singleton condition.

Since Lindblom (2011) does not provide any empirical evidence towards the biomechanical foundation of coarticulatory overlap, his approach is thus far an untested theory. In addition, the current formulation of this theoretical model cannot account for different patterns of intra-cluster timing (i.e. gestural C_R-C_A overlap). However, previous research reveals patterns suggesting that gestural C_R-C_A overlap may vary as a function of – among other factors – cluster composition (e.g. Bombien et al., 2013; Byrd, 1996; Chitoran & Cohn, 2009; Hoole et al., 2013; Kühnert et al., 2006). Manner of articulation of both C_R and C_A has been proposed to affect the gestural timing of both consonants. Consonant clusters in which C_R was a stop showed relatively less gestural overlap compared to clusters with a fricative in C_R, and consonant clusters in which C_A was the nasal /n/ showed less gestural overlap compared to clusters with /l/ or /s/ in the vowel-adjacent position (e.g. Bombien et al., 2013; Kühnert et al., 2006; Yip, 2013). If there is a causal link between jaw cycle and coarticulatory overlap, this would be somewhat surprising since clusters like /kn, pn/ should be preferable to /ks, ps/ clusters (cf. Lindblom, 1983). Recall the jaw opening/sonority hierarchy in which nasals are considered to be more sonorant than stops (i.e. /kn, pn/: preferred cluster), while sibilants are considered to be less sonorant than stops (i.e. /ks, ps/: dispreferred cluster). It remains unclear whether this discrepancy between Lindblom's hypothesis and the actually observed overlap pattern unveils a first shortcoming of this assumption. It is also conceivable that this discrepancy is a matter of the overlap measurement applied or that other determining factors beside the jaw cycle come into play. This will be tested in this study.

In sum, less is known about how the jaw affects syllable phonotactics and the temporal organization of syllables (i.e. intergestural timing). Hence, using Polish onset clusters we investigate whether Lindblom's (2011) biomechanical approach can account for emerging coarticulatory overlap. Polish serves as a perfect test environment since its syllable structure has been classified as largely unpredictable (Piotrowski, 1992) and literally all combinations of consonants can be bundled into onset clusters whether or not they obey the Sonority Sequencing Principle (cf. Rubach & Booij, 1990). Accordingly, we assembled a set of Polish syllables with labial+coronal onset clusters in which the vowel-adjacent consonants varies in terms of manner of articulation, hypothesized jaw positions and jaw coarticulation resistance (i.e. /mʃ̠a/[3], /pt̠a/, /pn̠a/, and /pl̠a/; see e.g. Lindblom, 1983; Mooshammer et al., 2007; Recasens, 2012b). For the segmental make-up of our clusters it is important to note that labial stops /p, b/ have been found to have a lower jaw position than /t, d/ (Hoole et al., 1990; Keating et al., 1994; Lee et al., 1994) but a higher jaw position compared to /n, l/ (Recasens, 2012b). Accordingly, at least some of those clusters under investigation (/mʃa/, /pta/) violate both the conventional as well as the mandible driven sonority hierarchies (e.g. Clements, 1990; Lindblom, 1983; cf. Figure 3.2); hence it remains an open question whether Lindblom's approach also holds for these clusters.

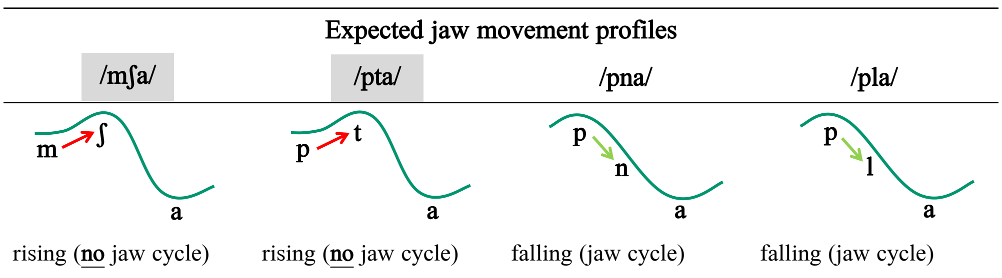

Figure 3.2: Schematic representation of the expected jaw movement profile according to the individual hypothesized jaw heights of C_R and C_A (see review above). Those clusters which are expected to violate the jaw cycle principle (i.e. 'rising') are highlighted in gray.

3.1.4 The jaw and incremental C_A shortening

Lindblom (2011) mentioned that consonant duration shortens incrementally as a function of increasing onset complexity. This may provide some support for a foundational role of the jaw in coarticulation. Indeed, there is some evidence for an overall tendency for consonants to be acoustically shorter in complex onsets than the same consonants in singleton onsets (e.g. American English: Haggard, 1973; O'Shaughnessy, 1974; Romanian: Marin & Pouplier, 2014), although the degree of relative duration changes may vary as a function of

[3]Recall that we use the symbol /ʃ/ (cf. Gussmann, 2007; Jassem, 2003) instead of the retroflex alternative /ʂ/ which is preferred by other researchers (e.g. Bukmaier & Harrington, 2016; Hamann, 2004; Żygis & Hamann, 2003).

manner of articulation, that is, sibilants and stops show relatively less acoustic shortening compared to laterals (cf. O'Shaughnessy, 1974; Syrika, Nicolaidis, Edwards, & Beckman, 2011). More specifically, Lindblom (2011) proposed that the general tendency towards incremental consonant shortening arises from the causal link between the jaw cycle and coarticulatory overlap: Since the internal consonant (C_A) is sandwiched between the vowel-remote consonant (C_R) and the vowel, C_A is acoustically shorter when part of an onset cluster than when it is the sole onset consonant. This is illustrated in Figure 3.1 by means of a shorter mid-gray portion in cluster (bottom) than in the singleton (top) panel. However, it remains unclear from previous studies as to whether incremental consonant shortening arises from the jaw cycle. Even though Redford (1999a) observed sonorant and obstruent consonants in vowel-adjacent position being shorter in cluster than in singleton condition (e.g. duration: /b$\underline{\text{l}}$-/ $<$ /$\underline{\text{l}}$-/ and /l$\underline{\text{b}}$-/ $<$ /$\underline{\text{b}}$-/; cf. Figure 3.1), only sonorant+obstruent clusters (e.g. /lb-/) showed a jaw cycle pattern (i.e. obstruents having a lower jaw than the sonorants), but obstruent+sonorant onsets (e.g. /bl-/) did not, contrary to expectations. In a subsequent study Redford (2004) argued in the same vein as Lindblom (2011) that long vs. short duration asymmetries typical for obstruent+sonorant clusters (e.g. Christie, 1977) are coarticulatorily rather than phonologically motivated. To probe the coarticulatory origin of the long-short duration pattern, she investigated Finnish obstruent+sonorant clusters spanning prosodic boundaries (i.e. clusters across syllable- and word-boundaries). Since in both prosodic conditions the sonorant was shorter than the obstruent, this was taken as evidence that consonant duration patterns are driven by a biomechanical constraint rather than by syllable structure. This means the sandwiched sonorant in an obstruent-sonorant-vowel sequence is compressed by the preceding obstruent and the following vowel to fit into the opening phase of the jaw cycle. In order to verify the biomechanical account for the long-short duration asymmetry, Redford (2004) also investigated the productions of obstruent+sonorant clusters using a jaw perturbation paradigm (e.g. bite block, clenched jaw). However, jaw perturbation turned out to be ineffective in terms of segmental duration patterns. Since obstruents emerged to be longer than sonorants despite the absence of jaw movements, Redford's study failed to demonstrate that consonant duration patterns were articulatorily driven.

In sum, Lindblom's (2011) suggestion that incremental consonant shortening arises from the causal link between the mandibular cycle and coarticulatory overlap is interesting, but lacks conclusive support thus far. Hence, we take up and elaborate his key ideas since they are potentially capable of linking several phonetic and syllable-related aspects. To do so we investigate for a variety of singleton onset consonants (i.e. /ʃ̠/, /t̠/, /n̠/, and /l̠/), to what extent they shorten in acoustic duration when they are part of a complex onset (i.e. /mʃ̠/, /pt̠/, /pn̠/, and /pl̠/). If we take previous results on incremental consonant shortening (cf. O'Shaughnessy, 1974; Syrika et al., 2011) and jaw coarticulation patterns (cf. Keating et al., 1994; Lindblom, 1983) into account, then a homogeneous pattern of susceptibility for incremental shortening and propensity to jaw coarticulation is predicted for the coronal consonants used in this study:

	— **increasing** →			
incremental shortening:	/ʃ/	/t/	/n/	/l/
degree of jaw coarticulation[4]:	/ʃ/	/t/	/n/	/l/

The parallel hierarchical ordering of coronal consonants predicts that the degree of incremental coronal shortening increases as the hypothesized degree of jaw coarticulation increases. By implication, the degree of incremental coronal shortening is highly dependent on the hypothesized jaw coarticulation resistance (i.e. the relevance of jaw contribution). Since Lindblom's (2011) hypothesized interaction between incremental consonant shortening and jaw cycle/coarticulatory overlap does not explicitly address inter-consonantal differences, the investigation of C_A = /ʃ, t, n, l/ in C_RC_AV sequences is expected to refine

[4]The degree of jaw coarticulation refers to the jaw's contextual variability in the *vertical* dimension. Indeed, there is some controversy as to whether /l/ is more variable than /n/. While Mooshammer et al. (2007) showed comparatively more contextual variability for /n/ than for /l/, the present ordering is motivated and supported by Lindblom (1983) as well as by cross-linguistic evidence (Swedish and English) presented in Figure 5 in Keating et al. (1994). In addition, Recasens (2012b) showed /n/ and /l/ at least to be equal in terms of vertical jaw variability (cf. Figure 2, bottom-right).

his rather general account.

3.1.5 Hypotheses

Given the considerations above, Lindblom's (2011) account is intriguing but appears somewhat premature for two reasons. Its major shortcoming is that it fails to consider that consonants differ with respect to jaw position (i.e. jaw opening) and the degree to which they accommodate jaw position to the demands of the phonetic surrounding (i.e. jaw coarticulatory resistance). Hence, if a consonant in the vowel-adjacent position (i.e. C_A) demands a comparatively higher jaw position than the vowel-remote consonant (i.e. C_R), then it is unlikely that C_A is acting in line with the jaw cycle principle. In order to do so, C_A would have to accommodate its jaw position to the new context. It is not clear that this is the case with consonants with high coarticulatory resistance. Redford's (1999a) study, where /bl/ does not conform to the jaw cycle suggests that consonants do not accommodate in this respect, so Lindblom's proposal may be incorrect in its strongest form. Secondly, Lindblom consciously proceeds from the simplified assumption that the time frame of an entire jaw cycle is invariant regardless of syllable complexity, that is, the frame is constant in duration for $C_AV(C)$ and $C_RC_AV(C)$ syllables (see Figure 3.1). For the example mentioned above in which the jaw position is expected to be higher for the vowel-adjacent than the vowel-remote consonant (e.g. /mʃ/), this conjecture is quite unlikely if we take the inertial mass of the jaw into account: raising ($C_R \rightarrow C_A$) and lowering ($C_A \rightarrow V$) the jaw mass presumably takes longer than simply lowering the jaw throughout the C_RC_AV syllable (as is expected for syllable /pla/). For our example, this means that theoretically the temporal frame determined by jaw movement is longer in some C_RC_AV than in C_AV and other C_RC_AV, i.e. the jaw temporal frame is expected to be longer in /mʃa/ than in /ʃa/ or /pla/. Hence, the degree of overlap is expected to be comparatively lower if the inertial mass of the jaw has to be lifted first before lowering for the vowel.

Taking these two shortcomings into account, we refine our predictions which contrast to some extent with those of Lindblom (2011):

(1) We argue that the emergence of the open/close jaw cycle is determined by the segmental make up of the onset cluster and not *vice versa*. In light of previous studies which investigated the jaw positions of consonants (e.g. Mooshammer et al., 2007; Recasens, 2012b) this means that the probability of an emerging jaw cycle decreases as the intrinsic jaw position (and the jaw coarticulation resistance) of the vowel-adjacent consonant (C_A) increases. For instance, if the vowel-adjacent consonant requires an (invariantly) higher jaw compared to the vowel-remote consonant, no jaw cycle is expected; if, however, the vowel-adjacent consonant requires a (flexible) lower jaw compared to the vowel-remote consonant, a jaw cycle is assumed to emerge. We therefore do not expect C_A jaw position to be accommodated to fit a jaw cycle, but rather we expect two jaw trajectory patterns to emerge (a raising and a falling one) as a function of intrinsic jaw position of the consonants involved.

(2) In the case of an apparent jaw cycle, we expect substantial coarticulatory overlap of C_R and C_A, i.e. the onset consonants should be closely coordinated (cf. Lindblom, 2011). If, however, no jaw cycle emerges due to C_A exerting a higher jaw position than C_R, then we expect a smaller degree of coarticulatory overlap (i.e. C_R and C_A gestures are produced further apart) than would be expected for an apparent jaw cycle pattern.

(3) Finally, we expect C_A duration to shorten acoustically as a function of the degree to which the vowel-adjacent consonant is overlapped by the vowel-remote consonant, i.e. if C_A is substantially overlapped by C_R, we expect a high degree of incremental C_A shortening with increasing onset complexity, while a lower degree of C_R-C_A overlap should have no influence on acoustic C_A duration.

3.1.6 Summary

Summing up, the present study investigates whether jaw movements may be a constraining factor for articulatory and acoustic changes as onset complexity increases. Lindblom (2011) suggested that a temporally constant jaw cycle movement ('frame') may be responsible for increasing gestural overlap and acoustic shortening of the vowel-adjacent consonant (C_A) when onset complexity increases (see Figure 3.1). Since related studies have unveiled some shortcomings in this theoretical approach, we argue instead that jaw movements throughout syllables with complex onsets are conditioned by the segmental make-up of the onset cluster and particularly the mandibular constraints of the vowel-adjacent consonant (i.e. jaw height and jaw coarticulation resistance). Our investigation of Polish labial+coronal onset clusters will be the first to examine the (inter)relation between jaw movements and coarticulatory C_RC_A overlap and their consequences for acoustic C_A duration. Hence, this study helps us to gain more knowledge about the biomechanical factors in syllable-related production variability.

3.2 Method

For this study we made use of articulography and acoustic data collected during the recording session described in Chapter 2. Therefore we refer to Section 2.2.1 for further details about our participants and the experimental setup.

The analyses applied in this chapter also required sets of singleton and cluster target words (e.g. set **pl** consists of target word [**l**atom] and [**pl**at͡som]). In the case of cluster target words the vowel-remote labials are assumed to have similar jaw height patterns (i.e. /m/ ≈ /p/), while the vowel-adjacent coronals differ in terms of jaw height and jaw coarticulation resistance (i.e /ʃ > t > n > l/). Apart from the presence/absence of the vowel-remote consonant, the phonetic structure of singleton and cluster target words was kept as similar as possible (see Table 3.1). To control for extrinsic variation, the phonetic and prosodic surrounding of the targeted clusters remained constant. Target words and carrier phrases were designed so that each cluster occurred in a vowel-symmetric /a#(C_R)C_Aa/ sequence, e.g. "*Jakub powtarz**a pla**com aktualnie.*" ('Jakub repeats placom currently.'). Further, we only used disyllabic target words in order to elicit lexical stress on the penultimate syllable, i.e. the syllable including the onset cluster (cf. Gussmann, 2007). Finally, to prevent variation in the realization of phrasal accent, we instructed the participants to accentuate the target words. These considerations were expected to ensure distinct closing and opening movements of the jaw from the first to second vowel (see, for instance, Erickson, 1998; Harrington, Fletcher, & Roberts, 1995; Summers, 1987 who found that low vowels in stressed word/sentence position were produced with a considerably lower jaw position than in unstressed position).

	Sets of cluster and singleton target words			
	mʃ	**pt**	**pn**	**pl**
Cluster	[m̲ʃ̲alik]	[p̲t̪̲akem]	[p̲n̲at͡ʂi]	[p̲l̲at͡som]
Singleton	[ʃ̲alik]	[t̪̲acim]	[n̲at͡ʂi]	[l̲atom]

Table 3.1: Sets of cluster and singleton target words are arranged according to expected jaw height of C_A (i.e. jaw height: /mʃ̲/ > /pt̲/ > /pn̲/ > /pl̲/).

The accentuation task caused one female speaker to often produce a distinct pause before the target word, possibly to put further emphasis on the accented target word. Particularly in the case of stop initial target words it was hence impossible to exactly identify the acoustic onset and the acoustic duration of the word initial stops. However, since these acoustic landmarks are crucial for most of our analyses (see below in Section 3.2.2), we excluded this speaker from the final data set. As a result, our analyses are based on data from only two female and three male Polish speakers.

Some further remarks should be made regarding the corpus set-up (Table 3.1). First, it should be noted that our stimuli include the non-word *pnaci* [pnat͡ʂi]. This is because the current data set was recorded as part of a larger corpus that imposed more requirements on the stimuli than those discussed above. However, *pnaci* is well-formed and quite similar to other lexical words in Polish (e.g. *pnący* [pnɔnt͡ʂɨ] ‘climbing (plant)’). Second, although the first consonant (C_R) in /mʃ/ is different from the other clusters /pt, pn, pl/, we gave preference to the /mʃ/ cluster in order to fulfill the requirement for a symmetric vowel context; /pʃ, ps/, which were also present in the larger corpus, were unsuitable in this regard because of /ɛ/ and /ɔ/ contexts, respectively. Thus the vowel preceding and following the cluster was always /a/.

In sum, the intended data set comprised four repetitions per target word and subject (8 words $\times$ 4 repetitions $\times$ 5 subjects = 160 utterances). However, occasional misreadings and technical difficulties during the recording session reduced the final data set to 148 items.

3.2.1 Articulography and acoustic data

In this study, articulogrpahy data are used for two reasons: a) to capture jaw movements throughout the first syllable of the target word (i.e. $C_R C_A V$) and b) to investigate the temporal relationship between the labial and the coronal gestures (i.e. $C_R C_A$ overlap)[5]. We employed Mark Tiede's *mview* algorithm to identify consonantal gestures in the movement time functions of the relevant articulator. C_R = /m, p/, which are produced with labial closure, were defined on the basis of lip aperture (LA), calculated as the Euclidean distance between the upper and the lower lip sensors. C_A = /ʃ, t, n, l/, which have a characteristic tongue tip constriction, were defined on the basis of the sensors attached to the tongue tip (TT). *mview* identifies semi-automatically articulatory landmarks of a given consonantal gesture on the basis of the filtered tangential velocity profile. That is consonant gestures are first manually determined in the respective articulator time series, then *mview* detects automatically the velocity maxima corresponding to the articulator movement towards (PVEL1) and away (PVEL2) from the constriction. The minimum articulator velocity within these landmarks is interpreted as the maximum constriction of the consonant (i.e. MAXC). The time points during the gesture's constriction movement at which the velocity increases beyond and falls below a 20%-threshold are referred to as the gesture onset (GONS) and the achievement of the articulatory target (NONS), respectively. Time points during the gesture's release movement at which the velocity exceeds falls below a 20%-threshold are referred to as the target release (NOFFS) and the gesture offset (GOFFS), respectively. We refer to the interval between NONS and NOFFS as the

[5]For further details on how time series of the articulatory movements were obtained see Section 2.2.1.

constriction plateau (cf. dark gray portions of /p/ and /l/ gestures in Figure 3.4).

In addition to the kinematic data, we recorded synchronized acoustic data at a sampling rate of 25 600 Hz. In the first instance, acoustic speech signals were automatically segmented by the MAuS algorithm (Schiel, 2004) before a phonetically trained annotator made manual corrections using Praat (Boersma, 2001). If manual corrections were necessary a combination of spectral and acoustic cues was applied to properly identify the segment boundaries.

To identify segment boundaries in /a#mʃa/ sequences, we used the end of the first vowel detectable in terms of changes in the periodic waveform (from /a/→/m/), the onset of aperiodic noise (from /m/→/ʃ/), and the onset of periodic oscillation (from /ʃ/→/a/). Sequences /a#pta/, /a#pna/, and /a#pla/ have in common that /p/ was identified on the basis of the acoustic offset of the heterosyllabic vowel (from /a/→/p/) and the release burst. This means that aspiration if present does not belong to the acoustic /p/ duration. Similarly, we segmented the vowel-adjacent /t/ in terms of the silence interval between the end of aspiration (from /p/→/t/) and the release burst (from /t/→/a/). Note that we excluded those tokens from our analyses, in which /p/ and /t/ could not be acoustically separated from each other due to an acoustically missing /p/ release. For the vowel-adjacent /n, l/ we took the onset of periodic oscillation (from /p/→/n, l/) and changes in the periodic waveform (from /n, l/→/a/) as acoustic markers to distinguish the sonorant portions from the phonetic surrounding.

In addition to the acoustic segmentation guidelines, one further remark should be made regarding the aspiration of voiceless stops. Similar to German and English stops, Polish singleton stops are produced with distinct aspiration before stressed vowels (cf. Wierzchowska, 1980 cited in Ruszkiewicz, 1990; Malisz & Żygis, 2015; Waniek-Klimczak, 2011a, 2011b). Therefore, in this study we restricted stop duration to the closure interval of the stop, while the corresponding aspiration is treated as a separate interval (i.e. duration of /t/ is the acoustic silence without aspiration). This was done for two reasons. Firstly, in this way, we obtain parallel acoustic segmentation patterns for all vowel-adjacent consonants (cf. Umeda, 1977). Secondly, since aspiration duration is generally longer for voiceless stops occurring in singleton than in cluster onsets (e.g. Klatt, 1975)[6], we prevent *a priori* that aspiration differences in singleton /t̠ʰ/ and cluster /pt̠/ confound our acoustic duration measurements (see Section 3.2.2).

3.2.2 Measurements

Jaw movements

Since our hypotheses are tied to the presence or absence of the so called jaw cycle, the first analysis of this study concerns the specific jaw movements throughout /mʃa/, /pta/, /pna/, and /pla/ syllables. We consider the jaw cycle as a trajectory that constantly decreases in jaw height from the syllable onset to the nucleus, that is, jaw opening is small during C_R and large during the vowel (V), while being intermediate during C_A.

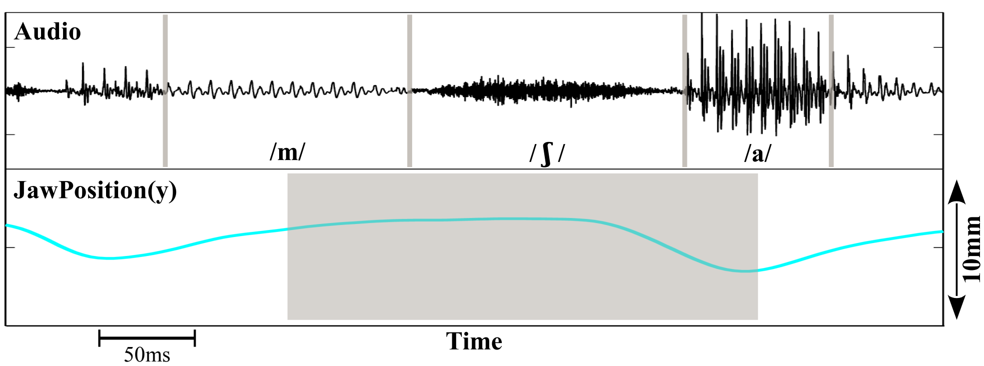

Figure 3.3: The top tier in this illustration shows the acoustic segmentation of /m/, /ʃ/, and the vowel /a/ of one particular [m̠ʃalik] item. The bottom tier shows the vertical position of the jaw as a function of time. The shaded box indicate the interval between the acoustic midpoints of /m/ and /a/ which is used for extracting the vertical jaw movements. The DCT curvature index for this example (= -1.002) indicates a ∩-shape during this particular jaw trajectory, i.e. no jaw cycle.

[6]To our knowledge there is no systematic investigation of whether VOT duration is shorter in cluster than singleton stops in Polish.

Based on the acoustic segmentation of the speech material, we extracted the vertical jaw movements between the acoustic midpoints of C_R and the heterosyllabic vowel (i.e. the shaded interval in Figure 3.3). In order to describe the jaw trajectories qualitatively, we performed a discrete cosine transform on these extracted time functions. The discrete cosine transformation is a mathematical decomposition of a signal into a set of sinusoids which reconstructs the original input signal if added together again. As Harrington (2010) pointed out, the first three DCT coefficients approximately quantify the mean, slope and curvature of a trajectory. Beňuš and Pouplier (2011), for example, used the third DCT coefficient (i.e. curvature) to establish whether the endpoints of a trajectory lie above or below the respective apex, that is, whether the trajectories were ∪-shaped (positive coefficient) or ∩-shaped (negative coefficient), respectively. This quantification was interpreted as indicating the presence or absence of a syllable-bound jaw cycle. Regarding our hypotheses, we expect that syllables with intrinsically low C_A jaw position should show curvature indices above or around 0 (e.g. for /pla/ syllables), while curvature indices should be negative for syllables in which C_A is produced with a comparatively higher jaw than C_R (e.g. /mʃa/; see Figure 3.3).

Gestural C_R-C_A overlap

The next analysis concerns the degree of coarticulatory overlap between onset cluster constituents. Previous research used different measures to investigate the temporal C_R-C_A organization. A frequently used measure is the C_R-C_A plateau lag (e.g. Bombien, Mooshammer, Hoole, & Kühnert, 2010; Byrd & Choi, 2010; Pouplier, 2012), but this measure cannot sufficiently evaluate the relationship between coarticulatory encroachment of adjacent segments and consonant shortening. Hence, in this study we use an overlap measure which is derived from that used in Chitoran, Goldstein, and Byrd (2002)[7]: Rather than quantifying the degree to which the C_R constriction plateau is overlapped by the onset movement of C_A, our measure determines the percentage of C_A constriction plateau (i.e. the interval

[7]See also Byrd (1996) and Byrd and Tan (1996) for a related measure for electropalatography (EPG) data.

between target attainment and target release) that is temporally overlapped by the C_R gesture's release phase (see cross-hatched portion of C_A plateau in Figure 3.4):

$$C_R\text{-}C_A \text{ overlap} = 100 * \frac{C_R(\text{GOFFS}) - C_A(\text{NONS})}{C_A(\text{NOFFS}) - C_A(\text{NONS})}$$

If the coarticulatory overlap index falls between 0% and 100%, the labial C_R gesture terminates during the constriction interval of C_A. If the index is >100%, C_R gesture terminates while C_A gesture has been articulatorily released. If, however, the index is <0%, the labial gesture (associated with C_R) is completed before C_A's tongue tip reaches its target.

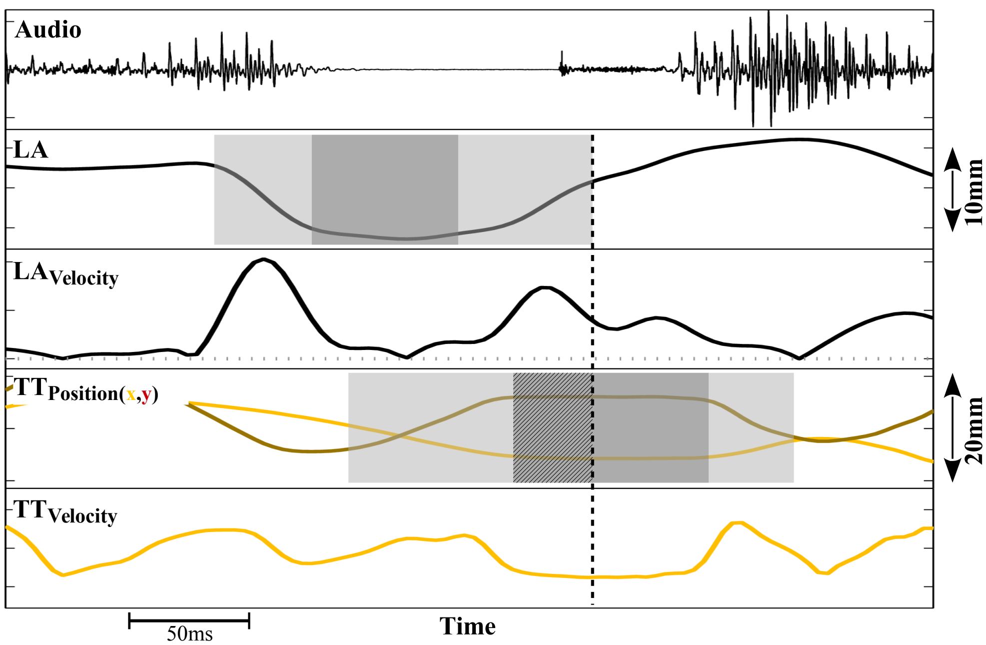

Figure 3.4: For a given instance of target word [pla͡tsom], this plot shows for /p/ and /l/ gestures the movement and velocity profiles for lip aperture (LA) and tongue tip (TT), respectively. For both gestures, the dark gray period indicates the constriction plateau, while the light gray periods prior to and following the plateau indicate the interval where the gesture waxes and wanes, respectively. The portion of C_A's constriction plateau, which is overlapped by C_R, is cross-hatched. In this particular example the percentage of coarticulatory overlap amounts 42.9%.

The advantage of the coarticulatory overlap measure over other timing measures (such as C_R-C_A lag) is that wax-and-wane intervals of speech gestures are taken into account, during which the influence of a speech gesture on the vocal tract increases and decreases, respectively (Joos, 1948). For the present study, these intervals are crucial since coarticulation patterns as well as acoustic shortening processes have been associated with a gesture's waxing and waning (Fowler, 1984; Fowler & Thompson, 2010).

Following our hypotheses, we expect large degrees of coarticulatory overlap, if C_RC_AV syllables arrange according to the jaw cycle pattern. On the other hand, syllables violating the jaw cycle principle are expected to show comparatively less C_RC_A overlap.

C_A duration ratio

In order to investigate the articulatory origins of acoustic consonant shorting in clusters, we applied our duration measurements to acoustic instead of the articulography data. We focus on the acoustic duration, since from an articulatory perspective labial (LA) and coronal (TT) gestures may overlap to a large extent without affecting the duration of the overlapped gesture. This means that if the duration for a C_A gesture is constant (regardless of onset complexity), increasing C_R–C_A overlap should cause a shortening of acoustic C_A duration while the articulatory C_A duration remains unchanged. Hence, we predict that the acoustic data are more informative in terms of the effect of articulatory processes on acoustic consonant shortening in clusters.

Based on the acoustic segmentations, the degree of C_A shortening is examined in terms of duration ratios of singleton and cluster C_A duration (henceforth, Dur_{Ratio}). Firstly, we measured the acoustic duration of C_A in [mʃ̠alik], [pt̺̠akem], [pn̠at͡s̹i], and [pl̠at͡som]. Secondly, we calculated the acoustic duration of the corresponding target words with singleton onsets to estimate the 'intrinsic' acoustic duration of [ʃ̠alik], [t̺̠acim], [n̠at͡s̹i], and [l̠atom] as a baseline for the Dur_{Ratio} calculation. The cluster/singleton C_A durations were then taken to determine the Dur_{Ratio} of each cluster token (e.g. [pl̠at͡som]) relative to the averaged durations of all corresponding singleton tokens (e.g. [l̠atom]). Finally, we centered the $\mathrm{Dur}_{\mathrm{Ratio}}$ indices to zero[8]. Positive $\mathrm{Dur}_{\mathrm{Ratio}}$ indices represent an acoustic lengthening of C_A with increasing onset complexity. Negative $\mathrm{Dur}_{\mathrm{Ratio}}$ indices represent a shortening of C_A in the cluster compared to the singleton condition. $\mathrm{Dur}_{\mathrm{Ratio}}$ indices around 0 suggest equal C_A durations in the singleton and cluster conditions.

In terms of C_A duration in the singleton and cluster conditions, we expect an effect of the degree of coarticulatory overlap on the degree of relative C_A shortening. That is, if there is no coarticulatory overlap between C_R and C_A, the acoustic duration of C_A should not be affected. If, however, C_A is largely overlapped by C_R, extensive C_A shortening is expected.

3.2.3 Statistical analyses

All statistical analyses were carried out in the R environment (R Core Team, 2013) by means of linear mixed models (*lme4* package: Bates, 2010; Bates et al., 2015). P-values were obtained by comparing one model with and one without the fixed factor the interaction of interest. If not declared otherwise, we included Speaker as a potential random factor; Repetition as additional random factor was also tested, but rejected if not required by the respective model. Post-hoc Tukey tests (*multcomp* package: Hothorn et al., 2008) were carried out to perform pairwise comparisons. To test whether jaw movement curvature, C_RC_A overlap, and incremental C_A shortening correlated with each other, we used non-parametric correlation tests (Spearman's rank correlation) since our data were not normally distributed.

3.3 Results

This section is separated into three subsections in which we present the results of the individual measurements, that is, jaw movement patterns: Section 3.3.1; coarticulatory C_RC_A overlap: Section 3.3.2; incremental C_A shortening: Section 3.3.3. Taking into account these individual results, we are then able to report on the possible interdependency of these

[8]This procedure is in principle identical with the one we used to calculate the lag ratios in Section 2.2.3.

three indices (Section 3.3.4). In reference to our hypotheses, we will discuss Hypothesis (1) in Section 3.3.1 and Hypotheses (2) and (3) in Section 3.3.4.

3.3.1 Jaw movement patterns (Index 1)

Recent accounts assume that the jaw cycle could be the articulatory basis for syllable organization (Lindblom, 2011; Redford, 1999a). Regarding phonotactic sequencing, this assumption allows for two interpretations: either segment-specific jaw height is integrated into a language's phonotactics so that C_R has consistently higher jaw positions compared to C_A, or C_A accommodates its jaw position individually so that the jaw cycle emerges. In accordance with jaw coarticulation studies (Mooshammer et al., 2007; Recasens, 2012b) we hypothesized, however, that a typical jaw cycle movement would fail to appear if C_A has an intrinsically high and hence coarticulation resistant jaw (as we hypothesize that C_A jaw position does not change to accommodate C_R and conform to a falling cycle).

Figure 3.5 shows the individual (dotted) and averaged (solid) jaw movements for /mʃa/ [mʃalik], /pta/ in [pt̪akem], /pna/ in [pnat͡s̢i], and /pla/ in [plat͡som]. The trajectories correspond to the jaw movements between the acoustical midpoints of the first consonant and the vowel, respectively, and are normalized for time (x-axis) and jaw height (y-axis). The averaged jaw trajectories (thick solid line) start at a similar height for all syllables during C_R = /m, p/ and reach a considerably lower position for the low vowel /a/. However, our focus lies on the interval in between, i.e. the articulatory path that the jaw takes from the syllable edge towards the syllable nucleus. Here, noticeable differences are apparent particularly between /pla/ and the remaining syllables /mʃa/, /pta/, and /pna/. In /pla/ the lateral has a lower jaw position compared with the initial labial (i.e. jaw cycle), while for the remaining syllables the jaw rises first from the initial (C_R) to the following consonant (C_A) before lowering for the vowel. Hence, the movement pattens for /mʃa/, /pta/, and /pna/ are at odds with the jaw cycle principle. Since the trajectories (Figure 3.5 top) do not allow the exact jaw positions of the three syllable constituents to be identified, the boxplot on the bottom of Figure 3.5 shows discrete jaw heights at the respective acoustic

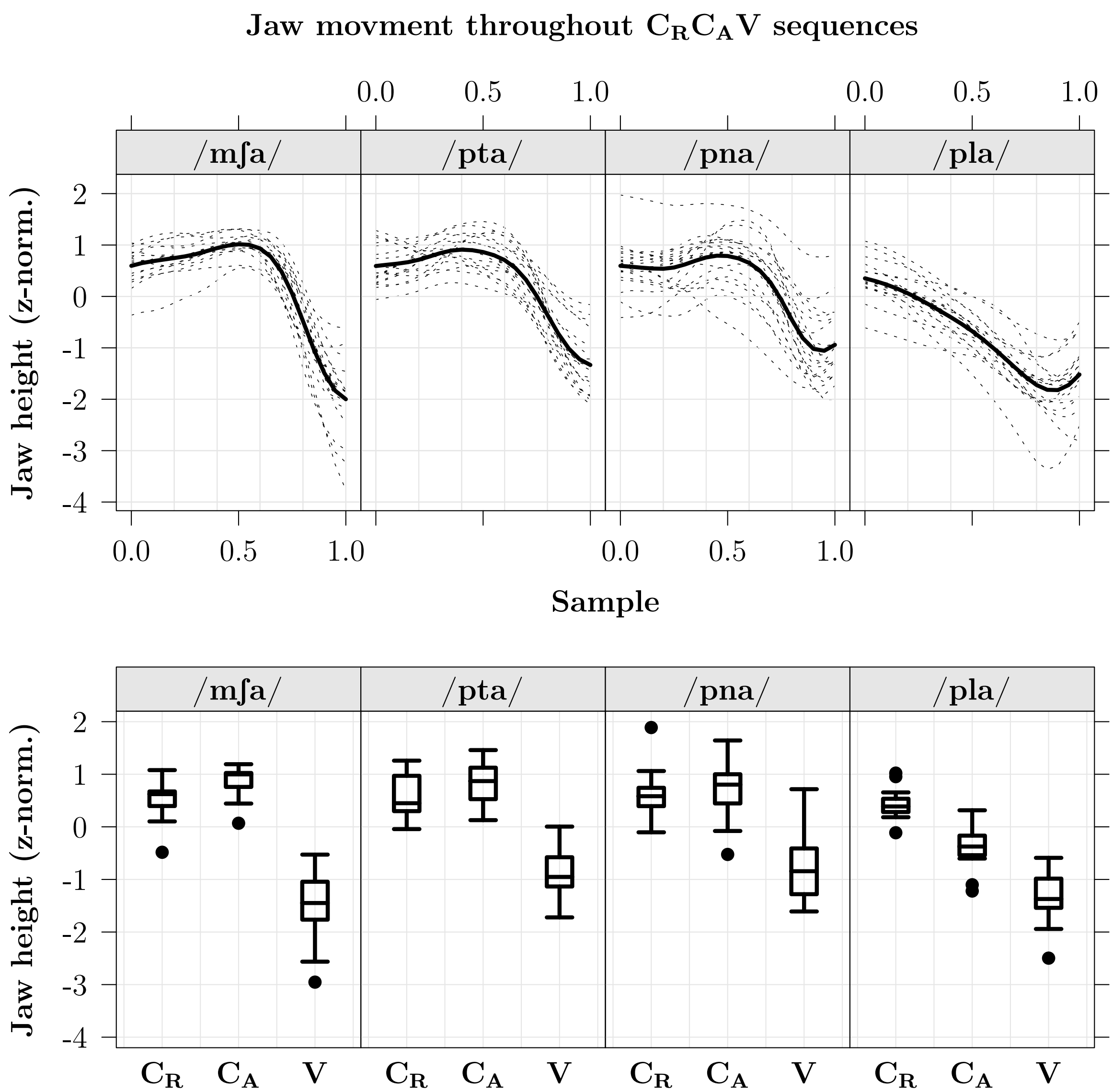

Figure 3.5: Top: For each of the four target syllables, the dotted lines indicate the z-normalized jaw height trajectories throughout CCV syllables. The temporal variability (x-axis) has been normalized as well; 0.0 and 1.0 indicate the acoustic midpoints of C_R and the vowel, respectively. The thick solid lines illustrate the averaged jaw trajectories for each particular syllable. Bottom: Discrete jaw positions extracted at the acoustic midpoints of C_R, C_A, and the vowel (V) are given as a function of the syllable.

midpoints of C_R, C_A, and the vowel.

In order to describe the jaw trajectories quantitatively, we performed a discrete cosine transformation (DCT) on the extracted time functions of vertical jaw movement. Particularly the third DCT coefficient was expected to give qualitative information about the curvatures of mandibular trajectories (Figure 3.6). The negative curvature coefficients of the jaw trajectories of /mʃa/, /pta/, and /pna/ suggest ∩-shaped jaw trajectories which is indicative of C_A having a higher jaw position than C_R and the vowel (i.e. no jaw cycle; see Figure 3.5), but the curvatures seem to vary to some extent. /pla/ is characterized by a curvature value around 0 which indicates a linear lowering of the jaw from C_R towards the vowel (i.e. jaw cycle). Even if /pna/ unexpectedly patterns with /mʃa, pta/ rather than with /pla/, this result is a first indication that a jaw cycle movement *only* occurs if the respective intrinsic jaw height of the syllable constituents (i.e. C_R, C_A, and V) facilitates it (cf. Hypothesis (1)).

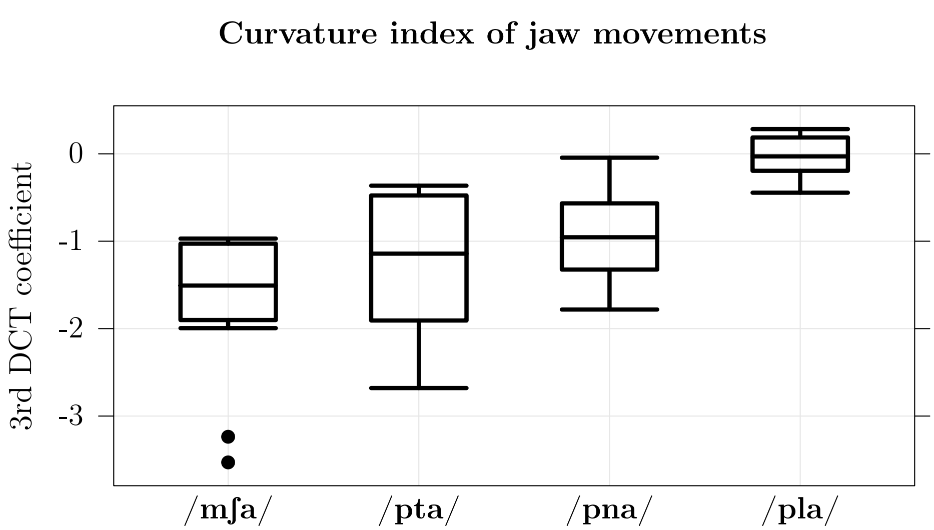

Figure 3.6: This plot shows the DCT curvature coefficients separated for syllables /mʃa/, /pta/, /pna/, and /pla/. Negative coefficients indicate a ∩-shaped jaw movement curvature; coefficients around 0 represent a linear decrease of jaw height throughout the syllable.

To investigate the differences of jaw trajectories in terms of DCT curvature coefficients with respect to segmental make-up of the syllable (Figure 3.6), we carried out a mixed model analysis with dependent variable Curvature and fixed factor Syllable (four levels: /mʃa/, /pta/, /pna/, and /pla/); Speaker served as random factor. There was a significant effect of Syllable on Curvature (χ^2[3]=66.0, p<0.001). This statistically corroborates our Hypothesis (1) that a jaw cycle arrangement is not obligatory, but depends on the segmental make-up of the syllable or more precisely on the intrinsic jaw position of C_A relative to C_R. Pairwise comparisons of the four syllables revealed highly significant differences in jaw trajectory curvature for most comparisons, but not for /mʃa/ vs. /pta/ (p=0.09) or /pta/ vs. /pna/ (p=0.4).

We have shown so far that a jaw cycle pattern only emerges for /pla/, but not for /mʃa/, /pta/, or /pna/. This is surprising insofar as the nasal in /pna/ was hypothesized to have a relatively lower (and hence less coarticulation resistant) jaw compared to the preceding labial. Due to the hypothesized falling jaw movement profile (cf. Figure 3.2), we expected that this syllable was likely to show a jaw cycle pattern. This leads us to the question of whether jaw position of the vowel-adjacent coronals changed as onset complexity increased. In order to estimate the accommodation of the jaw positions of the vowel-adjacent coronals with increasing onset complexity, we compared the vertical jaw position of coronals in singleton and cluster onsets (Figure 3.7). Accordingly, we ran a mixed model to investigate the interaction of Syllable (four levels: /mʃa/, /pta/, /pna/, and /pla/) and Complexity (two levels: singleton and cluster) on the jaw's vertical position of the vowel-adjacent consonant. Speaker was included as random factor. Although there was a significant interaction between both fixed factors (χ^2[3]=12.5, p<0.01), post-hoc Tukey tests revealed that none of the vowel-adjacent consonants accommodated jaw height with increasing onset complexity. This means that the jaw height of C_A does not adapt to a particular jaw cycle pattern; instead, vowel-adjacent consonants preserve their intrinsic jaw positions independent of onset complexity (cf. Figure 3.7). Taking together the results thus far, we can conclude that it is not the jaw cycle paradigm that predicts the jaw positions of the involved segments. Rather, it is the segment-specific jaw positions that predict the jaw cycle movement pattern (cf. Hypothesis (1)).

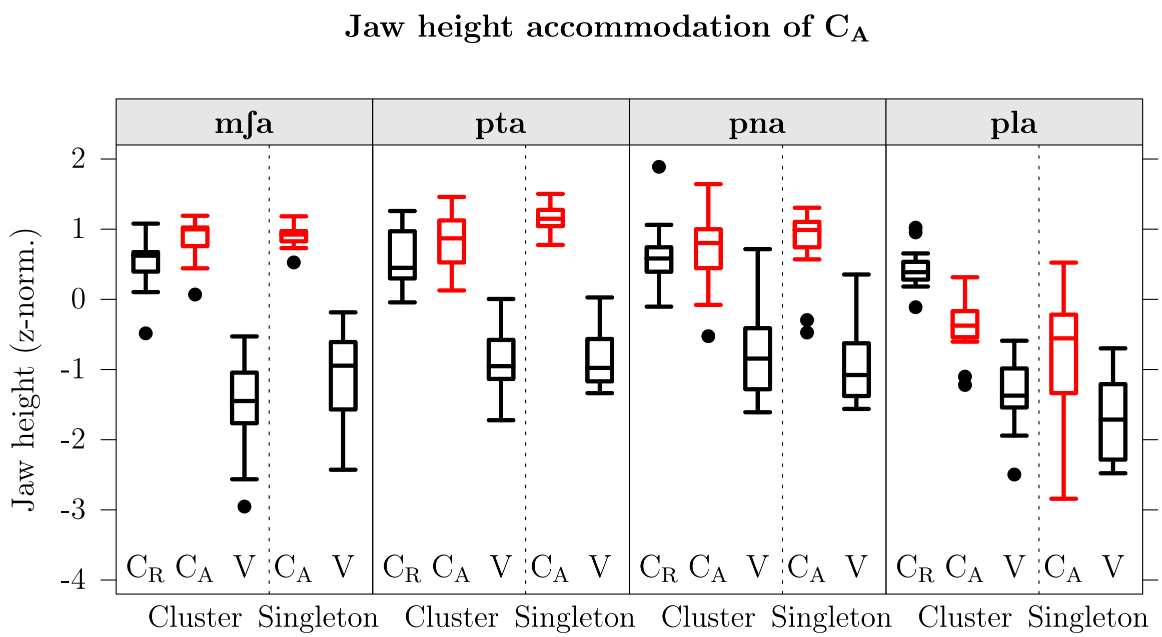

Figure 3.7: Z-normalized jaw position data for each segment in the target syllable (/mʃa/, /pta/, /pna/, and /pla/) and the corresponding singleton condition. In order to assess to what extent the vertical jaw position of the vowel-adjacent coronal accommodated as a function of increasing onset complexity, data for C_A are highlighted in red across complexity conditions.

3.3.2 Gestural C_R-C_A overlap (Index 2)

In this section we report the extent to which the initial consonant (C_R) overlapped the constriction plateau of the second consonant (C_A) and whether this particular overlap pattern varies as a function of cluster composition. Table 3.2 gives for each cluster the mean and standard deviation values of the coarticulatory overlap index (in %) and indicates different overlap patterns between clusters. Mean values indicate that constriction plateaus of the coronal was partially overlapped by the labials in /mʃ/ (36.9%) and /pt/ (27.4%) onset clusters. In the case of /pn/ the labial gesture terminated on average when the following nasal reached its constriction plateau (0.5%; no coarticulatory overlap). Finally, the labial gesture in /pl/ largely encroaches the constriction plateau of the lateral (200.6%; recall that values greater than 100% indicate that C_R finishes after C_A is articulatorily released, or in other words, the entire C_A plateau is encroached by C_R).

	Coarticulatory overlap index			
	/mʃ/	**/pt/**	**/pn/**	**/pl/**
mean	36.9%	27.4%	0.5%	200.6%
SD	58.7%	60.7%	40.3%	224.9%

Table 3.2: Mean coarticulatory overlap indices are given as a function of onset cluster make-up. In addition, standard deviation (*SD*) values are listed in order to indicate the data dispersion within each cluster type.

However, the standard deviation (*SD*) value of the /pl/ cluster (224.9%) hints at statistical outliers in the /pl/ productions which might be responsible for the considerably larger variance in the /pl/ overlap patterns compared to /mʃ/, /pl/, and /pn/. In order to identify possible outliers in the coarticulatory overlap indices we used a method that basically refers to the interquartile range (IQR = 3rdQuartile - 1stQuartile) of the entire data set to distinguish the outliers from the 'reasonable' tokens. By our definition, outliers are identified if they fall beyond a lower or upper threshold (i.e. 1stQuartile - 3∗IQR and 3stQuartile + 3∗IQR, respectively). By this means we could identify four /pl/ tokens of one single speaker (PL5). The evaluation of these tokens suggests that PL5 applied a different articulatory strategy to approach the tongue tip target position for /l/ in /pl/ clusters than the remaining subjects. This particular strategy caused a crucial bias in the articulatory labeling and hence in the determination of coarticulatory overlap indices of /pl/ clusters[9]. Therefore, based on the statistical and heuristic outlier examination, we excluded the four /pl/ tokens of subject PL5 from all analysis concerning coarticulatory overlap. Table 3.3 gives the mean and standard deviation values of the coarticulatory overlap index after outlier removal. To test whether Cluster (four levels: /mʃ/, /pt/, /pn/, and /pl/) differed in terms

[9]In Appendix A.2 we show two representative examples illustrating the different articulatory strategies and discuss the respective consequences for the coarticulatory overlap measure.

of Coarticulatory Overlap Index, we ran a mixed model with Speaker as random factor. The effect was significant ($\chi^2[3]=19.4$, $p<0.001$) and the post-hoc Tukey test revealed that cluster /pl/ differed significantly from /pt/ ($p<0.05$) and /pn/ ($p<0.001$), and was close to be significantly different from /mʃ/ ($p=0.0506$). The remaining comparisons yielded no significant differences. This means the degree of intergestural overlap is statistically similar in /mʃ/, /pt/, and /pn/; only /pl/ deviates to some extent from the remaining clusters, showing comparably more coarticulatory overlap of C_R and C_A.

	Coarticulatory overlap index			
	/mʃ/	**/pt/**	**/pn/**	**/pl/**
mean	36.9%	27.4%	0.5%	100.8%
SD	58.7%	60.7%	40.3%	86.2%

Table 3.3: Mean coarticulatory overlap indices (outliers excluded) are given as a function of onset cluster make-up. In addition, standard deviation (*SD*) values are listed in order to indicate the data dispersion within each cluster type.

3.3.3 Duration of vowel-adjacent consonant (Index 3)

This analysis examines to what extent the acoustic duration of the vowel-adjacent consonants (C_A) changes as a function of onset complexity. To do so we assembled singleton and cluster target words into sets. The following sets are considered: **mʃ** ([ʃalik] vs. [mʃalik]), **pt** ([t̪acim] vs. [pt̪akem]), **pn** ([nat͡s̨i] vs. [pnat͡s̨i]), and **pl** ([latom] vs. [plat͡som]). Figure 3.8 compares the acoustically measured duration of the vowel-adjacent consonant occurring in singleton (light gray) and cluster target words (dark gray) for each set (**mʃ**, **pt**, **pn**, and **pl**). Overall, the relative consonant duration patterns vary between sets. We discuss the set-specific results according to the order in which they appear in Figure 3.8 from left to right. In parentheses, we present the extent to which the acoustic consonant duration varied between singleton and cluster condition, i.e. the $\mathrm{Dur}_{\mathrm{Ratio}}$ which was calculated for each set and speaker. Negative $\text{Dur}_{\text{Ratio}}$ values indicate incremental C_{A} shortening with increasing onset complexity. Positive $\text{Dur}_{\text{Ratio}}$ values suggest longer C_{A} durations in the cluster than in the singleton condition. $\text{Dur}_{\text{Ratio}}$ values of 0 indicate no changes in terms of acoustic duration.

The results for set **mʃ** show almost identical sibilant durations in the singleton and the cluster condition (mean $\text{Dur}_{\text{Ratio}}$: -0.004). The absolute durations of singleton /$\underline{\text{ʃ}}$/ (mean: 134 ms) and cluster /m$\underline{\text{ʃ}}$/ (mean: 132 ms) are similar to the /$\underline{\text{s}}$/ and /p$\underline{\text{s}}$/ tokens found for adult Greek speakers reported in Syrika et al.'s (2011) work (see also Umeda (1977) on singleton /$\underline{\text{ʃ}}$/ durations). In contrast the remaining sets show shorter C_{A} durations in complex condition. For set **pt** we observe a considerably shorter duration for /t/ in the cluster than in the singleton condition (mean $\text{Dur}_{\text{Ratio}}$: -0.293). Recall that the acoustic analysis for the voiceless stop only considered the acoustic closure duration but not the VOT duration. Hence the incremental shortening found for /t/ (/$\underline{\text{t}}$/ $\rightarrow$ /p$\underline{\text{t}}$/) cannot be attributed to VOT shortening (see Klatt (1975) presenting VOT duration differences in English /t/ in singleton and cluster onsets), that is, the shortening of /t/ from singleton to cluster condition reflects the shortening of the closure duration only. For set **pn** a less clear pattern emerges from Figure 3.8. Although the pattern indicates incremental C_{A} shortening to a certain extent (mean $\text{Dur}_{\text{Ratio}}$: -0.137), the /$\underline{\text{n}}$/ vs. /p$\underline{\text{n}}$/ duration difference is not as clear as the one found for **pt**. Finally, set **pl** shows a similar pattern to **pt** (mean $\text{Dur}_{\text{Ratio}}$: -0.399), that is, singleton /$\underline{\text{l}}$/ is acoustically longer than cluster /p$\underline{\text{l}}$/ (cf. O'Shaughnessy, 1974).

To investigate whether the extent of relative duration change varied between sets, we carried out a mixed model analysis with dependent variable $\text{Dur}_{\text{Ratio}}$, fixed factor Set (four levels: **mʃ**, **pt**, **pn**, and **pl**), and random factor Speaker. The relative change of vowel-adjacent consonant duration from singleton to cluster condition differed significantly across Sets ($\chi^2[3]=48.7$, $p<0.001$). This may indicate that incremental C_{A} shortening varies as a function of C_{A}. Post-hoc Tukey tests revealed that $\text{Dur}_{\text{Ratio}}$ values patterned together for sets **pl**/**pt** as well as for sets **mʃ**/**pn**, that is, the comparisons **pl** vs. **pt** and **mʃ** vs. **pn** turned out not to be significant. The remaining comparisons reached significance levels.

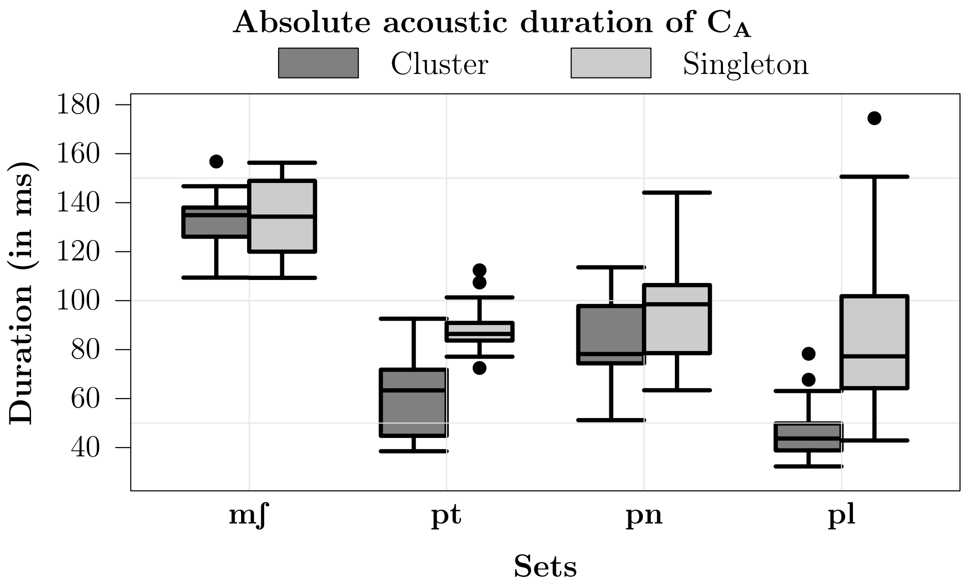

Figure 3.8: Absolute acoustic durations (in ms) of vowel-adjacent consonants are given with respect to set (**mʃ**, **pt**, **pn**, and **pl**) and onset complexity (cluster: dark gray; singleton: light gray).

In addition to the between-set analysis, a second mixed model analysis was carried out to provide further information on the extent of incremental C_A shortening. The interaction between Set (four levels: **mʃ**, **pt**, **pn**, and **pl**) and Complexity (two levels: singleton and cluster) on the dependent variable C_A Duration was tested with Subject as random factor. This mixed model showed a significant interaction between both fixed factors ($\chi^2[3]=20.9$, $p<0.001$), and the subsequent post-hoc Tukey comparisons confirmed significant acoustic C_A shortening only for sets **pt** ($p<0.001$) and **pl** ($p<0.001$). For sets **mʃ** and **pn** the acoustic duration of vowel-adjacent /ʃ, n/ did not shorten as a function of increasing onset complexity (**mʃ**: $p=1.000$; **pn**: $p=0.188$).

If we take together the results of both statistical analyses, the four sets can be split into two groups: the first group comprises sets **mʃ** and **pn** which have been found not to reduce acoustic C_A duration with increasing onset complexity. The second group comprises sets **pt** and **pl**. These sets consistently showed a reduced acoustic duration of the vowel-adjacent consonant in the cluster compared to the singleton condition. This raises the question of whether coarticulatory overlap could have caused the apparent C_A compression. At least informally, given that for instance **pt** and **pl** pattern differently on the two measures, we may expect a negative answer to this.

3.3.4 Interdependency of indices

With the results discussed above we laid the foundation for the forthcoming analyses. We go now into further detail on whether jaw movement patterns correlate with the coarticulatory overlap of C_RC_A cluster constituents and whether these results may have conditioned different degrees of incremental C_A shortening.

Correlation: Jaw movement × Gestural C_R-C_A overlap

By means of the data we presented in Section 3.3.1 and 3.3.2, we are able to investigate whether the degree of C_A plateau overlapped by C_R is determined by the jaw movement. To do so we calculated the Spearman's correlation coefficient between the curvature of the jaw movement (i.e. third DCT coefficient) and the coarticulatory overlap index.
Figure 3.9 shows the DCT curvature coefficients (x-axis) and the percentage of C_A plateau overlapped by C_R (y-axis). Recall that negative DCT curvature coefficients indicate a ∩-shaped jaw trajectory (i.e. the jaw raises from C_R to C_A before jaw lowering for the vowel), while DCT curvature coefficients around 0 represent a linear decrease of jaw height from C_R towards the vowel. In Figure 3.9 the x-axis ranges from negative (left end; ∩-shape) to slightly positive (right end; ∪-shape) DCT curvature indices. Since the regression line shows an upward trend from left to right, this suggests that the degree of gestural C_R-C_A overlap increases as the jaw movement throughout C_RC_AV syllables becomes more like a typical jaw cycle movement, i.e. constantly descending jaw position from the first consonant to the nucleus. The positive correlation between jaw movement curvature and C_R-C_A overlap indices ($r_s(68) = 0.43$, $p<0.001$) corroborates our Hypothesis (2) since ∩-shaped tokens show comparatively less gestural C_R-C_A overlap than tokens with linear or ∪-shaped jaw movement curvatures (i.e. jaw cycle pattern; predominantly /pla/ syllables).

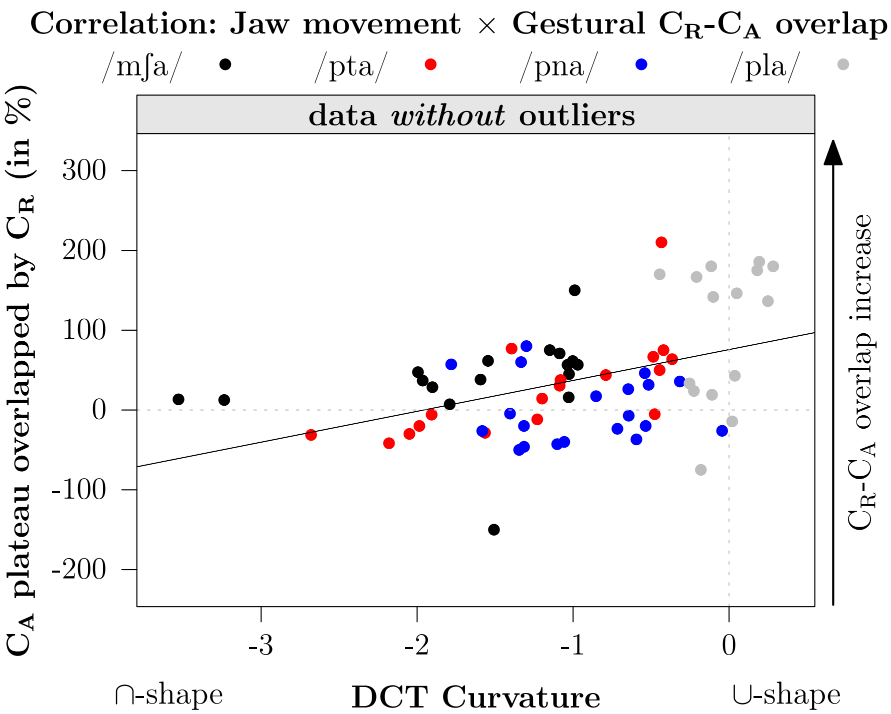

Figure 3.9: Scatterplot shows the correlation between DCT curvature and C_R-C_A overlap indices. The depicted data do not contain the outlier identified in Section 3.3.2.

Correlation: Gestural C_R-C_A overlap × Incremental C_A shortening

The indices obtained in Sections 3.3.2 and 3.3.3 allow us to investigate the hypothesized correlation between coarticulatory overlap indices and C_A duration ratios. More precisely, we expected an increasing degree of relative C_A shortening as gestural overlap of C_R and C_A increases.

Figure 3.10 shows the percentage of C_A plateau overlapped by C_R on the y-axis and the relative acoustic C_A duration values on the x-axis (i.e. Dur_{Ratio} values). Recall that negative Dur_{Ratio} values indicate that C_A is acoustically shorter in the cluster compared to the singleton condition (i.e. $C_R\underline{C_A}<\underline{C_A}$; left end), positive Dur_{Ratio} values indicate C_A is acoustically longer in the cluster compared to the singleton condition (i.e. $C_R\underline{C_A}>\underline{C_A}$; right end). Although the regression line suggests a downward trend (i.e. the degree of relative C_A shortening from singleton to cluster condition decreases with decreasing C_R-

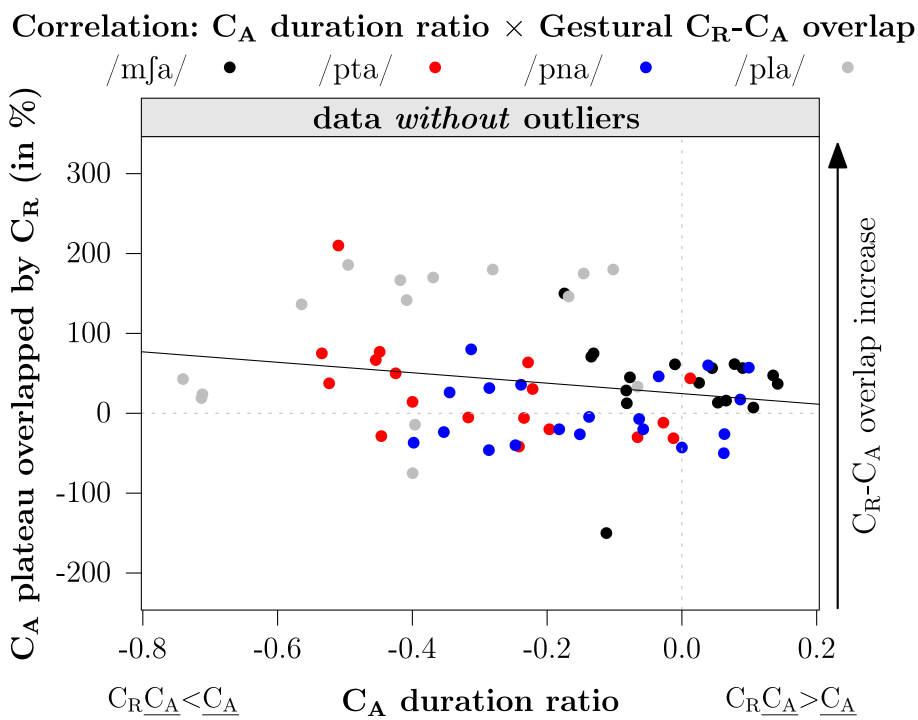

Figure 3.10: Scatterplot shows the correlation between C_R-C_A overlap and C_A's Dur_{Ratio} indices. The depicted data do not contain the outlier identified in Section 3.3.2.

C_A overlap), the Spearman correlation test had a very weak correlation coefficient and the observation was not significant ($r_s(68)$ = -0.16, p=0.18). In contrast to our Hypothesis (3), this evidences that incremental C_A shortening cannot be directly related to the degree of gestural C_R-C_A overlap.

Correlation: Jaw movement × Incremental C_A shortening

The final correlation analysis concerns the question as to whether jaw movement patterns determine C_A duration patterns. Even if our hypotheses did no concern the interaction of these two indices, it is possible that jaw movement patterns and C_A duration patterns are directly related to each other (cf. Lindblom, 2011). To take this possibility into account, we hypothesize *a posteriori* that if the jaw movement patterns follow the predictions of the jaw cycle paradigm there should be shorter C_A duration in the cluster than in the singleton condition.

Figure 3.11 shows on the x-axis the indices of the relative C_A shortening from singleton to cluster condition (Section 3.3.3) and on the y-axis the DCT curvature indices that describe the jaw movement trajectories (Section 3.3.1). As the slope of the regression line suggests, the corresponding Spearman correlation test revealed a negative correlation between the two indices ($r_s(72)$ = -0.52, p<0.001). For a C_RC_AV syllable this means that, on the one hand, if the jaw constantly descends from the syllable edge to the nucleus (i.e. DCT values of 0; jaw cycle), the duration of C_A in a cluster is shorter compared with C_A in a C_AV syllable (i.e. clearly negative C_A duration ratios). If, however, the jaw trajectory is characterized by initial raising ($C_R \rightarrow C_A$) and subsequent lowering ($C_A \rightarrow$ V), C_A is unlikely to be subject to acoustic shortening (e.g. /mʃa/ syllables).

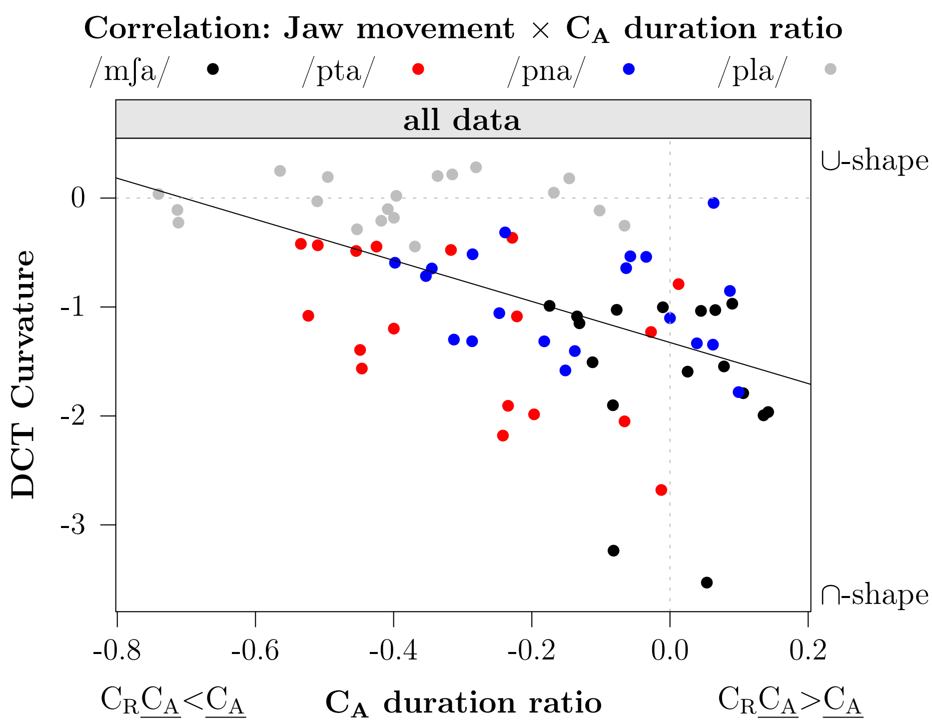

Figure 3.11: Scatterplots show the correlation between jaw movement patterns and C_A's Dur_{Ratio} indices.

3.4 Summary and Discussion

As shown in Chapter 2, syllable-related organization patterns cannot be entirely understood if syllable composition is not taken into account (see also Marin, 2013; Marin & Pouplier, 2014; Pouplier, 2012). Particularly the individual articulatory constraints of the consonants and vowels involved are significantly affecting the inter- and intra-segmental temporal patterns. In addition to the lingual constraints of the vowel-adjacent consonants investigated in previous study (Chapter 2), in the current study we systematically examined whether different degrees of articulatory overlap and incremental consonant shortening are directly attributable to jaw movement patterns. To evoke different types of jaw movements, we used a set of Polish labial+coronal clusters in which the coronal C_A varied in terms of its intrinsic jaw height and hence jaw coarticulation resistance. We measured on the one hand three independent indices (i.e. jaw movement, coarticulatory C_RC_A overlap, and incremental C_A shortening) and on the other hand assessed the interdependency between these indices to evaluate the jaw's relevance for syllable-related organization patterns. Ensuing from the basic hypothesis that mandible movements should differ with respect to cluster/syllable composition, we expected in turn that the degree of coarticulatory C_RC_A overlap and incremental C_A shortening should vary as a function of mandible movement. We proceed with a summary of the present results before continuing with the discussion of the ways in which these findings may help us to understand further articulatory correlates of syllable structure.

The idea for the present study arose from the suggestion that syllable phonotactics follow from the constraints of cyclic jaw movements (Lindblom, 1983). In the case of syllables with cluster onsets (i.e. C_RC_AV), this means that consonants with high jaw positions (e.g. coronal sibilants or stops; highly coarticulation resistant) should only occur in a position remote from the nucleus, while consonants with a lower jaw position (e.g. nasals, liquids, or glides; less coarticulation resistant) are allowed to occupy a position close to the vowel. From this perspective, /ʃm-/ would be a proper onset cluster. But what if a language's grammar also allows for syllables with reverse consonant ordering, for example /mʃ-/ onset cluster, in which the second consonant requires a comparatively higher and coarticulation resistant jaw than the initial consonant? Do the consonants' jaw positions adapt to the jaw cycle paradigm or does a jaw movement pattern emerge that is at odds with the jaw cycle paradigm? Finally, how does the presence/absence of a jaw cycle affect the hypothesized causality of jaw movement, articulatory overlap, and incremental consonant shortening (cf. Lindblom, 2011)?

The analysis of the jaw movements revealed significant differences as a function of segmental make-up of the syllables (cf. Figure 3.5 and 3.6). More precisely, only syllable /pla/ showed a continuously descending jaw position from the syllable edge to the nucleus (i.e. jaw cycle), while the remaining syllables /mʃa/, /pta/, and /pna/ exhibited higher jaw positions for the coronals compared with preceding labials (i.e. no jaw cycle). This conforms with our Hypothesis (1) that the segmental composition of the cluster and particularly the C_A's manner of articulation is the major determinant for the emergence of the jaw cycle. That is, intrinsic segment jaw positions determine the presence/absence of a jaw cycle rather than a jaw cycle determining an accommodation of jaw position. However, with reference to Figure 3.2 the present jaw movement patterns revealed an unexpected pattern for syllable /pna/: based on previous research (e.g. Hoole et al., 1990; Lindblom, 1983; Recasens, 2012a) we suggested that the jaw movement profile should be rising for /mʃa/ and /pta/ but falling for /pna/ as well as for /pla/. Unexpectedly, in our data /pna/ patterned with /mʃa/ and /pta/, but not with /pla/ which indicates that /n/ has a higher jaw position than /p/. Even more surprising is the observation that vowel-adjacent coronals /ʃ/, /t/, and /n/ make use of very similar vertical jaw positions, even if there is gradually increasing variability (cf. Figure 3.5, bottom). Since there is cross-linguistic evidence that /p/ tends to have a higher jaw than /n/ (e.g. Lindblom, 1983; Recasens, 2012a) and /n/ should have a lower jaw than /ʃ, t/ (e.g. Lindblom, 1983; Mooshammer et al., 2007; Recasens, 2012a), we evaluated whether this result could be attributed to increasing onset complexity. However, we could not confirm, for any of the vowel-adjacent coronals, that the presence of a vowel-remote labial affected C_A's jaw position, that is, labials do not exert coarticulatory force on the following coronal (i.e. jaw height: $\underline{C_A}V = C_R\underline{C_A}V$). Hence, it remains unclear what might have conditioned the unexpected high jaw position of the nasal. One explanation might be that Rocławski (1976) described Polish [n̪] as a laminal denti-alveolar nasal. It is noteworthy that Dart (1991) observed a high jaw position for laminal stops, while apical stops were articulated with a comparatively lower jaw position in order to provide enough space in the vocal tract for curling up the tongue tip. If this difference generalizes for laminal vs. apical alveolar nasals (cf. Dart, 1991; Ladefoged & Maddieson, 1996), we can assume that Polish [n̪] has an intrinsically higher jaw position than apical /n/ and possibly than /p/, it was a mistake to assume a falling jaw movement profile for the /pn/ cluster (cf. Figure 3.2).

Turning to the investigation of the inter-consonantal overlap of /mʃa/, /pta/, /pna/, and /pla/ syllables (Table 3.2), we noticed that the differences in the degree of coarticulatory overlap are in accordance with data from previous studies (cf. Bombien et al., 2013; Hoole et al., 2013): In /pla/ syllables the vowel-adjacent lateral is largely overlapped by the preceding labial, while in /pna/ the vowel-adjacent nasal achieves its constriction target at the same time as the labial gesture ends. The remaining clusters in /mʃa/ and /pta/ syllables showed an intermediate degree of C_A plateau overlapped by C_R with at least some coarticulatory overlap. Despite the gradual variability observed for /mʃa/, /pta/, and /pna/, the statistical analysis revealed no statistical difference between the corresponding overlap patterns. Since we suggested in Hypothesis (2) that the degree of coarticulatory overlap should vary as a function of absence or presence of a jaw cycle, the patterning of /mʃa, pta/ vs. /pla/ constitutes a confirmation of this assumption while /pna/ patterned unexpectedly.

The final analysis addressed the question of how the acoustic duration of the vowel-adjacent coronal changes under increasing onset complexity (i.e. $\underline{C_A}V \rightarrow C_R\underline{C_A}V$). We hypothesized that C_A duration should be temporally compressed as a function of the degree to which C_A is overlapped by C_R. In view of the coarticulatory C_RC_A overlap patterns summarized above, it was to be expected that only the lateral should be shortened from singleton to cluster condition, but not the remaining coronals /ʃ, t, n/ (cf. Hypothesis (3)). Indeed, for set **pl** the acoustic duration ratios ($\mathrm{Dur}_{\mathrm{Ratio}}$) indicate incremental /l/ shortening from singleton to cluster condition (i.e. duration: /l̲a/ > /pl̲a/). Further, the vowel-adjacent consonants in sets **mS** and **pn** did not change in duration as a function of onset complexity (i.e. duration: /n̲a/ = /pn̲a/, /ʃ̲a/ = /mʃ̲a/; cf. Syrika et al. (2011) on the preservation of sibilant duration). These findings suggest that lack of incremental nasal/sibilant shortening might be understood as a consequence of comparably small degree of coarticulatory overlap (cf. coarticulatory overlap pattern of /p̲n̲a/ and /m̲ʃ̲a/). However, in contrast to our set-specific expectations and previous research (cf. O'Shaughnessy, 1974) we obtained for set **pt** statistical evidence that the vowel-adjacent /t/ shortened acoustically as onset complexity increased (i.e. duration: /t̲a/ > /pt̲a/), and this was in spite of /pt/ having comparable overlap values to /mʃ/.

In sum, the results are predominantly in agreement with our hypotheses, even if the acoustic duration pattern observed for set **pt** prevents to some extent a straightforward allocation of the four target syllables into two separate groups, i.e. /mʃa, pta, pna/ vs. /pla/. However, regarding the last question we raised at the beginning of this summary as to whether jaw movement (Index 1), articulatory overlap (Index 2), and incremental consonant shortening (Index 3) are interrelated in a causal fashion, two generalizations can be made from the present data. If the mandible movement's curvature was consistent with the jaw cycle principle (i.e. /pla/ syllables), there was a considerable degree of C_R-C_A overlap and as a consequence incremental C_A shortening. However, if the jaw movement's curvature deviated from the typical jaw cycle movement (i.e. /mʃa/, /pta/, and /pna/ syllables), smaller degrees of C_R-C_A overlap were evident as compared to /pla/ tokens. With respect to incremental C_A shortening in /mʃa/, /pta/, and /pna/ tokens an unclear pattern remains. The statistical results of index correlations are reproduced in Table 3.4.

Due to the unclear origin of incremental shortening of the vowel-adjacent consonant, we focus in the following discussion on the evident interdependency between jaw movements (as conditioned by the jaw height and coarticulation resistance of C_A) and coarticulatory C_R-C_A overlap. By continuing the discussion of Chapter 2, we provide a proposal of how this interdependency can be embedded into the task-dynamic and gestural model (e.g.

Correlation matrix	Index 1	Index 2	Index 3
Index 1		0.43*	-0.52*
Index 2	0.43*		-0.16
Index 3	-0.52*	-0.16	

Table 3.4: Matrix of correlations between jaw movement, gestural C_R-C_A overlap, and incremental consonant shortening indices (Index 1, Index 2, and Index 3, respectively). Spearman's (r_s) coefficients marked with an asterisks indicate significant correlations.

Browman & Goldstein, 2000; Goldstein et al., 2009; Saltzman & Munhall, 1989). In addition we briefly address some implications for global syllable organization which could arise from this (e.g. Pouplier, 2012) and the emergence of transitional vowels in consonant clusters (e.g. Gafos, 2002; Gafos, Hoole, Roon, & Zeroual, 2010).

Given the present results, the question arises as to how to incorporate this observation into established theories of coarticulation/coproduction. This, however, appears to be quite difficult since theories which consider the jaw as a fundamental determinant for coarticulation are quite limited. Up to now the only theoretical account that advocates for a mandibular origin of coarticulation has been formulated by interpolating the suggestions of the Frame/Content theory (e.g. MacNeilage, 1998): Lindblom (2011) hypothesized that the coarticulatory overlap of neighboring speech gestures should increase with increasing segmental density within the opening movement of the jaw (i.e. from syllable onset to nucleus). In the same vein, Redford (2004) suggested for obstruent+sonorant clusters that the vowel-adjacent sonorant is produced during a jaw lowering movement and thus temporally (and articulatorily) compromised due to coproduction with its phonetic surrounding. From this perspective, if we take Lindblom's (1983) jaw opening/sonority hierarchy into account, some of the syllables used in this study were admittedly unlikely to show a jaw cycle pattern or – as hypothesized by Lindblom (2011) and Redford (2004) – substantial

$C_R C_A$ overlap: for /mʃa/ and /pta/, the vowel-adjacent /ʃ, t/ are cross-linguistically known for having a higher and rather invariant jaw target than the respective vowel-remote /m, p/ (e.g. Hoole et al., 1990; Lindblom, 1983; Recasens, 2012b). However, even more critical for a mandibular-driven account of coarticulatory overlap is that the syllables with initial obstruent+sonorant clusters did not show consistent jaw movement and coarticulation patterns as hypothesized by Lindblom (2011) and Redford (2004): only /pla/ showed the proposed patterns, but not /pna/, although both sonorants were expected to have a lower jaw position compared with the preceding obstruent. Since the aforementioned accounts are reliant on the presence of a typical jaw cycle pattern, this challenges the predictive power of the respective approaches that the origin of coarticulation is to be found in the oscillating movements of the jaw.

In contrast to the Frame/Content theory, the jaw plays only a minor role in the Articulatory Phonology framework, which incorporates the task dynamic and the gestural model (e.g. Browman & Goldstein, 1990, 2000; Saltzman & Munhall, 1989). As we pointed out in Chapter 2, Articulatory Phonology defines the phonological representation of a single speech sound in terms of spatiotemporal coordination of discrete vocal tract actions, i.e. articulatory gestures (cf. Browman & Goldstein, 1992). Each gesture is performed by one of the following *constriction organs* (cf. Nam et al., 2012): lips, tongue tip, tongue body, velum, and glottis. The absence of the jaw from this list makes clear that the jaw is not part of the phonological representation of vowels/consonants, but the jaw serves in a cooperative synergistic manner together with other articulators to facilitate the constriction organ (e.g. the tongue tip) to achieve the intended gestural target. To give an example: the attainment of a targeted constriction location or degree of a tongue tip gesture (e.g. for a coronal consonant: TTCL and TTCD) requires three articulators to fulfill this articulatory task: the tongue tip, tongue body, and the jaw (cf. Browman & Goldstein, 1990, 1992). The hierarchical differentiation between articulator and gestural level reveals the potential incompatibility of the Articulatory Phonology framework with the mandibular-driven account of coarticulation: In the gestural model of syllable organization the spatiotemporal overlap of neighboring gestures (i.e. coarticulation; cf. Fowler & Saltzman, 1993; Saltzman

& Munhall, 1989; Öhman, 1966) are defined in terms of coupling relations between gestures and *not* between articulators. As pointed out by Browman and Goldstein (2000; see also Goldstein & Fowler, 2003; Goldstein et al., 2009) variation in gestural overlap is assumed to emerge from variation in the underlying gestural coupling or bonding strength (see also Chapter 2 for a discussion of altering C-V coupling strength). Conclusively, since relative coupling of successive speech sounds is integrated at the gestural level, at least in theory the jaw cannot control for the spatiotemporal coordination of speech from the articulator level.

So, if neither the Frame/Content (e.g. MacNeilage, 1998; MacNeilage et al., 2000) nor the gestural/task-dynamic model (e.g. Browman & Goldstein, 2000; Fowler & Saltzman, 1993; Saltzman & Munhall, 1989) are capable to directly implement the mandibular-driven variability of coarticulation patterns, where does the apparent evidence come from that the degree of coarticulatory overlap is correlated with the biomechanical constraints of speech? We believe it might be possible to integrate this result into the gestural model – at least in an indirect fashion.

As we pointed out in the previous chapter, it is possible within the gestural model to fine-tune the intergestural timing of neighboring speech sounds (to adjust temporal overlap; e.g. Browman & Goldstein, 2000; Goldstein & Fowler, 2003; Goldstein et al., 2009) as well as the gestural blending (to adjust the degree of coarticulation; e.g. Iskarous et al., 2012; Pastätter & Pouplier, 2014). In fact, applying the TADA synthesizer (Nam et al., 2004) we were able to show that different degrees of C_AV overlap in a C_RC_AV sequence can be achieved by manipulating the pair-wise gestural coupling relations of C_A-V and C_R-V (see Section 2.4). Recall that in the gestural model, the C_RC_AV syllable timing is specified in terms of in-phase (C_A-V and C_R-V) and anti-phase (C_R-C_A) relations, which results in a temporal C-center organization of the onset cluster relative to the vowel if default settings are applied (e.g. Browman & Goldstein, 2000). To model onset-vowel timing patterns deviating from the C-center pattern, we took advantage of the fact that the respective coupling relations can be manipulated in terms of coupling strength (cf. Section 2.4 and Goldstein

& Fowler, 2003; Goldstein et al., 2009). This was the basis for our assumption that each onset consonant could be specified for a characteristic coupling strength value that should correspond to the individual degree of lingual coarticulation resistance. Accordingly, by assigning a relatively greater coupling strength to C_A = /s/ compared to C_R = /p/, we were able to model a /psɔ/ syllable with near-synchronous initiation of the sibilant and the vowel gestures, as would be expected for singleton onset /sɔ/ syllables (see for more details Section 2.4 of the previous chapter). Given this result of manipulating the strength of in-phase coupling relations, we hypothesize by implication that different degrees of coarticulatory C_R-C_A overlap could also be modeled by adjusting the consonantal coupling strength values in C_R-C_A anti-phase coupling relationships.

A first indication of how the C_R-C_A coupling strength values should be specified was found in Browman and Goldstein (2000). They proposed that the strength of C_R-C_A coupling relations must be greater than the strength of C_R-V/C_A-V phase relations to prevent synchronization of the onset consonants due to individual in-phase coordination with the vowel. It should be mentioned at this point that TADA simulations with default anti-phase specification (i.e. C_R-C_A coupling strength α=1.0) produced substantial degrees of C_R-C_A overlap, i.e. in terms of the overlap measure applied in this study: 100% of C_A plateau was overlapped by C_R. Accordingly, to model the articulatory C_R-C_A overlap patterns observed in this study, it is not reasonable to simply adopt the arbitrarily determined coupling strength values of the example in the previous chapter (i.e. α=0.8 for the coarticulation resistant sibilant and α=0.2 for the less resistant labial; cf. Section 2.4) since this would result in a more synchronized initiation of C_R and C_A gestures, i.e. degree of C_R-C_A overlap: > 100%. To model C_R-C_A overlap patterns lower than 100% (cf. Table 3.2: /mʃa/ = 36.9%; /pta/ = 27.4%; /pna/ = 0.5%), C_R and C_A should be specified with higher coupling strength values in C_R-C_A coupling relation (i.e. $\alpha > 1$).

In current TADA simulations, we took this into account and manipulated the C_R-C_A coupling relations to model the reduced degree of C_R-C_A overlap observed, for example, for /<u>m</u>ʃa/. In this example, the cluster-initial labial (/m/) was arbitrarily specified with strength value of α=1.2, while the vowel-adjacent sibilant (/ʃ/) had a comparatively higher strength value of α=1.8, in order to model the different degrees of coarticulation resistance. The simulation output confirmed that the gestural overlap in the C_R-C_A output was considerably reduced compared to the default strength settings. To express the reduced C_R-C_A overlap in terms of our overlap measure: 57.1% of the synthesized C_A=/ʃ/ plateau was overlapped by C_R=/m/. Even if the synthesized overlap pattern does not match the C_R-C_A overlap pattern we actually found, this demonstrates that intra-cluster timing can be modeled with respect to consonantal coarticulation resistance. Given the fact that the strength values used above were arbitrarily determined, the (still) open question is how the jaw could contribute towards modeling even more appropriate C_R-C_A overlap patterns?

We believe this could be achieved if a coalition of various scales of coarticulation resistance (e.g. jaw and tongue body) were to be used to fine-tune the anti-phase coupling strengths. This hypothesis originates from Articulatory Phonology's notion that consonant gestures involve a number of individual articulators which act in a synergistic manner to attain a gestural target, i.e. tongue tip gesture: tongue tip, tongue body, and jaw; labial gesture: upper lip, lower lip, and jaw (cf. Browman & Goldstein, 1990, 1992). The implementation of the task dynamics component in TADA allows for each gesture to assign for each articulator involved an individual weight value that aims to characterize the articulator's degrees of freedom in the functional synergy. Furthermore, Iskarous et al. (2013) argued that these weights also reflect the articulator's propensity for context-dependent variability (i.e. coarticulation). In sum, the articulator weights of a /ʃ/'s tongue tip gesture are substantially higher than those of /m/'s lip gesture (compare, for instance, jaw weight of /ʃ/ = 512 vs. /m/ = 8, which is indicative of substantially less coarticulatory variability in the sibilant than the labial; cf. Lindblom (1983); Mooshammer et al. (2007); Recasens (2012b)). As we pointed out earlier in this discussion the influence of the articulator weight parameter is, however, limited to the functional synergy within a single gesture and has *no* effect on the inter-gestural coordination. Hence, we formulate the hypothesis that the weights of all articulators could be used in a joint fashion to specify the coupling strength values of both in-phase (C_R-V and C_A-V) and anti-phase (C_R-C_A) coupling relations. This constitutes a reasonable expansion of the account we introduced in Section 2.4 which only considered the degree of lingual coarticulation resistance in the specification of coupling strength. For the two examples discussed above (/psɔ/ and /mʃa/) this would imply that a sibilant-specific strength value is composed of the individual degrees of coarticulation resistance of tongue tip, tongue body, and the jaw, while a labial-specific strength value is composed of the individual degrees of coarticulation resistance of upper and lower lips, and the jaw. Generally higher articulatory constraints of the sibilant compared with the labial (expressed in articulator weights) should result in comparatively higher coupling strength for the sibilant than for the labial. In terms of syllable organization patterns within labial-sibilant-vowel sequences, this type of coupling strength asymmetry would condition two concomitant timing patterns: first, the vowel-adjacent sibilant should start near-synchronously with the vowel; second, the vowel-remote labial and the vowel-adjacent sibilant should expose moderate C_R-C_A overlap (cf. TADA coupling manipulations earlier in this discussion and in the Discussion of Chapter 2). If these two timing patterns emerge simultaneously, this could imply that there is very limited temporal overlap of the vowel-remote labial and the vowel (cf. Figure 2.1). As a consequence, since the vocal tract is open and not actively constrained by any articulator for the period between C_R's constriction release and C_A's constriction attainment, this could make for an acoustic open transition (e.g. Catford, 1988; Gafos, 2002; Gafos et al., 2010). Indeed, in a number of instances the acoustic data from /mʃa/ target words showed for a number of instances a vocoid portion between the voiced/voiceless consonants. The idea that the occurrence of open transitions can be predicted by C_R/C_A's coupling strength values would speak for a phonetic rather than phonological status of open transitions (see e.g. Kirby, 2014 for a discussion on this topic).

Finally, we discuss the implication of how sonorants' articulator weights may contribute to gestural cohesion. Taking the different scales of consonantal coarticulation resistance into account (jaw: Lindblom, 1983; Mooshammer et al., 2007; Recasens, 2012b; tongue body: Recasens & Espinosa, 2005, 2009), the propensity for coarticulation is generally considered to be least for /ʃ/ and comparatively higher for /n, l/, i.e. sonorants are more compatible with the articulatory gestures of the neighboring segments and allow thus for more contextual overlap (cf. Fowler & Brancazio, 2000). Therefore we suggest a lower coupling strength for /n, l/ compared to /ʃ/ relative to the preceding labial and the following vowel, while the coupling strength values of labials should not vary as a function of the following coronal. Hence, the reduced coupling strength asymmetry of /pna, pla/ (as compared to /mʃa/) is expected to condition both a considerable overlap of the sonorant and the vowel (C_A-V) and the sonorant and the labial (C_R-C_A). These alleged timing patterns have been empirically shown at least for /pla/ syllables, i.e. the lateral overlapped with the vowel to a larger extent in the cluster condition than in the corresponding singleton /la/ condition (previous study; cf. Figure A.1), and the tongue tip gesture of the lateral was substantially overlapped by the labial gesture (present chapter). On the other hand, /pna/ showed a relative increase in C_AV overlap compared with the singleton /na/ condition (previous chapter; cf. Figure A.1), but the coarticulatory overlap measure investigated in this study confirmed that C_A plateau was on average not overlapped by C_R. This pattern is however consistent with previous studies (e.g. Bombien et al., 2013; Kühnert et al., 2006) which suggested that stop+nasal clusters show generally less gestural overlap in order to prevent that the perceptually crucial stop release and the lowering of the velum for the nasal come into a temporal conflict with each other. From this point of view the respective aerodynamic requirements of stop+nasal clusters seem to constitute an additional constraining factor towards intracluster coordination (just as tongue/jaw coarticulation resistance) which should also be considered in terms of coupling relations (and their weights) of the tongue tip and the velum at the gestural level (cf. Hoole et al., 2013).

In sum, we assume that articulator weights reflect the degrees of coarticulation resistance of the individual articulators (cf. Iskarous et al., 2013) and could therefore be used to determine in a predictive way the strength values of in-phase (C_R-V, C_A-V) and anti-phase (C_R-C_A) coupling relations alike. Accordingly, we conclude that the vowel-adjacent consonant's degree of coarticulation resistance may be predictive for the degree of C_A-V and C_R-C_A overlap. However, this theoretical conclusion is not compatible with Marin and Pouplier's (2014) observation that the presence or absence of a C-center organization is not reflected by large or small degree of C_R-C_A overlap (see also Pouplier (2012) and Marin

(2013)). This theoretical and empircal discrepancy will have to be addressed in future research.

To conclude, this study provided new insights into the biomechanical basis of different syllable coordination topologies. We showed that different patterns of jaw movements may have different effects on intra-cluster overlap and – as a consequence – on the degree of incremental shortening of the vowel-adjacent consonant. This led to the conclusion that the jaw cycle is merely an epiphenomenon of individual consonant and vowel productions rather than a determinant factor of speech. However, we expanded the theoretical virtue of coupling strength parameters in order to predict syllable-related timing patterns. More precisely, we proposed that different degrees of C_R-C_A overlap could be deduced from articulator weights which are hypothesized to reflect the respective articulator's propensity for context-dependent variability (e.g. tongue body and jaw coarticulation resistance). Alternatively, we assume that previous scales of articulator's coarticulation or invariance patterns could help for further investigation (cf. Recasens, 2012b; Recasens & Espinosa, 2009; see also Iskarous et al. (2013) for an enhanced account of Recasens' DAC values) since it is not yet clear how to transform the articulator weights or degrees of coarticulation resistance into appropriate coupling strength values.

Chapter 4

Investigating Conflicting Aerodynamic Requirements in CC Clusters

4.1 Introduction

The preceding chapters investigated the principled biomechanical constraints of the tongue and the jaw on the temporal organization of syllables with cluster onsets (i.e. $\mathrm{C_RC_AV}$ syllables). Both studies confirmed that degree of $\mathrm{C_AV}$ overlap (Chapter 2) and $\mathrm{C_RC_A}$ overlap (Chapter 3) is predicted by the propensity of the vowel-adjacent consonant ($\mathrm{C_A}$) to coarticulate with its phonetic neighbors. According to our expectations in Chapter 3, the results confirmed qualitatively less gestural overlap in /mʃ, pt/ than in /pl/ onsets, but unexpectedly the degree of consonantal overlap in /pn/ was found to be remarkably smaller to that in /pl/ (cf. Table 3.2). Even if our hypothesis did not predict different degrees of gestural overlap for stop+nasal and stop+lateral clusters, this pattern has been previously observed and discussed in terms of consonantal aerodynamic requirements (cf. Bombien et al., 2013; Kühnert et al., 2006). Since this particular overlap difference (/pl/ > /pn/) cannot be explained in terms of purely biomechanical constraints, in the final experimental chapter we investigate whether labial+alveolar onset clusters with conflicting aerodynamic requirements prevent increasing gestural overlap as a function of deaccentuation.

4.1.1 Empirical background on intra-cluster coordination

Previous research on intra-cluster gestural timing (i.e. the timing between consonants in a cluster) comprises various types of consonant clusters occurring in a variety of languages (e.g. English: Byrd and Choi (2010); French: Hoole, Bombien, Kühnert, and Mooshammer (2009); Kühnert et al. (2006); Georgian: Chitoran and Cohn (2009); Chitoran et al. (2002); German: Bombien et al. (2013); Hoole et al. (2009); Pouplier (2012); Modern Greek: Yip (2013); Moroccan Arabic: Gafos et al. (2010); Romanian: Marin (2013, 2014); Marin and Pouplier (2014); Russian: Marin, Pouplier, and Kochetov (2015)). Even if the different temporal patterns of intra-cluster coordination seem to be highly intricate, a few individual factors have emerged as being highly predictive of how closely the consonant gestures are articulated: for instance, the ordering of C_R and C_A place of articulation (e.g. Byrd, 1996; Chitoran et al., 2002; Kühnert et al., 2006; Wright, 1996; Yip, 2013; Zsiga, 1994) and the manner of articulation of both the first (C_R manner: e.g. Byrd, 1996; Kühnert et al., 2006; Yip, 2013) and the second consonant (C_A manner: e.g. Chitoran & Cohn, 2009; Kühnert et al., 2006; Yip, 2013). In light of these few factors, it appears that intra-cluster timing and hence the extent of gestural overlap within clusters is determined by their segmental composition.

As outlined above, one of the main effects influencing the timing of two successive consonantal gestures is the so-called *place order* effect on inter-consonantal overlap (i.e. C_R and C_A overlap). First evidence for this particular effect of cluster composition emerged from investigations of stop+stop sequences (e.g. Chitoran et al. (2002) plus subsequent studies by Chitoran and Cohn (2009); Gafos et al. (2010) on those types of clusters) which revealed more gestural overlap in front-to-back clusters (e.g. /pt, dg/, where anterior place of C_R articulation precedes a posterior place of C_A articulation) compared with reversely ordered clusters (e.g. /gd/; back-to-front clusters). These systematic gestural overlap differences have been interpreted as arising from the tug of war between perceptual recoverability, which reflects the needs of the listener, and parallel transmission (Mattingly, 1981), which arises from the need of the speaker to efficiently modulate articulation. Thus, on the one hand, parallel transmission requirements favor a greater degree of overlap, all other considerations being equal. On the other hand, perceptual recoverability requirements favor less overlap in certain configurations. Thus, it is likely in a front-to-back cluster (e.g. /gd/) that the essential perceptual cues of the velar stop /g/ (i.e. its burst release) would be masked by a more anterior alveolar constriction (/d/) overlapping it too much. Hence, in order to preserve the acoustic cues necessary for stop perception, speakers are compelled to reduce the gestural overlap in back-to-front clusters. In contrast, front-to-back clusters (e.g./dg/) are believed to allow comparatively greater overlap, since parallel transmission of acoustic information (Mattingly, 1981) is possible despite of the alveolar stop burst being superimposed by simultaneously occurring constriction in the velar region (cf. Chitoran et al., 2002). Beyond stop+stop sequences, the place order effect has also been attested for stop+sonorant, stop+fricative, and fricative+sonorant onset clusters (e.g. Kühnert et al., 2006; Pouplier, 2012; Yip, 2013) since perceptual masking should not be an issue in, for example, stop+sonorant clusters. The existence of the effect beyond stop+stop clusters challenges the assumption that perceptual recoverability is the crucial factor underlying this effect (cf. Kühnert et al., 2006; Yip, 2013). With reference to larger degrees of overlap in /pl/ than /kl/, and /pn/ than /kn/, Kühnert et al. argued against a purely perceptual account for this pattern, since for example the burst of the velar is recoverable even if a lateral or nasal encroaches it, and proposed instead an articulatory account for this finding: in contrast to the /kC/ clusters, extensive overlap within /pC/ clusters may have emerged since the lips and the tongue can execute their articulatory tasks with large overlap in the temporal domain without interfering with the other's gestural movement and – to a lesser extent – acoustic cues. On the one hand, at least for some cluster types (e.g. stop+sonorant clusters) this suggestion implies a weakening of the place order hypothesis, but, on the other hand, this articulatory account highlights the importance of manner of C_R and C_A articulation for intra-cluster timing.

An extensive body of research proposes that the temporal coordination of onset clusters varies as a function of C_R and C_A manner of articulation (e.g. Bombien et al., 2013, 2010; Hoole et al., 2009, 2013; Kühnert et al., 2006; Marin, 2014; Pouplier, 2012; Yip, 2013). The few systematic investigations into different manners of C_R articulation revealed that consonant clusters with a stop in C_R position allow for relatively less gestural overlap compared with consonant clusters with a fricative in C_R position (Kühnert et al., 2006; Yip, 2013). This observation can presumably be ascribed to the fact that the release burst of stop C_R has to be preserved due to its importance for perceptual recoverability (see discussion above). In contrast, fricatives are known to be acoustically salient and persistent; hence, the fricative C_R might allow partial overlap by C_A since this temporal interference would not cause the fricative to lose its perceptual cues (Henke, Kaisse, & Wright, 2012; Wright, 2004). To extend our knowledge to the effect of manner of C_R articulation, we investigate in this study a set of Polish stop and nasal initial onset cluster (i.e. C_R = /p/ vs. C_R = /m/). In line with the previously observed difference between fricative and stop initial clusters, we would expect that **mC** clusters should allow for more intra-cluster overlap than **pC** clusters, since acoustic cues for nasal perception should remain despite temporal overlap by the following consonant.

More studies have addressed the question of how C_A manner of articulation affects its timing relative to the preceding consonant (C_R). The systematic variation of different manners of alveolar consonants in th C_A position has been particularly revealing. Recent studies of Bombien et al. (2013) and Yip (2013) showed that /pl/ and /ps/ clusters have similar patterns of C_RC_A overlap in German and Greek, while Marin (2014) observed more C_RC_A overlap for /pl/ than for /ps/ in Romanian. In addition, stop+nasal sequences such as /pn/ have been shown to be coordinated further apart (i.e. with less articulatory overlap) compared to /pl/ (Hoole et al., 2009; Kühnert et al., 2006; cf. also Table 3.2 in the previous chapter for Polish). The fact that the /pn/ cluster repeatedly showed less articulatory overlap than the /pl/ counterpart has guided the discussions of Hoole and colleagues towards the individual aerodynamic requirements of the involved consonants (cf. also Vallée, Rossato, & Rousset, 2009): while the labial stop requires a closed velum in order to build up intraoral air pressure, the subsequent nasal (unlike the lateral) necessitates complete oral occlusion along with lowering of the velum. The release of the stop /p/ must occur before the velar lowering of /n/ is initiated, since otherwise nasal venting during stop closure would cause a loss of acoustic cues which are indispensable for successful stop perception, cf. Hoole et al.'s (2013) articulatory (TADA: Nam et al., 2004) and aerodynamic (HLSyn: Hanson & Stevens, 2002) modeling of stop+nasal clusters. To prevent this aerodynamic conflict, the cluster constituents overlap to a lesser extent in /pn/ than in /pl/ clusters. Since the lateral consonant requires neither a complete oral occlusion nor a lowering of the velum, the lateral may be produced during labial closure without interfering with the necessary aerodynamics.

It is also noteworthy that the overlap difference regarding C_A=/n/ vs. C_A=/l/ may also hold if C_R varies in terms of place of articulation: for instance, more overlap has been observed for /kl/ than in /kn/ onset clusters in German and French (cf. Bombien et al., 2013, 2010; Hoole et al., 2009; Kühnert et al., 2006), but no such difference was observed for /kl/ and /kn/ in Romanian (Marin, 2014). This was argued to be due to Romanian /kl/ patterning in terms of lag values with the German /kn/ rather than with German /kl/, which means that aerodynamic requirements do not automatically predict a difference between stop+/l/ and stop+/n/ clusters in a language where the consonants in general are timed farther apart from each other. Using /pl/, /pn/, and /pʃ/ onset clusters, we examine whether the C_A manner effect found for German (and French) generalizes cross-linguistically to Polish, a language generally assumed to have a different overlap pattern from German, with clusters being generally less overlapped (cf. Pouplier and Beňuš's (2011) suggestion that the range of consonant clusters permitted in a particular language interacts with the inter-consonantal overlap typology; that is, the larger the cluster inventory, the less the degree of intra-cluster overlap).

4.1.2 Conflicting aerodynamic requirements

The availability of **pC** (/pl/, /pn/, /pʃ/) and **mC** (/ml/, /mɲ/, and /mʃ/) clusters in Polish allows us to elicit potentially conflicting aerodynamic requirements determined by manner of both C_R and C_A, and we do so by means of prosodic variation. Thus far, the effect of prosodic variation on the articulatory coordination of onset clusters is quite underrepresented in the literature, but has recently gained increasing attention (e.g. Bombien et al., 2013; Byrd & Choi, 2010; Peters, 2015). Although, onset cluster overlap tends to be greater in prosodically weak (or deaccented) than in strong (or accented) positions, there is at the same time evidence for cluster-specific differences. That is, in German, the timing of /kn/ onset clusters turned out to be relatively less affected at prosodically weaker boundaries than /kl/ onsets (e.g. Bombien et al., 2013; Hoole et al., 2009). This pattern might suggest an interaction between prosodic variation and aerodynamic requirements on the intra-cluster timing. By taking this observation into account, we argue that two consonants with presumably conflicting aerodynamic requirements are resistant to prosodic variation, and hence undergo less/no increasing gestural overlap under deaccentuation than clusters with no such aerodynamic conflicts.

To test this hypothesis, we test two groups of onset clusters (**pC** and **mC**), each with one particular cluster expected to exhibit potentially conflicting aerodynamic requirements:

pC For /pn/, the C_R stop requires a sealed vocal tract so that intraoral air pressure can build up, while the C_A nasal needs a concomitant oral occlusion and a lowered velum (see above). If deaccentuation were to increase the degree of C_RC_A overlap beyond a certain level, the stop release would be obscured by premature nasal leakage that would be caused by the aerodynamic requirements of the nasal. The requirements of the stop and of the nasal are therefore at odds with each other and incompatible therefore with a great degree of overlap between them.

In contrast, for clusters /pl, pʃ/ the stop's perception is not expected to be compromised if /p/ is released into the following /l/ or /ʃ/: the incomplete linguo-palatal occlusion of /l, ʃ/ should preserve enough aerodynamic requirements of /p/, even if C_RC_A overlap increases under deaccentuation.

mC In /mʃ/, the C_R nasal necessitates a combination of labial occlusion and velar lowering and the C_A sibilant requires a precise jet of air striking an obstacle to produce high frequency noise. If deaccentuation were to increase the degree of C_RC_A overlap, this would increase the conflict between the aerodynamic requirements of the two consonants and – as a result – nasal leakage during the sibilant's production would impair the aerodynamic turbulence (characteristic for sibilants) due to insufficient intraoral air pressure.

In contrast, if inter-consonantal overlap increases under deaccentuation, the remaining clusters /ml, mɲ/ are assumed to render appropriate aerodynamic requirements to preserve the individual acoustic cues. Most obviously, since the second consonant in /mɲ/ is also a nasal, it would not suffer from nasal leakage. For /ml/, laterality cues are expected to retain despite of nasalization and reduced intraoral pressure.

4.1.3 Hypotheses

Altogether, it is assumed that manifold factors influence the articulatory timing of onset C_RC_A clusters. The present experiment aims to bring together various influencing factors within one single study. Therefore, we only investigate labial+alveolar consonant clusters (i.e. front-to-back typology) to invoke comparatively greater consonantal overlap as compared to back-to-front clusters (cf. Byrd, 1996; Chitoran et al., 2002; Kühnert et al., 2006). In addition, consonants in both positions vary as a function of manner of articulation, while the place of articulation is held constant: first, the vowel-remote consonant is either a stop or a nasal labial (C_R = /p, m/) to test for different perceptual recoverability constraints; and second, the vowel-adjacent consonant is either an alveolar sibilant, lateral, or nasal (C_A = /ʃ, l, n/) to elicit different aerodynamic requirements. Depending on how C_R and C_A consonants are clustered, it is expected that aerodynamic requirements of C_R and C_A may interfere with each other, particularly in the case of increasing gestural overlap under deaccentuation. Hence, we vary the prosodic conditions in which the consonant clusters occur (accented vs. deaccented position).

The strong hypothesis of this study is that conflicting aerodynamic requirements are expected to block increasing consonantal overlap in deaccented compared with accented position. We anticipate for both cluster groups (i.e. **pC** and **mC** clusters) an interaction between manner of articulation of the vowel-adjacent consonant (i.e. C_A Manner) and prosodic condition (i.e. Accent; accented vs. deaccented). For **pC** clusters, there should be an interaction between prosodic condition and manner of articulation since /pn/ is expected to avoid increasing overlap under deaccentuation; otherwise, premature lowering of the velum would hamper the acoustic cues of the stop's release burst. For **mC** clusters, there should be an interaction between prosodic condition and manner of articulation since /mʃ/ is expected to avoid increasing overlap under deaccentuation; otherwise, nasal leakage during the sibilant's production would cause a partial loss of the sibilant's high frequency noise due to insufficient intraoral air pressure. In sum, these clusters – /pn/ and /mʃ/ – should not change the degree of consonantal overlap under deaccentuation to preserve the acoustic cues necessary for consonant perception.

In light of the suggestions of Pouplier and Beňuš (2011), and of Marin (2014), and taking into account the observation that Polish may have a different overlap pattern than Germanic languages, we also think it is pertinent to formulate a weaker hypothesis. Namely, that the aerodynamic requirements may not be crucial if the clusters involved exhibit quite large timing lags between the consonants to begin with. In this case, the different aerodynamic requirements may be reflected in an overall intra-cluster timing difference between /pn/ vs. /pl, pʃ/, and /mʃ/ vs. /mn, ml/ but not in the prosodic variation.

4.2 Method

4.2.1 Speech material

In this study, we examine the temporal organization of a variety of labial+alveolar consonant sequences in word-initial position of disyllabic Polish target words (Table 4.1). To elicit different prosodic conditions, target words occurred in either accented or deaccented phrasal position. In addition, word-level stress fell on the word-initial syllable that contains the consonant clusters under investigation (cf. Gussmann, 2007). In order to investigate whether the interaction between prosodic variation and aerodynamic requirements can be generalized beyond stop-initial clusters, we assembled two groups of onset clusters with either C_R = /p/ (i.e. **pC**) or C_R = /m/ (i.e. **mC**) to test the perceptual recoverability constraint. Further, in relation to the question of whether the manner of the vowel-adjacent consonant's articulation interacts with prosodic condition, we varied the vowel-adjacent consonant in terms of articulatory laterality (/l/), nasality (/n/), and sibilance (/ʃ/).

Cluster target words					
pC			**mC**		
/**pl**/	[**pl**at͡som]	*placom*	/**ml**/	[**ml**ɛkax]	*mlekach*
/**pn**/	[**pn**at͡ʂ̻i]	*pnaci*	/**mɲ**/	[**mɲ**ixom][1]	*mnichom*
/**pʃ**/	[**pʃ**ɛraʑ̻]	*przeraź*	/**mʃ**/	[**mʃ**alik]	*mszalik*

Table 4.1: Target words with labial+alveolar consonant clusters in word-initial position as a function of vowel-remote /p/ (**pC**) and /m/ (**mC**), respectively. Those clusters which are expected to comprise potentially conflicting aerodynamic requirements are highlighted in gray.

4.2.2 Recording and data processing

The articulography data used in the study were collected during the same recording session introduced in Chapter 2. Therefore, we refer to Section 2.2.1 for further information on speakers, recording procedure, and post-processing and to Section 3.2.1 for more details on the algorithmic identification of consonantal gestures. However, in contrast to Chapter 3, the data from all Polish participants are usable since the measurements applied here (see

[1]Regarding the second onset consonant in [**mɲ**ixom], we follow the phonemic description of Jassem (2003) and Nowak (2006, citing G. Stone, 1990) who assigned an alveolo-palatal place of articulation to the nasal /ɲ/, but note that Gussmann, 2007; Sussex & Cubberley, 2006 assumed an exclusively palatalized articulation in the dorsal region of the vocal tract for /ɲ/.

below) are not sensitive to the emphatic pauses prior the accented target words produced by one female speaker.

To achieve consonant cluster productions under accentuation vs. deaccentuation, participants were asked to read carrier phrases with embedded target words from a display. The examples below show how attention was drawn to the word that had to be articulated with a phrasal accent (i.e. either the target word (here: *placom*) or the phrase initial word (here: *Jakub*)). In this sense, the word intended to be accented was denoted with red and underlined font on the subject's display, while the respective target word appeared within single quotes. Whether or not the participants mastered the reading task was auditorily evaluated on the basis of pitch accent position, yet no acoustic measures were applied.

accented: **Jakub powtarza 'placom' aktualnie.**
(Jakub repeats 'placom' currently.)

deaccented: **Jakub powtarza 'placom' aktualnie.**
(Jakub repeats 'placom' currently.)

Given a targeted data set of n = 288 (6 words × 2 prosodic conditions × 6 speakers × 4 repetitions) we had to exclude 13 utterances due to unambiguously wrong prosodic productions, occasional misreadings and technical difficulties during the recording session, leaving 275 items for analysis.

4.2.3 Measurements

In line with previous research (e.g. Bombien et al., 2013; Kühnert et al., 2006; Pouplier, 2012), we assess the effect of segmental composition on consonant cluster productions in terms of how the cluster constituents are timed with each other. Since our hypotheses concern the preservation of acoustic cues (e.g. the stop's release burst) despite of increasing gestural overlap under deaccentuation, we assessed the relative consonantal timing as the temporal lag between $\mathrm{C_R}$ target release (NOFFS) and $\mathrm{C_A}$ target attainment (NONS), henceforth referred to as plateau lag (Figure 4.1). In terms of this particular measure, larger plateau lag values indicate that the articulatory release and target of C_R and C_A are timed further apart (i.e. they have low overlap), while small plateau lag values denote comparatively tighter consonant coordination.

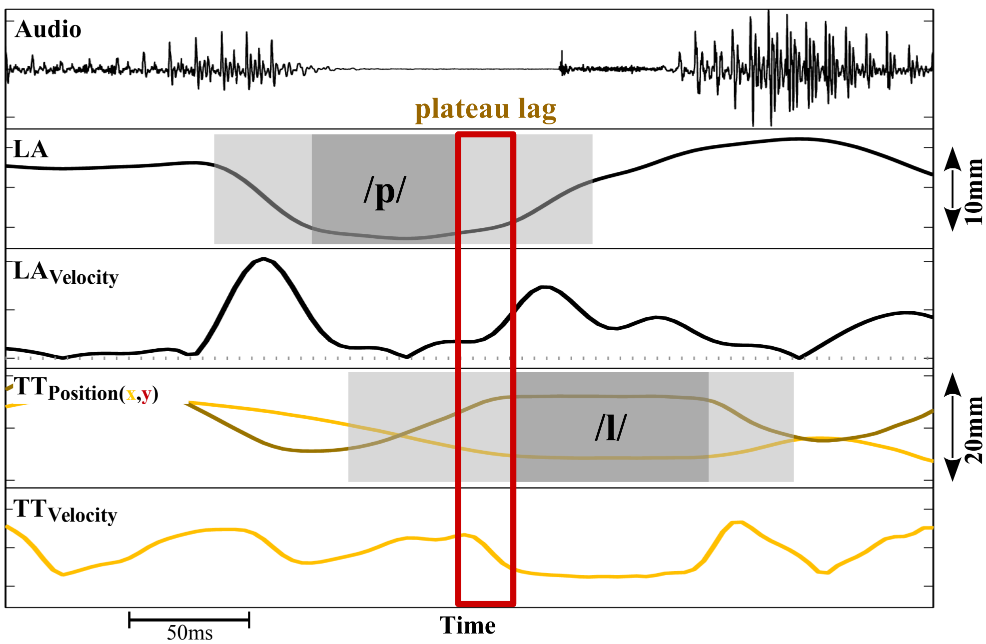

Figure 4.1: For a given instance of target word [plat͡som], this plot shows for /p/ and /l/ gestures the movement and velocity profiles for lip aperture (LA) and tongue tip (TT), respectively. For both gestures, the dark gray period indicates the constriction plateau, while the light gray periods prior to and following the plateau indicate the interval where the gesture waxes and wanes, respectively. The red box indicate the temporal lag between C_R's constriction release and C_A's constriction attainment (i.e. plateau lag). In this particular example the plateau lag amounts 24 ms.

In reference to our hypotheses, we expect that prosodic variation may generally be a conditioning factor on the temporal organization of word-initial clusters towards tighter coordination patterns (i.e. smaller plateau lags or more overlap) in prosodically weak (deaccentuation) compared to strong (accentuation) positions (cf. Bombien et al., 2013). However, this effect should be blocked or limited if aerodynamic requirements of both C_R and

C_A would be endangered by increasing gestural overlap. By implication, increasing gestural overlap is only hypothesized for onset clusters in which no conflicting aerodynamic requirements are present.

4.2.4 Statistical analyses

Using the R environment (R Core Team, 2013), we calculated linear mixed models (*lme4* package: Bates, 2010; Bates et al., 2015) in order to statistically examine the main effects of Manner of C_R and C_A as well as their interaction with the prosodic condition (Accent) on the dependent variable Plateau Lag. To test the effect of C_A Manner (and its interaction with Accent), we run two separate mixed models for **pC** and **mC** clusters. Speakers were a random factor in all analysis, and Cluster was a random factor in the model testing for C_R Manner and Accent. To obtain *p*-values, we compared one model with and one without the fixed factors and interactions of interest. Post-hoc analyses were carried out using package Multcomp (Hothorn et al., 2008) to perform pairwise comparisons.

4.3 Results

The central question we address is whether increasing gestural overlap in onset clusters is avoided in deaccented conditions for those clusters exhibiting conflicting aerodynamic requirements. Since both groups of labial+alveolar sequences (**pC**={/pl/, /pn/, /pʃ/} vs. **mC**={/ml/, /mɲ/, /mʃ/}) contain clusters with potential conflicting aerodynamic requirements, this would necessitate a three-way interaction analysis (factors: $C_R \times C_A \times$ Accent) to answer the question above. However, as a total corpus size of only 275 data points is presumably not sufficient for a three-way interaction analysis, we performed separate statistical analysis for the C_R and C_A fixed factors. Furthermore, for the C_A factor, we performed separate analyses for **pC** and **mC** clusters, respectively. A second reason for the separate treatment of the data was that we were also interested in whether the interaction of prosodic variation and aerodynamic requirements can be generalized beyond /p/-initial clusters. Recall that cluster-initial /p/ requires that the burst be successfully perceived as an acoustic cue, while this is not the case for /m/; however, cluster-initial /m/ might itself prevent **mC** overlap, since a lowered velum might endanger the acoustic properties of second consonant (C_A). Therefore, this section includes first a report on global differences between **pC** and **mC** clusters, followed by the consideration of group-specific interaction patterns between C_A and Accent.

Our analyses in this section concern the intra-cluster organization in terms of temporal lag measures between the target release of C_R and target attainment of C_A. By these means we are able to assess whether conflicting aerodynamic requirements predict a particular gestural lag pattern, and furthermore if they impose less temporal flexibility under prosodic variation. First, we tested the general differences between C_R=/p/ (**pC**) and C_R=/m/ (**mC**) clusters, and specifically the interaction between C_R manner of articulation and prosodic condition (i.e. C_R Manner × Accent) on Plateau Lag values (random factors: Subject and Cluster) in order to investigate whether prosodic variation affects equally **pC** and **mC** clusters. Table 4.2 shows the C_R results in terms of mean lag values. Although **mC** exhibits generally smaller lag values (i.e. more overlap) compared to **pC** clusters, the effect of prosodic alteration on both groups is similar, that is, smaller lag numbers for **pC**/**mC** clusters in the deaccented than in accented condition. In terms of statistics, the main effect of Manner (/p/ vs. /m/) was not statistically significant, but there was a significant main effect of Accent ($\chi^2[1]$=63.1, p<0.001). The statistical model showed no significant interaction between C_R Manner and Accent ($\chi^2[1]$=1.03, p=0.31); that is, the gestural organization within **pC** and **mC** clusters behaves similarly under deaccentuation.

The next analyses concern the cluster-specific patterns of gestural overlap as a function of C_A manner of articulation (C_A Manner) and as a function of prosodic variation (Accent), separately for **pC** and **mC** clusters. In line with our hypotheses, we expect an interaction between C_A Manner × Accent on Plateau Lag values for both group of clusters. Figure 4.2 shows the results of these analyses as a function of group, C_A manner of articulation,

	Accent		
C_R Manner	accented		deaccented
pC	56.36 *(23.85)*	>	39.20 *(27.39)*
mC	44.64 *(22.71)*	>	24.34 *(20.69)*

Table 4.2: Table assembles averaged plateau lag values (in ms) as a function of C_R Manner and Accent. Standard deviation values (*SD*) are given in parentheses.

and prosodic condition: the larger the plateau lag numbers on the y-axis, the larger the temporal lag between the constriction plateaus of C_R and C_A; that is, the degree of gestural overlap decreases with increasing Plateau Lag values.

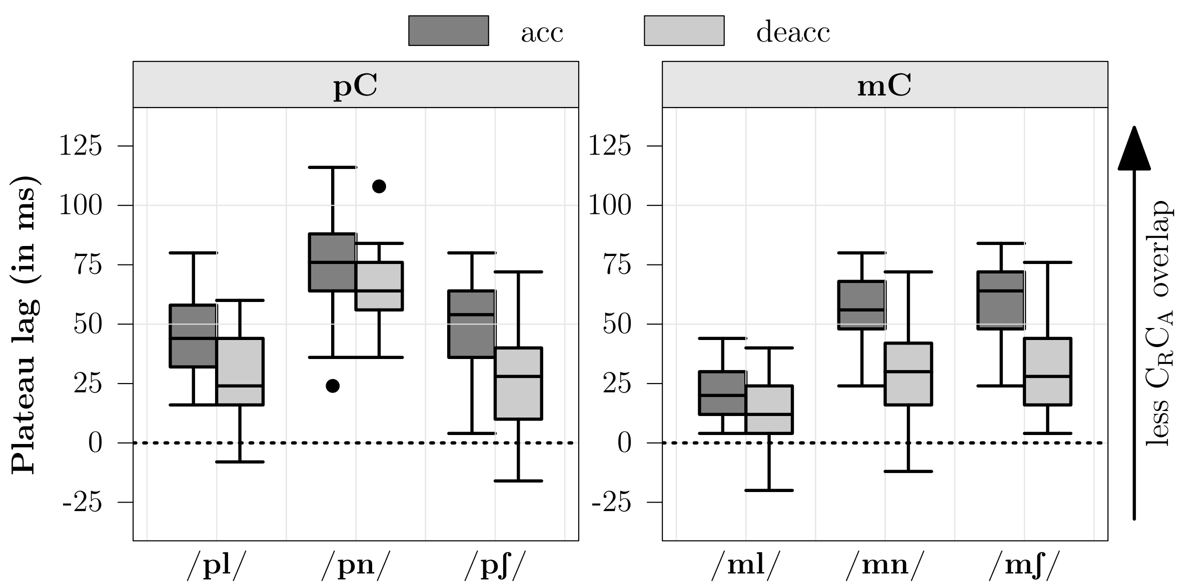

Figure 4.2: Plateau lags (in ms) are given for **pC** (left) and **mC** (right) with respect to C_A and prosodic condition.

Regarding **pC** clusters, we expected an interaction between C_A Manner and Accent on Plateau Lag. According to previous findings (e.g. Bombien et al., 2013), this interaction should be carried by /pn/, that is, in contrast to /pl, pʃ/, /pn/ was not expected to exhibit different overlap patterns in accented and deaccented sentence position. In contrast to our hypothesis, the mixed model carried out on this set of data (**pC**) revealed no significant interaction, while both factors separately treated had a significant effect on Plateau Lag (C_A Manner: $\chi^2[2]=61.8$, $p<0.001$; Accent: $\chi^2[1]=15.0$, $p<0.001$). However, as Figure 4.2 (left) suggests, our expectation concerning /pn/ receives qualitative confirmation, in that onset /pn/ shows comparatively less variation as a function of Accent than /pl, pʃ/. Moreover, cluster /pn/ exhibits the largest plateau lags and this indeed may be due to the conflicting aerodynamic constraints present for cluster /pn/ but not for /pl/ or /pʃ/. Statistical confirmation for this observation comes from post-hoc Tukey tests for the main effect of C_A Manner (/l/ vs. /n/ vs. /ʃ/) which reveal a overall plateau lag difference between /pn/ vs. /pl/ ($p<0.001$) and /pn/ vs. /pʃ/ ($p<0.001$), but not between /pl/ vs. /pʃ/. Thus, while **pC** clusters do not confirm our hypothesis in its strong form, they do it in its weaker form though.

Regarding **mC** clusters, we hypothesized again a statistically significant interaction between C_A Manner and Accent. For this group of nasal-initial clusters, we assumed for /ml, mɲ/ that aerodynamic requirements may allow for prosodically induced increasing overlap. For /mʃ/ onsets, however, we anticipated that growing gestural overlap could result in conflicting aerodynamic requirements. Hence, we expected that /mʃ/ should be timed further apart in both prosodic conditions since otherwise the lowered velum for /m/ would cause a lack of turbulence during /ʃ/ due to insufficient intraoral air pressure. The mixed model carried out in analogy to the previous analysis revealed the predicted C_A Manner $\times$ Accent interaction effect ($\chi^2[2]=11.3$, $p<0.01$) with both main effects turning out significant (C_A Manner: $\chi^2[2]=49.5$, $p<0.001$; Accent: $\chi^2[1]=30.4$, $p<0.001$). This means that the degree of relative Plateau Lag changes (accented vs. deaccented) differed between **mC** clusters. However, the plateau lag patterns in Figure 4.2 (right) disclose clearly that the interaction was due to a lack of prosodic variation for /ml/ instead of /mʃ/. This implies that the already high degree of consonantal overlap apparent for /ml/ is not sensitive to prosodic alteration, while /mɲ, mʃ/ show distinctly more gestural overlap in deaccented than in accented condition. Post-hoc Tukey tests reveal that /mɲ, mʃ/ plateau lags decreased with deaccentuation (both: p<0.001; i.e. increasing $C_R C_A$ overlap), but /ml/ overlap remained unchanged under prosodic variation.

4.4 Summary and Discussion

Despite numerous cross-linguistic investigations into consonant clusters, the complexity of interacting factors responsible for different intra-cluster timing patterns is not entirely understood. Previous research suggests that different $C_R C_A$ timing typologies may emerge from cluster composition, reflecting perceptual and aerodynamic requirements of the consonants involved. Prosodic alternations have also been shown to have an effect on consonant timing, but the interaction between these factors, namely aerodynamics and prosody has not been examined yet. Therefore, we systematically investigated whether onset clusters with potentially conflicting aerodynamic requirements prevent an increasing gestural overlap under deaccentuation. The measurements (i.e. temporal lag between C_R target release and C_A target attainment) applied to **pC** and **mC** onset clusters unveiled overlap patterns which are only partly in agreement with our hypotheses. Hence we discuss now how these result may be integrated into the complex interactions of language and/or cluster-specific effects determining $C_R C_A$ overlap. We begin by summarizing the current results.

Although **pC** and **mC** onset clusters showed tendentially different overlap patterns in both prosodic conditions (accented vs. deaccented), with less $C_R C_A$ overlap for **pC** than for **mC** clusters (cf. Table 4.2) this result was not robust statistically. So although we hypothesized different perceptual recoverability constraints for /p/ vs. /m/, these constraints did not effect a statistically different timing pattern in our data. In contrast to the marginal C_R manner effect, the intra-cluster organization differed significantly as a function of prosodic alteration, that is, the extent of inter-consonantal overlap increased if **pC**/**mC** clusters occurred in deaccented than in accented position (cf. Table 4.2). This confirms an overall effect of prosodic variation on the degree of intra-cluster overlap (cf. Bombien et al., 2013; Byrd & Choi, 2010; Peters, 2015). As to C_A effects, we present them separately for **pC** and **mC** clusters.

pC clusters exhibited different plateau lag values as a function of **pC** cluster composition; that is, /pn/ cluster revealed larger plateau lag numbers compared to /pl/, while /pʃ/ and /pl/ clusters exhibited virtually equal plateau lag numbers. With respect to C_A manner this relative overlap pattern conforms with previous results (i.e. C_RC_A overlap /pl/ > /pn/: Hoole et al., 2009; Kühnert et al., 2006; /pl/ $\ngtr$ /pʃ/: cf. Bombien et al., 2013; Yip, 2013). In addition, the different degrees of gestural overlap between /pl, pʃ/ > /pn/ persist across prosodic conditions. Thus although the different aerodynamic requirements did not block prosodic-determined timing changes in either of the clusters, the aerodynamic conflicting /pn/ exhibited much larger timing lags between consonants, confirming thus the aerodynamic hypothesis in its weaker instantiation.

As to **mC** timing patterns smaller plateau lag values (i.e. more C_RC_A overlap) were observed for /ml/ compared to /mɲ, mʃ/. Also in contrast to /mɲ, mʃ/, the degree of gestural overlap in /ml/ did not change under deaccentuation. Since we expected the degree of /mʃ/ overlap to remain constant under deaccentuation (due to potential aerodynamic conflicts), this result contradicts our hypothesis concerning **mC** clusters. Indeed, we assumed that if deaccentuation would cause a /mʃ/ overlap increase, then velar lowering during sibilant's constriction plateau would cause a loss of aerodynamic characteristics and concomitantly a shortening of the steady-state frication. But note that sibilants require a certain duration of stationary noise for successful place of articulation perception (Hughes & Halle, 1956; Jongman, 1989). A subsequent analysis of acoustic sibilant durations revealed, however, that the prosodically conditioned overlap increase did not cause the amount of steady-state frication to fall below a critical threshold of 30 – 50 ms (cf. Hughes & Halle, 1956; Jongman, 1989). Considering that the mean frication portions were of 140 ms in accented /mʃ̲/ and of 118 ms in deaccented /mʃ̲/, our assumption for /mʃ/ clusters appears to have been too strong since essentially much more overlap should have occurred to cause a loss of acoustic sibilant cues. In addition, we have not anticipated that /ml/'s degree of consonantal overlap would be insensitive to prosodic variation. However, given the already large degree of C_RC_A overlap in accented /ml/ clusters, it is conceivable that perceptual recoverability constraints might have conditioned a floor effect, which prevented a further overlap increase under deaccentuation.

In sum, the present results only partly support the notion that conflicting aerodynamic requirements predict gestural overlap differences in onset clusters. We assumed for /mʃ/ and /pn/ onset clusters that possible aerodynamic requirements would block a prosodically conditioned change in intra-cluster lag. However, this was not the case. For cluster /mʃ/, it appears that the lags observed in both prosodic conditions were large enough to allow for large enough steady state frication intervals (over 100 ms in both conditions, well above the critical 30 – 50 ms proposed in the literature). Cluster /pn/ on the other hand provided evidence for the role of aerodynamic constraints in shaping intra-cluster lag, albeit in a different manner than the one we strongly hypothesized. Thus, /pn/ was the cluster that exhibited significantly larger lags than any other cluster even in the deaccented condition. From Figure 4.2 we can observe that /pn/ in deaccented condition had larger lags than /pl, pʃ, mn, mʃ/ in accented condition. The large lags in accented condition presumably afforded /pn/ to exhibit a prosodically conditioned lag decrease and still maintain its aerodynamic requirements. From this two conclusions can be drawn: first, the concept of conflicting aerodynamic requirements is more relevant for **pC** than for **mC** clusters; second, the conclusion that can be drawn is that the presence of potential aerodynamic conflicts does not necessarily constitute a restraint of a cluster to increase its overlap in the deaccented compared to the accented condition as long as the cluster is not very overlapped to begin with. Altogether, our results make it clear that complex interactions of factors have to be considered at the planning and/or executing stages of speech production. This implies that speakers have obvious knowledge about the articulatory constraints of speech gestures and how they have to be coordinated in larger units of speech (e.g. consonant clusters or syllables), so that essential acoustic information is provided to the listener for successful consonant perception. The Articulatory Phonology framework (Browman & Goldstein, 1990, 1992) and the incorporated gestural coupling model (e.g. Browman & Goldstein, 2000; Goldstein et al., 2009; Nam et al., 2009) have recently been assumed to be capable to implement of the intricate interactions of segmental composition (varying place/manner of C_R and C_A articulation) and aerodynamic requirements (cf. Bombien et al., 2013; Hoole et al., 2013). The gestural coupling model defines the intergestural timing in terms of coupling relations; that means that in a C_RC_AV sequence, both consonants are coupled in-phase with the vowel (in-phase: C_R-V and C_A-V) while the consonants are coupled anti-phase to each other (anti-phase: C_R-C_A). The default specification of sequentially (i.e. anti-phase) coupled C_R and C_A gestures causes the cluster consonants to be coordinated with a certain amount of plateau lag. But if stop+nasal clusters are modeled by means of TADA (Nam et al., 2004) the default coupling relations between C_R (e.g. /p/) and C_A (e.g. /n/), this yields inappropriate acoustic characteristics of the stops release burst despite a certain amount of plateau lag (Hoole et al., 2013). Since the acoustic output of TADA simulations is ultimately generated by the pseudo-articulatory synthesizer HLSyn (Hanson & Stevens, 2002), Hoole and colleagues examined the aerodynamic profiles of the simulated stop+nasal clusters and could demonstrate that intraoral pressure declined prematurely during the stop constriction and before the actual stop release. This was attributed to aerodynamic impairment due to nasal leakage. To circumvent the conflicting aerodynamic requirements occurring as a consequence of gestural overlap of the stop and the following nasal, Hoole et al. (2013) manually adjusted the C_R-C_A phasing to result in a larger temporal lag of the corresponding gestures. And indeed, they obtained thus more appropriate stop cues since the intraoral pressure remained at a high level throughout the stop's occlusion and declined only after the stop's release. Although the modeling work of Hoole et al. proposes that different degrees of gestural overlap in onset clusters can be straightforwardly achieved by varying the phasing of the consecutive consonantal gestures, it inherently challenges one of the core properties of the gestural coupling model, that is, that phasing relations should be independent of the gestures involved.

In order to revise this restriction, recent studies discussed the capability of alternative coupling relations between cluster consonants (i.e. C_R-C_A phasing) to model systematically occurring timing differences. These attempts rely on the assumption that consonant gestures consist of two individually controlled gestures which sequentially execute the closure

(CLO) and the release (REL) movement of a given consonant (Browman, 1994). Accordingly, this split-gesture account makes two coupling nodes accessible for each consonant gesture which could be addressed in terms of gestural coupling (Nam, 2007). Goldstein et al. (2009) were the first to employ the split-gesture account and could show that, all else being equal, the degree of C_R-C_A overlap varies according to which part of the C_R gesture (i.e. CLO or REL) is underlyingly linked with the vowel. Thus, a substantially lower C_R-C_A overlap was observed, if the release (REL) instead of the closure (CLO) gesture of C_R was coupled with the vowel. Based on this finding, Goldstein et al. (2009) concluded that such coupling differences can also be used to understand the emergence of systematically different gestural overlap patterns as repeatedly reported for stop+/n/ vs. stop+/l/ clusters. Relatedly, Bombien (2011) mentioned another coupling topology to model the evidently smaller degree of C_RC_A overlap for /kn/ compared to /kl/ – namely, a relative coupling of the closure gesture of C_A (/n/) to the release gesture of C_R (/k/) in order to reduce the degree of C_RC_A overlap. One question is how to motivate, from an articulatory and/or phonological perspective, a revision of the gestural coupling model as suggested by Goldstein et al. (2009) and Bombien (2011). Note that the gestural compositions of /kn/ and /kl/ (as well as of /pn/ and /pl/) are to the greatest extent identical; only the active lowering of the velum differentiates the gestural scores of /pn, kn/ clusters from /pl, kl/, which makes the former more complex in terms of articulatory coordination compared to the latter. Accounting for such effects will require an extension of the basic model of gestural coordination as originally proposed by Browman and Goldstein (1989).

The observation that oral-velum timing differs as a function of prosodic position (Krakow, 1989, 1993, 1999) has played a fundamental role in the gestural model of syllable structure (Browman & Goldstein, 1989, 1995). Yet which specific coupling topologies may underly this well-known prosodically conditioned variation is far from clear. Recent studies (e.g. Byrd et al., 2009) replicated and extended Krakow's findings that nasals in syllable onsets (i.e. /#nV/) showed near-synchrony of velum lowering and tongue tip raising, while tongue tip raising occurred later than velum lowering in syllable coda nasals (i.e./#Vn/). Byrd et al. (2009) suggested two possible coupling topologies in order to specify the underlying coordination patterns of /#nV/ sequences: either, the gestures associated with /n/ (i.e. tongue tip and velum gesture) are coupled in-phase with each other, while only the tongue tip gesture is directly linked in-phase to the vowel's tongue body gesture (cf. Figure 4.3 left); or, all involved gestures of this sequence (i.e. the tongue tip, the velum, and the vowel's tongue body gesture) are coupled in-phase in all pairwise combinations (cf. Figure 4.3 right). As the authors stated both variants result in a close tongue-velum coordination of the nasal initiated simultaneously with the vowel.

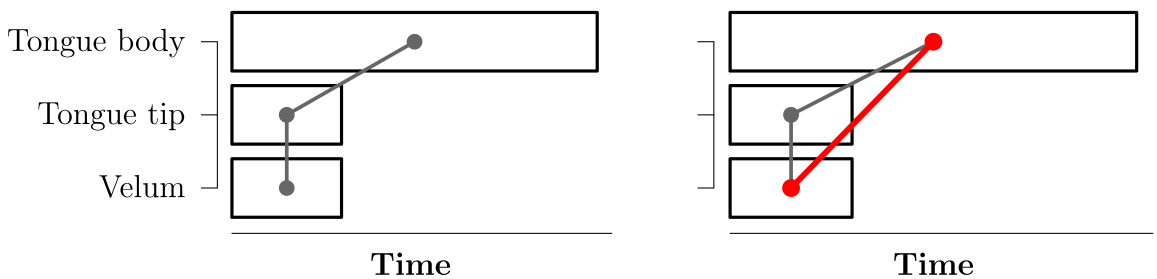

Figure 4.3: TADA simulations of the two possible coupling graphs discussed in Byrd et al. (2009) in order to model syllable-onset nasal+vowel sequences.

Although Byrd et al.'s study does not include clusters containing nasal stops, the hypothesized coupling topologies can nonetheless be relevant for the timing of /pn/ (or other stop+/n/) onset clusters. If, for instance, the second topology is used for the underlying coupling relations of /pn/, a multiply-linked structure emerges which is strongly reminiscent of the one for English /pl/ onsets suggested by Goldstein et al. (2009). Accordingly, since both the nasal's tongue tip and velum gesture are coupled in-phase the vowel's tongue body gesture, this should result in a roughly synchronous coordination of the vowel-adjacent nasal and the vowel. The TADA simulations in Figure 4.4 qualitatively confirms this approach. The top panel shows the gestural score of /pna/ synthesized with the default C_RC_A coupling topology which conditions C_R and C_A to symmetrically shift away and towards the vowel, respectively. Proceeding from this topology, the introduction of an additional in-phase link between the nasal's velum and the vowel's tongue body gesture (i.e. the red solid connector in the mid panel of Figure 4.4) results in an asymmetric shift pattern. That is, the shift of the labial gestures (lips and glottis) is comparatively larger than the shift of

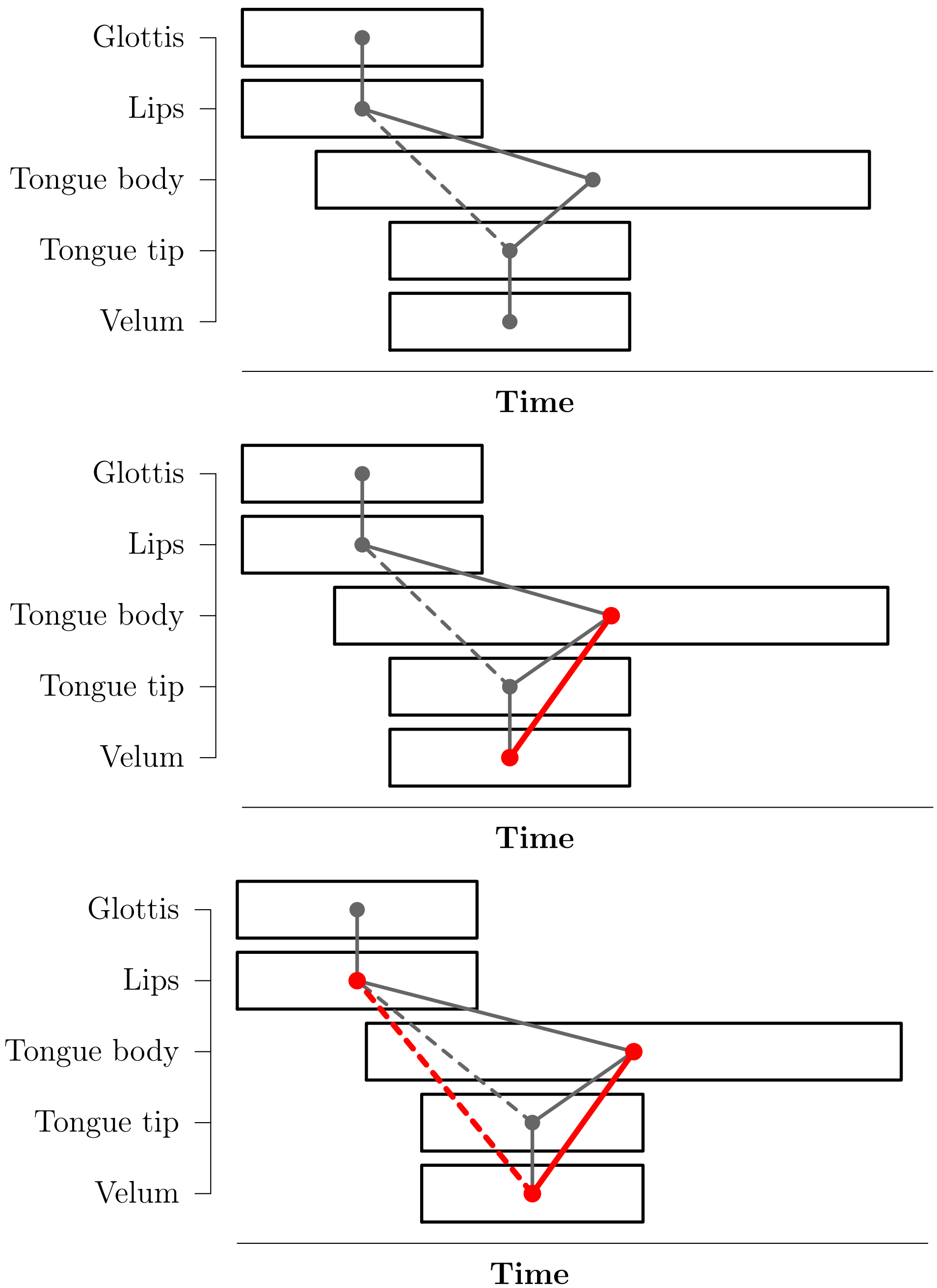

Figure 4.4: TADA simulation outputs of /pna/ syllables based on simply- (top panel) and multiply-linked coupling topologies (mid and bottom panel). Solid and dashed connectors represent in-phase and antiphase coupling relation modes, respectively. (The gestural scores at hand are simplified for illustration purposes, e.g. the labial's velum ist not actived here.)

nasal gestures (tongue tip and velum) which involves the nasal to be initiated more closely with vowel's tongue body gesture than they do in the top panel. However, since the present multiply-linked structure only affects the temporal coordination of the consonants relative to the vowel but not between consonants, it might be reasonable to further introduce an anti-phase coupling between the nasal's velum and the labial's lip gesture to constrain a low overlap pattern. As indicated in the bottom panel of Figure 4.4, the effect of such a coupling relation (i.e. the red dashed connector) is twofold. First, the gestures associated with /p/ and /n/ are coordinated further apart than in the upper gestural scores. Such a topology is apparently effective to prevent aerodynamic conflicts (i.e. premature lowering of the velum during the stop's closure; cf. Hoole et al., 2013). Second, the shift asymmetry grows further so that the nasal is initiated even closer to the onset of the vowel's tongue body gesture than in the mid panel (due to a later start of the vowel). This suggests that considering multiply-linked onset structures may help to model individual overlap patterns for different types of onset clusters (Goldstein et al., 2009), particularly for clusters with potential conflicting aerodynamic requirements.

To conclude, the results of this chapter do not support our hypothesis that intra-clusters timing patterns should not be affected by deaccentuation if conflicting aerodynamic requirements are present. From a theoretical point of view, they however partially supported our hypothesis that aerodynamics shape intra-cluster timing in that /pn/, a cluster with aerodynamic competing requirements between its consonants, exhibited a timing in which consonants were timed further apart from each other compared to all other clusters analyzed. But /mʃ/ patterned against our predictions based on aerodynamic factors. Whether aerodynamic factors have a higher relevance for stop-initial clusters specifically will have to be addressed in future research. We also observed that in general base-line intra-cluster timing may determine whether prosodic conditioned timing differences are observed or are blocked: thus, the very large timing lags of an aerodynamic conflicting cluster such as /pn/ did not block prosodic timing differences, while the very short timing lags of /ml/ blocked further overlap increase, although its consonant members' perceptual and aerodynamic requirements are not conflicting. We discussed how the repeatedly shown difference between stop+/n/ vs. stop+/l/ clusters could be modeled within the gestural coupling model, with a special emphasis on the importance of velum-oral organization. The suggestions made here should receive further attention in future modeling work.

Chapter 5

Discussion and conclusion

The present dissertation examined the role of biomechanical (lingual: Chapter 2; mandibular: Chapter 3) as well as aerodynamic (Chapter 4) constraints of particular consonants and consonant combinations influencing the spatiotemporal organization of onset clusters in relation to the following vowel as well as the timing of the consonants themselves within the cluster. This chapter reviews the central observations of the three experimental chapters of this thesis, summarizing the spatiotemporal organization patterns that emerged by varying, respectively, the lingual coarticulatory resistance of the vowel-adjacent consonant (Chapter 2), the jaw coarticulatory resistance and intrinsic jaw height of the vowel-adjacent consonant (Chapter 3), and the aerodynamic requirements of the consonants composing an onset cluster (Chapter 4).

5.1 Biomechanical constraints on onset-vowel and intra-cluster timing

Previous research on articulatory correlates of syllable structure suggests that the temporal organization of syllables varies as a function of the segmental make-up of the onset cluster and particularly the identity of the consonant adjacent to the vowel (i.e. C_A). Since there is currently little understanding of the details conditioning this variation, we hypothesized that lingual and jaw coarticulatory resistance of C_A are among the factors possibly conditioning the variation of syllable timing.

The clusters used in Chapter 2 vary systematically in terms of lingual (i.e. tongue body) coarticulatory resistance of C_A in C_RC_AV clusters. The continuum of C_A coarticulatory resistance ranges from most likely to resist (group *sibilants*: /s, ʃ/) to least likely (group *labials*: /m, p/); group *alveolar sonorants* /n, l/ was categorized at an intermediary level of coarticulatory resistance. We started from the proposal (Browman & Goldstein, 1988, 2000) that typical onsets exhibit a so-called C-center organization, whereby as onset complexity increases, the vowel-adjacent consonant increases its overlap with the following vowel, and whereby the center of the onset maintains a constant timing to the vowel regardless of onset complexity. Based on previous qualitative observations (Marin, 2013), we hypothesized that this increasing overlap with the vowel in cluster vs. singleton onsets may be conditioned by the degree of coarticulatory resistance of C_A. In order to determine the effect of C_A coarticulatory resistance on the relative change of onset-vowel overlap from singleton (C_AV) to cluster (C_RC_AV) condition, we employed two measurements as indices of onset-vowel overlap change: 1) a temporal lag measure which compares the timing differences between the vowel-adjacent consonant and a constant anchor point in a singleton and cluster condition ($\underline{C_AV}$ vs. $C_R\underline{C_AV}$), and 2) a spatial measure which determines the degree of contextual variability of the vowel's tongue body position conditioned by C_R. As expected, the temporal and spatial analyses confirmed that the degree of onset-vowel organization varied as a function of lingual coarticulatory resistance of C_A, showing relatively more C_AV overlap change for group *labials* compared to group *sibilants*, and an intermediate pattern for group *alveolar sonorants*. Since onset-vowel overlap gradually increased as lingual coarticulatory resistance of the C_A decreased, this demonstrates that lingual coarticulatory resistance (in terms of tongue body control) is one of the presumed factors conditioning the different patterns in onset-vowel timing.

Methodologically similar, Chapter 3 systematically varied jaw height and jaw coarticulatory resistance of the vowel-adjacent (C_A) consonant in order to unveil the interrelation between jaw movement, coarticulatory overlap of C_R and C_A, and incremental C_A shortening. We hypothesized that intra-cluster gestural overlap and – as a consequence – the degree of incremental shortening of C_A is determined by the vertical jaw movement throughout /a#$\underline{C_R C_A a}$/ syllables where $C_R C_A$ was either /mʃ/, /pt/, /pn/, or /pl/. Since the intrinsic jaw height of C_A was hypothesized to decrease in the progression /ʃ/ > /t/ > /n/ > /l/, we expected that the probability for a 'typical' jaw cycle movement (i.e. a constantly descending jaw position from C_R towards the syllable nucleus) increases with decreasing intrinsic jaw height of the vowel-adjacent consonant. That is, we expected that /pl/ would show a more 'typical' jaw cycle movement than /mʃ/. We observed that jaw movement curvature qualitatively differed as a function of intrinsic jaw height of the vowel-adjacent consonant; statistically /mʃa/, /pta/, and /pna/ showed similar patterns (i.e. the jaw raised first from C_R to C_A before lowering from C_A to V), while /pla/ showed a 'typical' jaw cycle pattern. Correlation tests revealed that the degree of C_R-C_A overlap depended on jaw height of C_A, i.e. syllables in which C_A exerted a lower jaw position than C_R (i.e. in syllables where a continuous opening of the jaw was evident), the onset consonants overlapped to a greater extent than in syllables in which C_A exerted a higher jaw position compared to C_R (i.e. in syllables with a cap-shaped jaw movement). Further analyses confirmed that incremental C_A shortening (i.e. C_A was acoustically shorter in the cluster than in singleton condition) was negatively correlated with jaw movement patterns. This means that the degree of incremental C_A shortening decreased with increasing intrinsic jaw height of C_A. These findings, together with the suggestion that jaw height is directly related to jaw coarticulatory resistance (e.g. Lindblom, 1983; Mooshammer et al., 2007; Recasens, 2012b), indicate that jaw coarticulatory resistance is generally capable of conditioning different degrees of C_R-C_A overlap, i.e. variability in intra-cluster timing.

5.2 Aerodynamic constraints on intra-cluster timing

In addition to the biomechanical factors conditioning different intra-cluster timing patterns, previous research reported that $C_R C_A$ coarticulation depends on cluster composition, as well as on prosodic alternations (e.g. cluster consonants overlap to a greater extent in unaccented than in accented position). For instance, stop+/n/ clusters were repeatedly shown to exhibit less inter-consonantal overlap than stop+/l, s/ clusters (cf. Bombien et al., 2013; Kühnert et al., 2006). This has been hypothesized to be conditioned by the aerodynamic incompatibility of stops and nasals, since an early lowering of the velum would interfere with the acoustic stop burst (cf. Bombien et al., 2013; Hoole et al., 2013; Kühnert et al., 2006). Based on these findings, we hypothesized in Chapter 4 that clusters with conflicting aerodynamic requirements (/pn/ and /mʃ/) should show less C_RC_A overlap and a decreased temporal flexibility under prosodic variation compared to clusters without pronounced aerodynamic conflict (/pl, pʃ/ and /mn, ml/). For /pn/ we observed qualitatively less variation as a function of prosody than for /pl, pʃ/, and an overall greater temporal lag than for all other clusters. However, against our expectations we found for /mʃ/ a clearly increased C_RC_A overlap in unaccented position in spite of the potential aerodynamic conflict. These findings suggest that conflicting aerodynamic requirements do not alone predict C_RC_A timing. Instead, it is more likely that complex interactions of various constraints (e.g. tongue body/jaw coarticulatory resistance, as well as perceptual constraints discussed but not tested here) have to be considered at the planning and/or executing stages of speech production.

5.3 Theoretical implications

This dissertation concerned syllable organization from two perspectives, i.e. onset-vowel (Chapter 2) and intra-cluster timing (Chapter 3 and 4). The final remarks concern the inclusion of the individual results in order to extend and improve the gestural model of syllable organization.

Based on observed systematic differences in onset-vowel and intra-cluster timing as a function of C_A's tongue body and jaw coarticulatory resistance, respectively, we argued that coarticulatory resistance should be comprised in the gestural model in terms of the coupling strength parameter. On the basis of TADA simulations we proposed that systematic coupling strength manipulations allow for modeling different degrees of spatiotemporal overlap of gestures in the case of competing phase relations. Thus far, all coupling strength settings are by default specified to be equal in TADA, i.e. in $\#C_RC_AV$ sequences the strength of C_R-V and C_A-V (i.e. in-phase) and C_R-C_A (i.e. anti-phase) coupling relations are α=1). This particular coupling strength setting conditions C_R and C_A to symmetrically shift away and towards the vowel respectively (i.e. C-center organization), with considerable overlap between the onset consonants. However, to model the coarticulatory-resistance-conditioned onset-vowel and intra-cluster timing differences, we suggested instead that consonants should be specified with context-independent coupling strength values derived from the hypothesized degree of tongue body/jaw coarticulatory resistance, i.e. consonants should be specified with increasing coupling strength as tongue body/jaw coarticulatory resistance increases. By means of TADA simulations in which different coupling strength values were applied to C_R and C_A we could show that the relative strength of C_R and C_A is a major determining parameter for the extent to which C_R and C_A overlap with the vowel (i.e. onset-vowel organization) *and* with each other (i.e. intra-cluster organization). In this context it is noteworthy that the degree of consonantal coupling strength has inverse implications on in-phase and anti-phase coupling relations. Thus, increasing C_A coupling strength in a C_A-V in-phase relationship causes increasing bonding of the consonant and the vowel (i.e. increasing synchronicity of C_A and V), while at the same time increasing C_A coupling strength in a C_R-C_A anti-phase relationship conditions the consecutive consonants to decrease in gestural overlap (cf. more C_AV and C_RC_A overlap in /pl/ compared to /mʃ/). In this principled fashion, we are able to bring together patterns of onset-vowel and intra-cluster organization into a global account of syllable organization (cf. Marin, 2013; Marin & Pouplier, 2014; Pouplier, 2012) by providing a novel interpretation to the coupling strength parameter so as it relates to the notion of coarticulatory resistance.

However, the results of /pna/ syllables complicate the already intricate interaction of cluster composition and syllable organization. Recall that /pna/ and /pla/ showed comparable onset-vowel but different inter-consonantal timing patterns, although we expected similar degrees of tongue body/jaw coarticulatory resistance for alveolar nasals and laterals. In addition, from a gestural account, the emergence of different C_RC_A timing patterns is surprising since the gestural compositions of /pna/ and /pla/ syllables are mostly identical. Only the velum gesture associated with the vowel-adjacent nasal differentiates the gestural compositions of /pna/ from /pla/. Based on this difference we hypothesized that an additional coupling relation between the nasal's velum and the labial's lip gesture could help to implement the conflicting aerodynamic requirements which are expected to be responsible for comparatively less $C_R C_A$ overlap in /pn/ than in /pl/ clusters. However, it is thus far not clear how the introduction of the velum gesture into a multiply-linked coupling topology interacts with coupling strength specifications and which implications arise thereof for the spatiotemporal syllable organization. A complex, controlled modeling experiment would be called for to settle this issue.

The results presented in Chapter 2 to 4 provide empirical evidence that spatiotemporal organization of syllables depends on the segmental make-up of the onset clusters. Particularly the degree of tongue body/jaw coarticulatory resistance of the vowel-adjacent consonant turned out to be reliable predictors for how cluster consonants are coordinated with the vowel and with each other. The gestural model is in principle ideally suited to model C-to-V coarticulation in terms of spatially and temporally overlapping gestures in $C_A V$ syllables, but previous to this thesis this principled capability has not been sufficiently expanded to explain differences among various complex onset types (i.e. $\#C_R C_A V$ with different consonant composition). In this thesis we could demonstrate that building coarticulatory resistance into the coupling strength parameter (whereby the gestural model of syllable organization assigns different bonding weights to competing phase relations) is a promising amendment of the gestural model, which thus allows coarticulatory resistance to condition the temporal overlap of gestures in the case of competing phase relationships. However, since we used arbitrarily specified coupling strength values for our TADA simulations, the exact evaluation and determination of coupling strength values should be addressed in further research. Preliminary TADA simulations suggest that these individual, context-independent coupling strengths need not be specified arbitrarily, but rather they can be derived in a principled way from specific articulator weights associated with each gesture.

If future systematic work confirms this, it would prove to be a valuable advancement for the gestural model and for our understanding of articulatory timing and speech production in general.

5.4 Future directions

One limitation of this dissertation is that we had to rely on different studies (e.g. Hoole et al., 1990; Lindblom, 1983; Mooshammer et al., 2007; Recasens, 2012b; Recasens & Espinosa, 2005, 2009; Recasens et al., 1997) to deduce individual degrees of tongue body and jaw coarticulatory resistance of Polish consonants since, to our knowledge, there is no independently established scale of coarticulatory resistance for Polish consonants (and vowels). This is particularly important since the degree of contextual variability of consonants and vowels is supposed to be language and speaker-specific to a certain extent (cf. Kühnert & Nolan, 1999). Since the available corpus of Polish articulography data does not allow for establishing an independent inventory of consonantal/vowel coarticulatory resistance, we are convinced that the recording and analysis of appropriate data is called for to better understand the language and segment-specific degrees of contextual variability, and how the respective degrees of contextual variability interact with the spatiotemporal organization of syllables. Furthermore, this would also contribute to our understanding of cross-linguistic typologies.

The production of C_RC_AV syllables is hypothesized to involve competing coupling relations of all consonants to the vowel (C_R-V and C_A-V), and the coupling of the consonants to one another (C_R-C_A) (Browman & Goldstein, 2000; Goldstein et al., 2009; Nam et al., 2009). Due to our suggestion that the context-independent coupling strength of C_R, C_A, and V is determined by the individual degrees of coarticulatory resistance, the first step would be to evaluate the contextual variability of Polish consonants and vowels. Following the methodology used in previous studies (e.g. Chen, Chang, & Iskarous, 2015; Iskarous et al., 2013; Recasens & Espinosa, 2009; Öhman, 1966), we suggest the recording of articulatory (EMA) data of (V)#CV sequences with varying consonants (e.g. labial /p/, /m/, apical/laminal /s/, /ʃ/, /n/, /l/ and velar /k/) and vowels (e.g. /i, a, u/). The data thus obtained could then be used to evaluate for different articulators (e.g. the jaw, the tongue body) the consonant- and vowel-specific degrees of contextual variability. Linear regression (e.g. Iskarous, Fowler, & Whalen, 2010), articulatory variability (DAC; e.g. Recasens, 2012b; Recasens & Espinosa, 2009), and most recently mutual information (MI; Iskarous et al., 2013) analyses have been previously applied to articulatory data to determine the degree of coarticulatory resistance/variance of respective target segments (cf. Chen et al., 2015 for a parallel application of these three analyses). Although all these analyses have been shown to be capable to assess the degree of contextual variability, Iskarous et al. (2013) claimed that MI constitutes an improvement over the DAC values: MI objectively assigns each segment a list of numbers for each articulator and component, quantitatively derived from speech data (e.g. vertical and horizontal variability of the tongue tip or body). In contrast, the DAC indices are empirically derived from coarticulatory variability measures, and then subjectively assigned to each segment with only one number per segment. Another advantage of MI for our purpose is that articulator independence (as measured by MI) could be a measure of articulator weights (Iskarous et al., 2013) which are used in Task Dynamics (Saltzman & Munhall, 1989) to specify synergies of various articulators to fulfill an articulatory task. The hypothesized link between MI and articulator weight is that the more important a particular articulator is to a constriction formation, the more likely it is to resist contextual encroachment. Since Goldstein et al. (2009) suggested that coupling strengths in competing phase relations may function like articulator weights in the task dynamics of constriction formation, we assume by extension that MI are in principle applicable for the determination of coupling strength values for consonants and vowels.

The EMA recordings of (V)#CV sequences should also include (V)#C_RC_AV sequences produced by the same speakers in which the consonants used in the singleton condition (i.e. #CV) should be combined into heterorganic consonant clusters with varying vowel contexts (e.g. /i, a, u/). Although the onset-vowel and intra-cluster timing measures used in Chapter 2 to 4 have proved successful in revealing commonalities and differences between conditions, we suggest, however, some expansions of the measurements outlined in Chapter

2 in order to estimate the temporal and spatial relations between both cluster consonants and the vowel (i.e. C_R-V and C_A-V). Recall that the degree of coarticulatory resistance of both consonants (C_R and C_A) and the vowel are expected to condition onset-vowel timing. Regarding the temporal lag measures this means for a syllable like /pla/ that relative C_R-V (i.e. /p/-/a/) and C_A-V (i.e. /l/-/a/) timing should be compared with the respective singleton conditions (i.e. /p/-/a/ in /pa/ and /l/-/a/ in /la/ syllables, respectively) to learn more about the consonant-specific relative timing changes as a function of increasing onset complexity (e.g. Honorof & Browman, 1995; Marin & Pouplier, 2010). In addition, deriving from the methodologically novel spatial measure, we assume that the cluster-vowel overlap could also be measured in terms of V-to-C_R and V-to-C_A coarticulation by means of landmark statistics (cf. Pastätter & Pouplier, 2015). By defining various 'landmarks' (i.e. measurement points) throughout consonant clusters in different vowel contexts (e.g. /pli/ vs. /pla/ vs. /plu/), we believe that it is possible to determine at which point in time the vowel starts to encroach C_R (or C_A in the case of comparatively lower onset-vowel overlap degrees). This would be informative in the context of Chen et al.'s (2015) recent suggestion that coarticulatory resistance is a concept that characterizes consonants whereas aggression is a concept that characterizes vowels. If the degree of temporal overlap between the vowel-remote consonant (C_R) and the vowel is spatially quantifiable by this measure, this would then provide an opportunity to deduce the instantiation of the C_R-V phasing in actual timing as a function of the identity of the consonant and vowel involved.

To our knowledge, no methodological approach has been yet developed to determine coupling strength values of consonants and vowels from temporal (or spatial) coordination pattern. Therefore we suggest an evaluation and determination of the respective coupling strength values through back-fitting of the segment-specific coupling strength values based on quantitative modeling of syllable organization patterns by using TADA (Nam et al., 2004). Methodologically, this approach would involve the modeling of speaker and syllable-specific onset-vowel and intra-cluster timing patterns (obtained from spatiotemporal measures on articulatory data) by means of quantitative variation of C_R-V, C_A-V, and C_R-C_A coupling relations (cf. Goldstein et al., 2009). Note that individual degrees of coarticulatory resistance/variance should also be considered during quantitative modeling; hence, we expect, using this approach, to derive exact (i.e. segment-specific) coupling strength values for different types of consonants and vowels.

We conclude from this that parallel examination of coupling strength and coarticulatory resistance/variance may help us to associate in a principled way segmental coupling strength values with the corresponding degrees of coarticulatory resistance, which would enrich the gestural model of syllable organization with predictive power. Specifically, the coupling relations specified at the planning level would be allowed to vary in a predictable, principled way as a function of coarticulatory resistance of the consonants and vowels involved. Although our proposed concept of context-independent coupling strength values is in principle capable to explain inter-speaker variability in syllable organization (i.e. due to speaker-specific differences in coarticulatory behavior, due in turn possibly to anatomical differences), it is not clear how to account for syllable timing variability within one speaker (i.e. less stable coordination patterns). Therefore, it seems suitable to ask how syllable organization patterns (i.e. in terms of C_R-V, C_A-V, C_R-C_A coupling relations) interact with prosodic variation (cf. Chapter 4). Further, since our remarks only consider syllables with onset clusters, it is currently unclear what the implications of our theoretical enhancements are for the organization of coda clusters. Previous studies on vowel-coda organization reported fairly complicated timing patterns, partly deviating from the gestural models' prediction (cf. Byrd, 1995; Marin, 2013; Marin & Pouplier, 2010, 2014; Pouplier, 2012). It may be that incorporating coupling strength values may also help explain those coda timing patterns deviating from the predictions of the model in its simpler instantiation. We therefore recommend further investigation of syllables with complex codas for future research (accompanied by modeling work) to gain a deeper understanding of how coarticulatory resistance of coda consonants interacts with vowel-coda timing.

Appendix A

Chapter supplements

A.1 Chapter 2

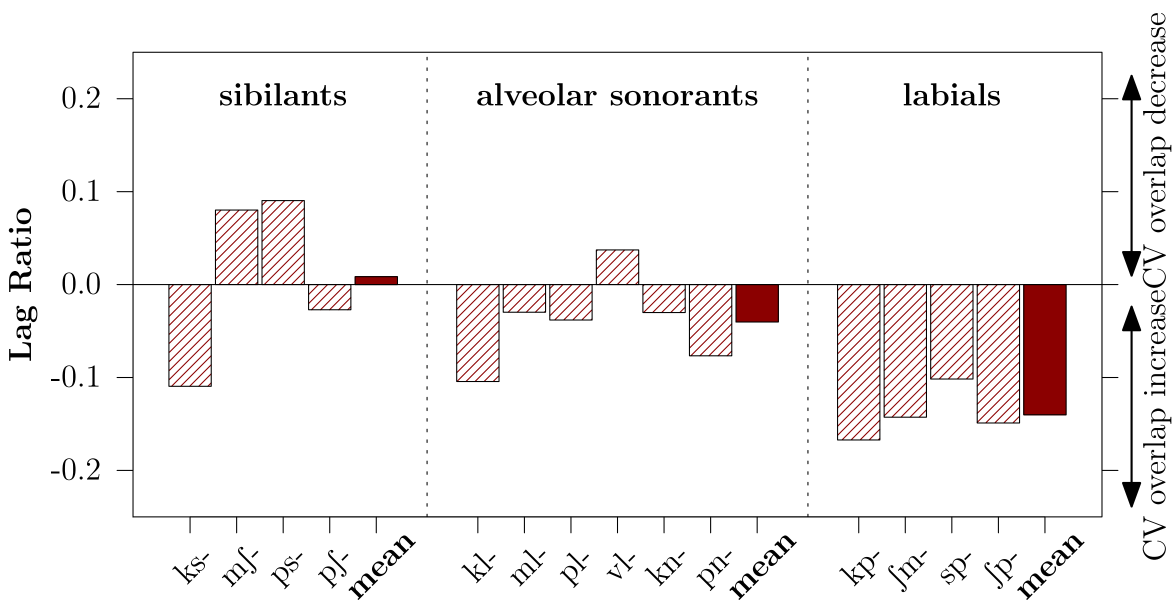

Figure A.1: Lag ratios for the individual sets. Negative lag ratios indicate increased CV overlap in the cluster relative to the singleton condition; positive lag ratios indicate that the CV overlap decreased between cluster and singleton. Filled bars represent the means shown in Figure 2.5.

A.2 Chapter 3

In this Appendix we show and discuss two representative examples illustrating the different articulatory strategies and their respective influences on the coarticulatory overlap measure.

Representative for the 'reasonable' tokens, the top panel of Figure A.2 shows a /pl/ token of subject PL4. In this example the tangential velocity profile of the tongue tip sensor (i.e. $\mathrm{TT}_{\mathrm{Velocity}}$) exhibits two particular peaks corresponding to the tongue tip movement towards (1) and away (2) from the targeted constriction location. These two peak velocities (i.e. PVEL1 and PVEL2) allow for an appropriate establishment of the constriction plateau in the tongue tip sensor trajectory (i.e. dark gray portion in the $\mathrm{TT}_{\mathrm{Position(x,y)}}$ tier). In this example, the degree to which /p/ ($\mathrm{C_R}$) overlapped the constriction plateau of /l/ ($\mathrm{C_A}$) amounts to 141.7%.

The bottom panel of Figure A.2 shows an outlier /pl/ token of subject PL5. In this example, three peaks can be clearly detected in the $\mathrm{TT}_{\mathrm{Velocity}}$ tier of which (1) and (2) relate to the global movements of the tongue tip towards and away from the constriction location (cf. (1) and (2) in top panel for the 'reasonable' /pl/ token). However, the distinct peak (3) corresponds to an intermediate adjustment of the tongue tip position in the horizontal plane, i.e. after approaching the intended vertical tongue tip position, the tongue tip moves forward to achieve the target /l/ position. Instead of choosing (1) and (2), the *mview* algorithm automatically detected peak (1) and (3) as PVEL1 and PVEL2, respectively, and determined accordingly a very short constriction plateau (i.e. dark gray portion in the $\mathrm{TT}_{\mathrm{Position(x,y)}}$ tier). Manual intervention had no satisfactory effect. Therefore, the degree to which /p/ ($\mathrm{C_R}$) overlapped the constriction plateau of /l/ ($\mathrm{C_A}$) in this example amounts 600.0%.

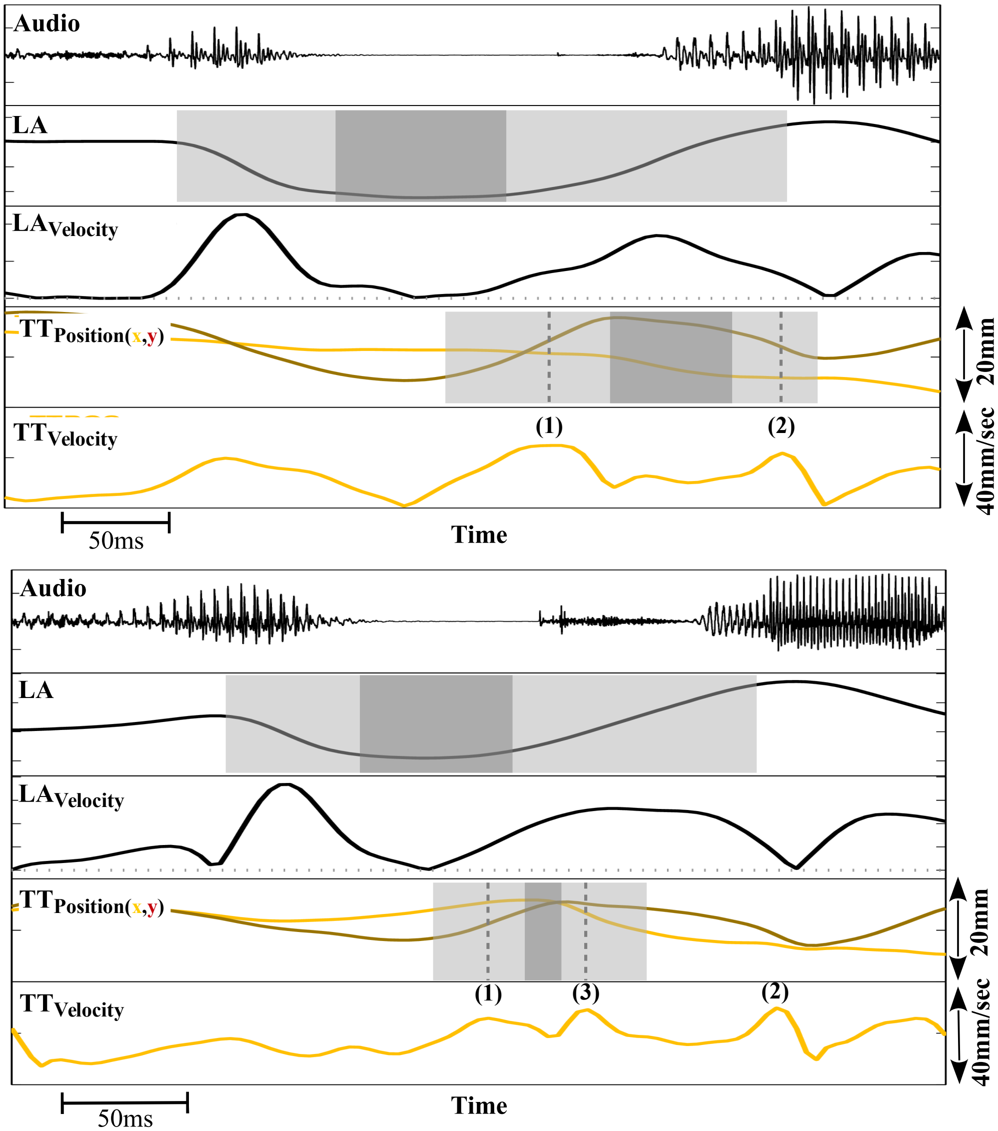

Figure A.2: Articulatory data of two /pla/ items are given in order to illustrate the origin of the peculiar mean coarticulatory overlap indices and *SD* values for /pla/ presented in Table 3.2. Both panels show representative overlap examples for 'reasonable' (top; speaker PL4) and outlier (bottom, speaker PL5) /pl/ items. The different articulatory strategies in approaching the tongue tip target position (cf. respective $TT_{Position(x,y)}$ and $TT_{Velocity}$ tiers) have biased the articulatory labeling and and the coarticulatory overlap measure.

Appendix B

Speech material

		Experimental corpus	
Onset		**Target word**	**Chapter**
/mʃ/	[mʃalik]	*mszalik*, ‘missal’ (nom.sg.)	2, 3, 4
/ʃ/	[ʃalik]	*szalik*, ‘scarf’ (nom.sg.)	2, 3
/pʃ/	[pʃɛraʑ]	*przeraź*, ‘to terrify sb with sth’ (imp.perf.)	2, 4
/ʃ/	[ʃerek]	*szereg*, ‘row’ (nom.sg.)	2
/ps/	[psɔtʲɲe]	*psotnie*, ‘mischievously’ (inf.)	2
/s/	[sɔtʲɲa]	*sotnia*, ‘term for (historic) military unit’ (nom.sg.)	2
/ml/	[mlɛkax]	*mlekach* ‘milk’ (loc.pl.)	2, 4
/l/	[lekaʃ]	*lekarz*, ‘doctor’ (nom.sg.)	2
/sp/	[spɔdʲɲɛ]	*spodnie*, ‘trousers’ (nom.pl.)	2
/p/	[pɔdʲɲɛt]	*podniet*, ‘incentive’ (gen.pl.)	2
/vl/	[vlit͡ʃɨ]	*wliczy*, ‘to factor in’ (3ps.sg.pres)	2
/l/	[lit͡ʃi]	*liczi* ‘lychee’ (nom.sg.)	2
/ʃp/	[ʃpɛrat͡ɕ]	*szperać* ‘to rummage in sth.’ (inf.)	2
/p/	[pɛrɔ̃n]	*peron*, ‘station platform’ (nom.sg.)	2

Table B.1 – *Continued on next page*

Table B.1 – *Continued from previous page*

/pl/	[plat͡som]	*placom*, ‘square, public open space’ (dat.pl.)	2, 3, 4
/l/	[latom]	*latom*, ‘years’ (dat.pl.)	2, 3
/ʃm/	[ʃmata]	*szmata*, ‘rag’ (nom.sg.)	2
/m/	[mata]	*mata*, ‘mat’ (nom.sg.)	2
/ks/	[ksɛrɔ]	*ksero*, ‘copy’ (nom.sg.)	2
/z/	[zɛrɔ]	*zero*, ‘zero’ (nom.sg.)	2
/kl/	[klut͡ʃɨk]	*kluczyk*, ‘(diminutive) key’ (nom.sg.)	2
/l/	[lud͡ʑik]	*ludzik*, ‘(diminutive) man, figurine’ (nom.sg.)	2
/kp/	[kpʲinom]	*kpinom*, ‘mockery’ (dat.pl.)	2
/p/	[pʲinom]	*piniom*, ‘pine’ (dat.pl.)	2
/pn/	[pnat͡ɕi]	<logatom>	2, 3, 4
/n/	[nat͡ɕi]	*naci*, ‘(carrot) tops’ (gen./dat.sg.)	2, 3
/kn/	[knurɨ]	*knury*, ‘boar’ (nom.pl.)	2
/n/	[nurek]	*nurek*, ‘diver’ (nom.sg.)	2
/pt/	[pt̪akem]	*ptakem*, ‘bird’ (instr.sg.)	3
/t/	[t̪acim]	*takim*, ‘such’ (masc.loc.sg., among others)	3
/v/	[vit͡sɛ]	*wice*, ‘vice-’ (nom.sg.)	2
/mn/	[mɲixom]	*mnichom*, ‘monk’ (dat.pl.)	4

Table B.1: List of singleton and cluster target words used in this thesis.

Appendix C

Previous works

The author of this thesis has previously contributed to following research:

Pouplier, M., Tiede, M. & Pastätter, M. (2011). Using sine-wave speech to examine the perception-action link in speech. *Supplement 6th International Conference on Speech Motor Control* (p. 46). Groningen, Netherlands.

Pouplier, M., Pastätter, M. & Schiel, F. (2011). Spontaneously occurring speech errors in the BAS corpora. *Supplement 6th International Conference on Speech Motor Control*, (p. 93). Groningen, Netherlands.

Pastätter, M. & Pouplier, M. (2013). Temporal coordination of sibilants in Polish onset clusters. *Journal of the Acoustical Society of America*, *134*, 4201.

Pastätter, M. & Pouplier, M. (2014). The articulator modeling of German coronal consonants using TADA. In *Proceedings of the 10th International Seminar on Speech Production* (pp. 308–311). Cologne, Germany.

Pastätter, M. & Pouplier, M. (2014). The temporal coordination of Polish onset and coda clusters containing sibilants. In *Proceedings of the 10th International Seminar on Speech Production* (pp. 312–315). Cologne, Germany.

Pouplier, M., Hoole, Ph., Cederbaum, J., Greven, S. & Pastätter, M. (2014). Perceptual and articulatory factors in German fricative assimilation. *Proceedings of the 10th*

International Seminar on Speech Production (pp. 332–335). Cologne, Germany.

Pastätter, M. & Pouplier, M. (2015) Onset-vowel timing as a function of coarticulation resistance: Evidence from articulatory data. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, Scotland.

Pastätter, M. & Pouplier, M. (2017) Articulatory mechanisms underlying onset-vowel organization. *Journal of Phonetics, 65*, 1–14.

Bibliography

Amerman, J., Daniloff, R., & Moll, K. (1970). Lip and jaw coarticulation for the phoneme /æ/. *Journal of Speech and Hearing Research*, *163*, 147–161.

Bates, D. (2010). *lme4: Mixed-effects modelling with R*. Springer.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Beňuš, Š., & Pouplier, M. (2011). Jaw movement in vowels and liquids forming the syllable nucleus. In *Proceedings of the 12th INTERSPEECH* (pp. 389–392). Florence, Italy.

Bladon, R. A. W., & Al-Bamerni, A. (1976). Coarticulatory resistance in English /l/. *Journal of Phonetics*, *4*, 137–150.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Bombien, L. (2011). *Segmental and prosodic aspects in the production of consonant clusters – On the goodness of clusters* (Doctoral dissertation). Ludwig-Maximilians-University Munich.

Bombien, L., Mooshammer, C., & Hoole, P. (2013). Articulatory coordination in word-initial clusters of German. *Journal of Phonetics*, *41*(6), 546–561.

Bombien, L., Mooshammer, C., Hoole, P., & Kühnert, B. (2010). Prosodic and segmental effects on EPG contact patterns of word-initial German clusters. *Journal of Phonetics*, *38*, 388–403.

Browman, C. (1994). Lip aperture and consonant releases. In P. A. Keating (Ed.), *Phonological structure and phonetic form: Papers in Laboratory Phonology III* (pp.

331–353). Cambridge: Cambridge University Press.

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

Browman, C., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, *45*, 140–155.

Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*, 201–251.

Browman, C., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 340–376). Cambridge: Cambridge University Press.

Browman, C., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*, 155–180.

Browman, C., & Goldstein, L. (1995). Gestural syllable position effects in American English. In F. Bell-Berti & L. J. Raphael (Eds.), *Producing speech: Contemporary issues. for Katherine Safford Harris* (pp. 19–34). Woodbury, NY: AIP Press.

Browman, C., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée*, *5*, 25–34.

Bruni, J. (2011). *Sonorant voicing specification in phonetic, phonological and articulatory context* (Doctoral dissertation). Universität Stuttgart.

Brunner, J., Geng, C., Sotiropoulou, S., & Gafos, A. (2014). Timing of German onset and word boundary clusters. *Laboratory Phonology*, *5*(4), 403–454.

Bukmaier, V., & Harrington, J. (2016). The articulatory and acoustic characteristics of Polish sibilants and their consequences for diachronic change. *Journal of the International Phonetic Association*, *46*(3), 311–329.

Byrd, D. (1995). C-centers revisited. *Phonetica*, *52*, 285–306.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, *24*(3), 209–244.

Byrd, D., & Choi, S. (2010). At the juncture of prosody, phonology, and phonetics – The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. In C. Fougeron, B. Kühnert, M. P. d'Imperio, & N. Vallée (Eds.), *Laboratory Phonology X: Variation, Detail and Representation* (pp. 31–60). Berlin, New York: Mouton de Gruyter.

Byrd, D., & Tan, C. C. (1996). Saying consonant clusters quickly. *Journal of Phonetics*, *24*, 263–282.

Byrd, D., Tobin, S., Bresch, E., & Narayanan, S. (2009). Timing effects of syllable structure and stress on nasals: A real-time MRI examination. *Journal of Phonetics*, *37*, 97–110.

Catford, J. S. (1988). *A practical introduction to phonetics*. Oxford: Clarendon Press.

Chen, W., Chang, Y., & Iskarous, K. (2015). Vowel coarticulation: Landmark statistics measure vowel aggression. *Journal of the Acoustical Society of America*, *138*(2), 1221–1232.

Chitoran, I., & Cohn, A. (2009). Complexity in phonetics and phonology: Gradience, categoriality, and naturalness. In F. Pellegrino, E. Marisco, I. Chitoran, & C. Coupé (Eds.), *Approaches to Phonological Complexity* (pp. 21–46). Berlin: Mouton de Gruyter.

Chitoran, I., Goldstein, L., & Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology VII* (pp. 419–448). Berlin: Mouton de Gruyter.

Christie, W. M. (1977). Some multiple cues for juncture in English. *General Linguistics*, *17*, 212–222.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology i* (pp. 283–333). Cambridge: Cambridge University Press.

Dart, S. (1991). *Articulatory and acoustic properties of apical and laminal articulations*. UCLA Phonetics Laboratory.

Dave, H. (2012). *Frequency word list for Polish 2012*. https://invokeit.wordpress.com/

frequency-word-lists/. ([Online; accessed 06/02/2014])

Davis, B. L., & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research*, *38*, 1199–1211.

de Jong, K. J. (2003). Temporal constraints and characterizing syllable structure. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI* (pp. 253–268). Cambridge University Press.

Erickson, D. (1998). Effects of contrastive emphasis on jaw opening. *Phonetica*, *55*, 147–169.

Farnetani, E. (1990). V-to-V lingual coarticulation and its spatiotemporal domain. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (Vol. 55, pp. 93–130). Dordrecht: Kluwer Academic Publishers.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, *36*(4), 359–368.

Fowler, C. A., & Brancazio, L. (2000). Coarticulatory resistance of American English consonants and its effects on transconsonantal Vowel-to-Vowel coarticulation. *Language and Speech*, *43*, 1–41.

Fowler, C. A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, *36*, 171–195.

Fowler, C. A., & Thompson, J. (2010). Listeners' perception of compensatory shortening. *Attention, Perception & Psychophysics*, *72*, 481–491.

Gafos, A. (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory*, *20*, 269–337.

Gafos, A., Hoole, P., Roon, K., & Zeroual, C. (2010). Variation in timing and phonological grammar in Moroccan Arabic clusters. In C. Fougeron, B. Kühnert, M. P. d'Imperio, & N. Vallée (Eds.), *Laboratory Phonology X: Variation, Detail and Representation* (pp. 657–698). Berlin, New York: Mouton de Gruyter.

Gahl, S. (2008). "Thyme" and "Time" are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*, 474–496.

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

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. A. Arbib (Ed.), *Action to Language via the Mirror Neuron System* (pp. 215–249). Cambridge: Cambridge University Press.

Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In N. O. Schiller & A. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 159–207). Berlin: Mouton de Gruyter.

Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. In G. Fant, H. Fujisaki, & J. Shen (Eds.), *Frontiers in Phonetics and Speech Science* (pp. 239–250). Beijing: The Commercial Press.

Gussmann, E. (2007). *The phonology of polish*. Oxford: Oxford University Press.

Haggard, M. (1973). Abbreviation of consonants in english pre- and post-vocalic clusters. *Journal of Phonetics*, *1*(1), 9–24.

Hamann, S. (2004). Retroflex fricatives in Slavic languages. *Journal of the International Phonetic Association*, *34*(1), 53–67.

Hanson, H., & Stevens, K. (2002). A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn. *Journal of the Acoustical Society of America*, *112*, 1158–1182.

Harrington, J. (2006). An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics*, *34*, 439–457.

Harrington, J. (2010). *Phonetic analysis of speech corpora*. Chichester: Wiley-Blackwell.

Harrington, J., Fletcher, J., & Roberts, C. (1995). Coarticulation and the accented/unaccented distinction: Evidence from jaw movement data. *Journal of Phonetics*, *23*, 305–322.

Henke, E., Kaisse, E. M., & Wright, R. (2012). Is the Sonority Sequencing Principle an epiphenomenon? In S. Parker (Ed.), *The Sonority Controversy* (pp. 65–100). Berlin: Mouton de Gruyter.

Hermes, A., Mücke, D., & Grice, M. (2013). Gestural coordination of Italian word initial clusters – the case of 'impure s'. *Phonology*, *30*(1), 1–25.

Honorof, D., & Browman, C. (1995). The center or edge: How are consonant clusters organized with respect to the vowel? In *Proceedings of the 13th International Congress of Phonetic Sciences* (pp. 552–555). Stockholm, Sweden.

Hoole, P., Bombien, L., Kühnert, B., & Mooshammer, C. (2009). Intrinsic and prosodic effects on articulatory coordination in initial consonant clusters. In G. Fant, H. Fujisaki, & J. Shen (Eds.), *Frontiers in phonetics and speech science* (pp. 275–286). Beijing: Commercial Press.

Hoole, P., Gfroerer, S., & Tillmann, H. G. (1990). Electromagnetic articulography as a tool in the study of lingual coarticulation. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München*, *28*, 107–122.

Hoole, P., Pouplier, M., Beňuš, Š., & Bombien, L. (2013). Articulatory coordination in obstruent-sonorant clusters and syllabic consonants: Data and modelling. In L. Spreafico & A. Vietti (Eds.), *Rhotics. New Data and Perspectives* (pp. 81–97). Bolzano University Press.

Hooper, J. B. (1972). The syllable in phonological theory. *Language*, *48*, 525–540.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363.

Hughes, G. W., & Halle, M. (1956). Spectral properties of fricative consonants. *Journal of the Acoustical Society of America*, *28*, 303–310.

Iskarous, K., Fowler, C. A., & Whalen, D. H. (2010). Locus equations are an acoustic expression of articulator synergy. *Journal of the Acoustical Society of America*, *128*(4), 2021–2032.

Iskarous, K., McDonough, J., & Whalen, D. H. (2012). A gestural account of the velar fricative in Navajo. *Laboratory Phonology*, *3*(1), 195–210.

Iskarous, K., Mooshammer, T., Recasens, D., Hoole, P., Shadle, C., Saltzman, E., &

Whalen, D. H. (2013). The coarticulation/invariance scale: Mutual information as a measure of coarticulation resistance, motor synergy, and articulatory invariance. *Journal of the Acoustical Society of America*, *134*(2), 1271–1282.

Iskarous, K., Pouplier, M., Marin, S., & Harrington, J. (2010). The interaction between prosodic boundaries and accent in the production of sibilants. In *Proceedings of the 5th International Conference on Speech Prosody.* Chicago, IL.

Iskarous, K., Shadle, C. H., & Proctor, M. I. (2011). Articulatory-acoustic kinematics: The production of American English /s/. *Journal of the Acoustical Society of America*, *129*(2), 944–954.

Jassem, W. (2003). Polish. *Journal of the International Phonetic Association*, *33*(1), 103–107.

Jongman, A. (1989). Duration of frication noise required for identification of English fricatives. *Journal of the Acoustical Society of America*, *85*(4), 1718–1725.

Joos, M. (1948). Acoustic phonetics. *Language*, *24*, 1–136.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. L. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 229–254). Amsterdam: Benjamins.

Keating, P. (1983). Comments on the jaw and syllable structure. *Journal of Phonetics*, *11*, 401–406.

Keating, P., Lindblom, B., Lubker, J., & Kreiman, J. (1994). Variability in jaw height for segments in English and Swedish VCV sequences. *Journal of Phonetics*, *22*, 407–422.

Kirby, J. P. (2014). Acoustic transitions in Khmer word-initial clusters. In *Proceedings of the 10th International Seminar on Speech Production* (pp. 234–237).

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

Krakow, R. A. (1989). *The articulatory organization of syllables: A kinematic analysis of labial and velar gestures* (Unpublished doctoral dissertation). Yale University, New Haven.

Krakow, R. A. (1993). Nonsegmental inuences on velum movement pattems: Syllables, sentences, stress, and speaking rate. In M. K. Huffman & R. A. Krakow (Eds.), *Nasals, nasalization, and the velum (phonetics and phonology* (Vol. 5, pp. 87–114). New York: Academic Press.

Krakow, R. A. (1999). Physiological organization of syllables: A review. *Journal of Phonetics*, *27*, 23–54.

Kühnert, B., Hoole, P., & Mooshammer, C. (2006). Gestural overlap and C-center in selected French consonant clusters. In *Proceedings of the 7th International Seminar on Speech Production* (pp. 327–334).

Kühnert, B., Ledl, C., Hoole, P., & Tillmann, H. G. (1991). Tongue-jaw interactions in lingual consonants. In *Perilus xiv.*

Kühnert, B., & Nolan, F. (1999). The origin of coarticulation. In W. J. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques* (pp. 7–30). Cambridge: Cambridge University Press.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages.* Oxford: Blackwell.

Lee, S., Beckman, M., & Jackson, M. (1994). Jaw targets for strident fricatives. In *Proceedings of icslp* (pp. 37–40). Yokohama, Japan.

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*, 239–269.

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed.), *The Production of Speech* (Vol. 55, pp. 217–245). New York: Springer.

Lindblom, B. (2011). The units of speech – A bio-cultural perspective. In *Perilus 2011, symposium on language acquisition and language evolution* (pp. 1893–1896). Stockholm, Sweden.

Löfqvist, A., & Gracco, V. L. (1999). Interarticulator programming in VCV sequences: Lips and tongue movements. *Journal of the Acoustical Society of America*, *105*(3), 1864–1876.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production.

Behavioral and Brain Sciences, *21*, 499–546.

MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, *288*, 527–531.

MacNeilage, P. F., & Davis, B. L. (2001). Motor mechanisms in speech ontogeny: phylogenetic, neurobiological and linguistic implications. *Current Opinion in Neurobiology*, *11*, 696–700.

MacNeilage, P. F., Davis, B. L., Matyear, C. L., & Kinney, A. (2000). The motor core of speech. *Child Development*, *71*, 153–163.

Malisz, Z., & Żygis, M. (2015). Voicing in Polish: Interactions with lexical stress and focus. In *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow, Scotland.

Marin, S. (2011). Organization of complex onsets in Romanian. In *Proceedings of the 9th International Seminar on Speech Production* (pp. 179–186).

Marin, S. (2013). The temporal organization of complex onsets and codas in Romanian: A gestural approach. *Journal of Phonetics*, *41*, 211–227.

Marin, S. (2014). Intra-cluster timing in Romanian stop-initial onsets. In *Proceedings of the 10th International Seminar on Speech Production* (pp. 265–268).

Marin, S., & Pouplier, M. (2010). Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control*, *14*(3), 380–407.

Marin, S., & Pouplier, M. (2014). Articulatory synergies in the temporal organization of liquid clusters in Romanian. *Journal of Phonetics*, *42*, 24–36.

Marin, S., Pouplier, M., & Kochetov, A. (2015). Timing patterns of word-initial obstruent-sonorant clusters in russian. In *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow, Scotland.

Mattingly, I. G. (1981). Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech* (pp. 415–420). Amsterdam: North Holland.

Menzerath, P., & de Lacerda, A. (1933). *Koartikulation, Steuerung und Lautabgrenzung:*

Eine experimentelle Untersuchung. Berlin, Bonn: Ferd. Dümmlers Verlag.

Mermelstein, P. (1973). Articulatory model for the study of speech production. *Journal of the Acoustical Society of America*, *53*(4), 1070–1082.

Mooshammer, C., Hoole, P., & Geumann, A. (2007). Jaw and order. *Language and Speech*, *50*, 145–176.

Munson, B., & Solomon, N. P. (2004). The influence of phonological neighborhood density on vowel articulation. *Journal of Speech, Language and Hearing Research*, *47*, 1048–1058.

Nam, H. (2007). Syllable-level intergestural timing model: Split-gesture dynamics focusing on positional asymmetry and moraic structure. In J. Cole & J. I. Hualde (Eds.), *Papers in Laboratory Phonology IX: Phonology and Phonetics* (pp. 507–534). Berlin: Mouton de Gruyter.

Nam, H., Goldstein, L., Giulivi, S., Levitt, A. G., & Whalen, D. H. (2013). Computational simulation of CV combination preferences in babbling. *Journal of Phonetics*, *41*(2), 63–77.

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. In F. Pellegrino, E. Marsico, I. Chitoran, & C. C. (Eds.), *Approaches to phonological complexity* (pp. 299–328). Berlin/New York: Mouton de Gruyter.

Nam, H., Goldstein, L., Saltzman, E., & Byrd, D. (2004). TADA: An enhanced, portable task dynamics model in MATLAB. *Journal of the Acoustical Society of America*, *115*(5,2), 2430.

Nam, H., Mitra, V., Tiede, M., Hasegawa-Johnson, M., Espy-Wilson, C., Saltzman, E., & Goldstein, L. (2012). A procedure for estimating gestural scores from speech acoustics. *Journal of the Acoustical Society of America*, *132*(6), 3980–3989.

Nowak, P. M. (2006). *Vowel reduction in Polish* (Doctoral dissertation). University of California, Berkeley.

O'Shaughnessy, D. (1974). Consonant durations in clusters. *IEEE Transactions on Acoustics Speech and Signal Processing*, *22*, 282–295.

Pastätter, M., & Pouplier, M. (2014). The articulator modeling of German coronal consonants using TADA. In *Proceedings of the 10th International Seminar on Speech Production* (pp. 308–311).

Pastätter, M., & Pouplier, M. (2015). Onset-vowel timing as a function of coarticulation resistance: Evidence from articulatory data. In *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow, Scotland.

Pastätter, M., & Pouplier, M. (2017). Articulatory mechanisms underlying onset-vowel organization. *Journal of Phonetics*, *65*, 1–14.

Perkell, J. (1969). *Physiology of speech production.* Cambridge, Mass.: MIT Press.

Peters, S. (2015). *The effects of syllable structure on consonantal timing and vowel compression in child and adult speakers of German* (Doctoral dissertation). Ludwig-Maximilians-Universität, München.

Peters, S., & Kleber, F. (2014). Articulatory mechanisms underlying incremental compensatory vowel shortening in German. In *Proceedings of the 10th International Seminar on Speech Production* (pp. 316–319).

Piotrowski, M. (1992). Polish yers in non-linear phonology. In W. Dressler, H. C. Luschutzky, O. E. Pfeiffer, & J. R. Rennison (Eds.), *Phonologica 1988* (pp. 215–227). Cambridge: Cambridge University Press.

Pouplier, M. (2012). The gestural approach to syllable structure: Universal, language- and cluster-specific aspects. In S. Fuchs, M. Weirich, D. Pape, & P. Perrier (Eds.), *Speech planning and dynamics* (pp. 63–96). Peter Lang.

Pouplier, M. (2015). Between 'whims of fashion' and 'phonetic law': Performance constraints in speech production in the face of linguistic diversity. In M. Vayra, C. Avesani, & F. Tamburin (Eds.), *Language Acquisition and Language Loss: Acquisition, Change and Disorders of the Language Sound Structure* (pp. 41–57). Milano, Italy: AISV.

Pouplier, M., & Beňuš, Š. (2011). On the phonetics status of syllabic consonants: Evidence from Slovak. *Laboratory Phonology*, *2*, 243–273.

Pouplier, M., Marin, S., Hoole, P., & Kochetov, A. (in press). Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics*.

Przepiórkowski, A. (2013). *Distributable subcorpus of National Corpus of Polish.* http://metashare.dfki.de/repository/browse/distributable-subcorpus-of-national-corpus-of-polish/d4e5be386aff11e284b6000423bfd61c757435ed76e74b098068569fe55d6bb9/. ([Online; accessed 02/19/2014])

Pęzik, P. (2011). *PELCRA Polish spoken corpus.* http://metashare.elda.org/repository/browse/pelcra-polish-spoken-corpus-cc-by-nc/e47ff96863f111e2bff4525400d76147c34334267af0462daa1f95b542c203d9/. ([Online; accessed 06/02/2014])

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/ (ISBN 3-900051-07-0)

Recasens, D. (2012a). A cross-language acoustic study of initial and final allophones of /l/. *Speech Communication*, *54*, 368–383.

Recasens, D. (2012b). A study of jaw coarticulatory resistance and aggressiveness for Catalan consonants and vowels. *Journal of the Acoustical Society of America*, *132*, 412–420.

Recasens, D., & Espinosa, A. (2005). Articulatory, positional and coarticulatory characteristics for clear /l/ and dark /l/: Evidence from two Catalan dialects. *Journal of the International Phonetic Association*, *35*, 1–25.

Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *Journal of the Acoustical Society of America*, *125*(4), 2288–2298.

Recasens, D., Pallars, M. D., & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America*, *102*, 544–561.

Redford, M. A. (1999a). *An articulatory basis for the syllable* (Doctoral dissertation).

University of Texas, Austin.

Redford, M. A. (1999b). The mandibular cycle and reversed-sonority onset clusters in Russian. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1893–1896). San Francisco, CA.

Redford, M. A. (2004). Origin of consonant duration patterns. In A. Agwuele, W. Warren, & S.-H. Park (Eds.), *Proceedings of the 2003 Texas Linguistics Society Conference* (pp. 54–61). Somerville, MA: Cascadilla Proceedings Project.

Redford, M. A., & van Donkelaar, P. (2008). Jaw cycles and linguistic syllables in adult English. In B. L. Davis & K. Zajdo (Eds.), *Syllable Development: The Frame/Content Theory and Beyond* (pp. 355–376). London, UK: Taylor & Francis.

Rochet-Capellan, A., & Schwartz, J.-L. (2007). An articulatory basis for the labial-coronal effect: /pata/ seems a more stable articulatory pattern than /tapa/. *Journal of the Acoustical Society of America*, *121*(6), 3740–3754.

Rochoń, M. (2000). *Optimality in complexity: The case of polish consonant clusters*. Berlin: Akademie Verlag.

Rocławski, B. (1976). *Zarys fonologii, fonetyki, fonotaktyki i fonostatystyki współczesnego języka polskiego*. Gdańsk: Wydawnictwo Uczelniane Uniwersytetu Gdańskiego.

Rubach, J., & Booij, G. (1990). Syllable structure assignment in Polish. *Phonology*, *7*(1), 121–158.

Ruszkiewicz, P. (1990). Aspiration in English and Polish: An overview. *Papers and Studies in Contrastive Linguistics*, *25*, 147–161.

Saltzman, E. (1986). Task dynamic co-ordination of the speech articulators: A preliminary model. In H. Heuer & C. Fromm (Eds.), *Generation and Modulation of Action Patterns* (pp. 129–144). Berlin: Springer-Verlag.

Saltzman, E., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, *1*(4), 332–382.

Scarborough, R. A. (2004). *Coarticulation and the structure of the lexicon* (Unpublished doctoral dissertation). UCLA, Los Angeles, CA..

Schiel, F. (2004). Maus goes iterative. In *Proceedings of the 4th International Conference* *on Language Resources and Evaluation* (pp. 1015–1018).

Selkirk, E. (1982). The syllable. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representation* (pp. 337–383). Dordrecht: Foris.

Shadle, C. (1985). *The acoustics of fricative consonants* (Ph.D. thesis). MIT, Los Angeles, CA..

Smith, C. L. (2004). Prosodic patterns in the coordination of vowel and consonant gestures. In B. Connell & A. Arvaniti (Eds.), *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence* (pp. 205–222). Cambridge: Cambridge University Press.

Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, *21*(3), 291–311.

Stone, G. (1990). Polish. In B. Comrie (Ed.), *The World's Major Languages* (pp. 349–366). New York and Oxford: OUP.

Stone, M., & Vatikiotis-Bateson, E. (1995). Trade-offs in tongue, jaw, and palate contributions to speech production. *Journal of Phonetics*, *23*, 81–100.

Summers, W. V. (1987). Effects of stress and final consonant voicing on vowel production: articulatory and acoustic analyses. *Journal of the Acoustical Society of America*, *82*(3), 847–863.

Sussex, R., & Cubberley, P. (2006). *The slavic languages*. Cambridge University Press.

Syrika, A., Nicolaidis, K., Edwards, J., & Beckman, M. E. (2011). Acquisition of initial /s/-stop and stop-/s/ sequences in Greek. *Language and Speech*, *54*(3), 361–386.

Tomaschek, F., Wieling, M., Arnold, D., & Baayen, R. H. (2013). Word frequency, vowel length and vowel quality in speech production: An EMA study of the importance of experience. In *Proceedings of the 14th INTERSPEECH* (pp. 1302–1306). Lyon, France.

Turvey, M. T. (1990). Coordination. *American Psychologist*, *77*, 938–953.

Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America*, *61*(3), 846–858.

Vallée, N., Rossato, S., & Rousset, I. (2009). Favoured syllabic patterns in the world's languages and sensorimotor constraints. In F. Pellegrino, E. Marisco, I. Chitoran,

& C. Coupé (Eds.), *Approaches to Phonological Complexity* (pp. 111–139). Berlin: Mouton de Gruyter.

Waniek-Klimczak, E. (2011a). Aspiration and style: A sociophonetic study of the VOT in Polish learners of English. In M. Wrembel, M. Kul, & K. Dziubalska-Kolaczyk (Eds.), *Achievements and perspectives in SLA of speech: New Sounds 2010* (pp. 303–316). Frankfurt: Peter Lang.

Waniek-Klimczak, E. (2011b). Aspiration in Polish: A sound change in progress? In M. Pawlak & J. Bielak (Eds.), *New perspectives in language, discourse and translation studies* (pp. 3–11). Berlin and Heidelberg: Springer Verlag.

Wierzchowska, B. (1980). *Fonetyka i fonologia języka polskiego*. Wrocław: Zakład Narodowy im. Ossolińskich.

Wood, S. (1979). A radiographic analysis of constriction location for vowels. *Journal of Phonetics*, *7*, 25–43.

Wright, R. (1996). *Consonant clusters and cue preservation in Tsou* (Doctoral dissertation). UCLA.

Wright, R. (1997). Lexical competition and reduction in speech: A preliminary report. In *Research on spoken language processing: Progress Report No. 21* (pp. 471–485). Bloomington: Indiana University.

Wright, R. (2004). Factors of lexical competition in vowel articulation. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in Laboratory Phonology VI* (pp. 26–50). Cambridge: Cambridge University Press.

Yip, J. C. K. (2013). *Phonetic effects on the timing of gestural coordination in Modern Greek consonant clusters* (Doctoral dissertation). University of Michigan.

Zsiga, E. C. (1994). Acoustic evidence for gestural overlap in consonant sequences. *Journal of Phonetics*, *22*(2), 121–140.

Öhman, S. E. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, *39*(1), 151–168.

Żygis, M., & Hamann, S. (2003). Perceptual and acoustic cues of Polish coronal fricatives. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 395–398).

Barcelona, Spain.

Acknowledgements

This thesis has been supported by the ERC under the EU's 7th Framework Programme (FP/2007-2013)/Grant Agreement n. 283349 "Stability and change in sound patterns of the world's languages: typological, physiological, and cognitive factors".

First of all, I am deeply grateful to Marianne Pouplier who brought me into her ERC project. Her patient support with excellent comments and reflections were so many times helpful to get me back on track. She never got tired to guide me (despite my perseverance) and to push me further and further into abstract worlds of gestures and algorithms. All this has been indispensable for me and this work.

Further, I would like to thank the current (Stefania Marin, Lia Saki Bučar Shigemori) and former (Susanne Waltl) colleagues during the ERC project for their collaboration in different perspectives and the numerous IPS members who aided me with valuable comments and suggestions, not only in the course of the PhD/Postdoc meetings. Phil Hoole and Jonathan Harrington are here particularly mentioned.

Beside the scientific area, I am very obliged to all my endearing friends inside and outside of Munich who took the load off me, who lent a sympathetic ear, who gave me moral support, who distracted me, and who provided me good reasons to leave all the work far, far behind me. I will never forget the recreative times I had together with all of you in such pleasant spots like Fålaboda in summer, Zell am See in winter, Zottegem and Prestatyn somewhere between, and of course Cologne, Berlin, and Frankfurt where I always found a shelter to escape. To be honest, I could not have done this without your support. You mean everything to me, cheers!